PhD-FSTM-2020-47
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 04/09/2020 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN Informatique

by

Andrea CAPPONI

Born on 4 February 1984 in La Spezia (Italy)

ENERGY-EFFICIENT MOBILE CROWDSENSING SOLUTIONS FOR SMART CITIES

Dissertation Defence Committee

Dr Ulrich Sorger, Dissertation Supervisor
Professor, Université du Luxembourg

Dr Luca Foschini
Professor, University of Bologna

Dr Pascal Bouvry, Chairman
Professor, Université du Luxembourg

Dr Burak Kantarci
Professor, University of Ottawa

Dr Dzmitry Kliazovich, Vice Chairman
ExaMotive, Luxembourg

Abstract

This thesis proposes energy-efficient mobile crowdsensing (MCS) solutions for smart cities. Specifically, it focuses on sensing and communications processes in distributed computing paradigms and complex urban dynamics in city-wide scenarios. MCS is a data collection paradigm that has gained significant attention in recent years and has become appealing for urban sensing. MCS systems rely on contributions from mobile devices of a large number of participants or a crowd. Smartphones, tablets, and wearable devices are deployed widely and already equipped with a rich set of sensors, making them an excellent source of information. Mobility and intelligence of humans guarantee higher coverage and better context awareness if compared to traditional sensor networks. At the same time, individuals may be reluctant to share data for devices' battery drain and privacy concerns. For this reason, MCS frameworks are specifically designed to include incentive mechanisms and address privacy concerns.

Despite the growing interest in the research community, MCS solutions still need a more in-depth investigation and categorization on many aspects that span from sensing and communication to system management and data storage. This Ph.D. thesis focuses not only on sustainable MCS solutions to challenging problems in urban environments but also on a comprehensive study aiming to clarify concepts, aspects, and inconsistencies in existing literature from a global perspective. Specifically, this manuscript proposes the following contributions:

- Present the MCS paradigm as a four-layered architecture divided into application, data, communication, and sensing layers, proposing novel taxonomies related to each layer. The detailed taxonomy aims to shed light on the current landscape, covering all MCS aspects and allowing for a simple and clear classification of applications, methodologies, and architectures.

- A significant improvement of the previously developed simulation environment CrowdSenSim by implementing a set of novel features. The novelties include easy-to-use city-wide street networks, more realistic pedestrian mobility models, and real battery drain measurements over several other features.

- An analysis of energy efficiency that poses the basis for sustainable MCS data collection frameworks (DCFs). It includes both a theoretical methodology to assess different DCFs and real energy measurements conducted in a laboratory, simulated in large scale urban environments.
- A study that exploits crowdsensed data for a learning-driven estimation of local businesses' attractiveness in cities to show how MCS systems can support urban planning.
- A novel efficient edge data centers deployment in real urban environments based on human mobility and traffic generated from mobile devices. The citizens' mobility is developed by feeding CrowdSenSim with crowdsensed data.

Acknowledgements

This manuscript comes after four years of ups and downs, challenges, experiences, and adventures. During this journey, I have met influential people who have contributed to my work and, more importantly, my life. To all of them, I am and will always be very grateful.

First of all, I would like to thank my supervisor Uli who gave me this opportunity. From the first day, he showed me what a supervisor should be like: a kind person who gives suggestions but leaves you free and independent to draw your path. I feel genuinely grateful for that. I would also like to thank the other committee members and Professors that always supported me. Pascal, for his endless support since I was an intern during my Master Thesis and for always pushing me into different research activities around the world. Burak, who I met in conferences and contributed from remote with precious advice. Luca, who hosted me in his department and has been much more than a Professor, guiding me not only for research but also for delicious and endless food in Bologna. I would also like to thank Prof. Zheng, who hosted me in Canada in a period that will always be unforgettable. If I've arrived here, it is also because my former advisor, Prof. Giordano, has always believed in me and strongly supported me to start an internship abroad.

I really don't know how I could appropriately thank Dzmitry. More than five years ago, he gave me the opportunity for an internship in Luxembourg and always supported, guided, and helped me. Before, as a Master student in his Thesis and a graduate looking for a PhD. Then, as a PhD student during his studies. Later, as a PhD candidate searching the inspiration to find an exciting way out of academia.

I can't forget the people who work at the University and had an impact on my studies. Catherine and her incredible support, I would never have been in Canada or China without her fantastic organizational help, as well as for courses and credits. Geneviève and her help with bureaucratic procedures. Antoine and his constant and endless IT support in really everything. Fabienne, who was very tolerant of me and my frequent requests.

I will never thank Claudio enough for all he has done since the first day we met at Kirchberg. He has been a tutor, a colleague, a friend, and a supporter. He has always and tirelessly helped me, even when it wasn't necessary, and I've learnt so much that I will never forget. Most of all, he taught me that when you feel overwhelmed and think you'll never be able to do something, just start doing anything, and everything will come by itself. And it did.

Special thanks to my Master students and colleagues Mattia and Piergiorgio. They have significantly contributed to this work and have soon become friends more than students. We started with an interview, we ended up skiing together and enjoying the sunset in Hawaii. With them, I want to thank the PhD students met around the world and become friends. Most of all Nat, met in Lipari, for all PhD experiences we shared, starting with an octopus.

A great thank you to the people that were the "Italians in Lux", but have become much more than that. In particular, Emiliano the boss of Caprera, Rubens and his child birthdays as Italian, Filippo my front door mate, Maria and our Portos. A particular thank to Luca, the first person I met in Luxembourg, who was fundamental in helping me to adapt to the shock.

What could I ever say to thank Ilaria enough? No idea. Simply, the second person I met at my arrival in a hostel in Luxembourg and the person who has contributed the most to my personal growth over the past 6 years. She has helped me face many challenges while she had much bigger challenges than me. I've learnt so many things from her that I could never list all of them. The only grateful thought I could say to her is: keep strong and get what you deserve, because you deserve it! Whatever it takes [contemporary cit.].

My friends from Spezia and surroundings. Berta, Dario, Mario, Pinetto, Matte, Daniele, Fra, Krull, Via Istria 12, and others. Last year was difficult and it's not clear how the next one will be. I just miss you. So much. People change, little people grow up, they are far and I lose moments that will never come back but I'm everyday thankful for their support wherever I am.

An immense thank you, with love, to my family. They have always supported me, even when supporting me was difficult. I would have never arrived where I am without their patience. Being away for many years and having that support felt so strongly is not easy, but they do it great, and I couldn't be more grateful. I dedicate this manuscript to them.

The most important thank goes to Gianlu. Deskmate and so many other things, who has always tried to make me a better person (a little better, it's hard work) and who most pushed me not to get angry when I don't need to. My biggest thought goes to Gaia, Damiano, and his family. My biggest hope is that better times will come.

Per correr miglior acque alza le vele
omai la navicella del mio ingegno,
che lascia dietro a sé mar sì crudele;

***A** ne ghe credo.*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

In recent years, mobile crowdsensing (MCS) has become a promising and appealing data collection paradigm to monitor urban environment phenomena. MCS systems rely on contributions from the smart devices of a large number of citizens. Smartphones, tablets, and wearable devices are deployed widely and are already equipped with a rich set of sensors, making them an excellent source of information. Mobility and intelligence of humans guarantee higher coverage and better context awareness if compared to traditional sensor networks. The capillary spread of smart devices for our daily activities and the rich set of built-in sensors are certainly the main key enablers leading to the success of the MCS paradigm [1], [2]. According to Gartner statistics, whether the first quarter of 2020 has seen a 20% decrease in the number of worldwide smartphone sales due to COVID-19 impact, in 2019, it was 1.52 billion units [3]. The number of worldwide shipped wearable devices is estimated to reach 453.19 million in 2022 [4]. Smart objects (e.g., glasses, watches, helmets) are continually increasing their market and correspond to a revenue predicted to reach USD 95.3 billion by 2021 [5]. Also, the crowd analytics market has a compound annual growth rate of 24.3%, and it is projected to reach USD 1 142.5 million by 2021, rising from USD 385.1 million of 2016 [6].

Many popular applications using smartphone sensors have been developed and are currently in use. To illustrate representative examples, MCS can support studies for infrastructure management in civil engineering, such as monitoring structural vibrations of a bridge. A real use case in the Harvard Bridge (Boston, US) is given in [7], where data gathered from smartphones' accelerometer used by taxi driver for navigation systems provide information on the modal frequencies of the bridge. Glutensensor collects data to foster healthy food by sharing images between celiac people and extracting context information to map and rate restaurants and places [8]. Safestreet aggregates data from smart devices to monitor road surface conditions for safer driving and less risk of car accidents [9]. GasMobile [10], HazeWatch [11], and Third-Eye [12] rely on active citizen participation to monitor air pollution. Creekwatch [13], developed by the IBM Almaden research center, permits the monitoring of the conditions of the 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 [14] and WasteApp [15] allow monitoring the content of different bins to improve the recycling program.

Smart devices can contribute an unlimited amount of sensed data, which needs to be stored. As resources are locally limited, reporting data to a central collector for processing and analysis represents a win-win solution to enforce crowd intelligence [16], [17]. Distributed computing systems enable access to gathered data and shared resources easily. On the one hand, the cloud computing paradigm represents a central shared infrastructure that provides a ubiquitous approach for efficient data management [18]. On the other hand, moving the intelligence closer to the end-users by exploiting paradigms like fog [19] and multi-access edge computing (MEC) [20] is a win-win strategy to enhance the performance of MCS applications.

MCS can significantly improve citizens' everyday life and provide new perspectives to urban societies, being an essential enabler for building smart cities of the future by exploiting ICT solutions [21], [22]. Moreover, citizens' active participation can improve the spatial coverage of deployed sensing systems with no need for further investments. While urbanization is intensifying, smart cities face significant deficits in infrastructure services. In this context, MCS represents a promising approach to involving humans to improve urban infrastructures' monitoring and maintenance.

1.2 Motivation

MCS systems rely on users that contribute data gathered from their smart devices and delivered to a central collector. Smartphones, tablets, and wearable devices are widely deployed and equipped with a rich set of sensors, making them an excellent source of information. An MCS campaign requires citizens' broad participation to be effective. Still, individuals may be reluctant to join a campaign or contributing data, and the motivation is mainly due to costs for sensing and delivery operations. Hence, collecting and reporting data must not drain devices' batteries to foster users' extensive participation. It is crucial to devise energy-efficient data collection frameworks (DCFs) for a successful campaign and assess their performance according to different key performance indicators (KPIs), depending on the campaign organizer. The trade-off between the amount of gathered data and the sensing process's energy consumed is one of the most challenging issues that still require investigation. The desired objectives are to obtain a high amount and quality of contributed data (i.e., to maximize the utility of sensing) with low energy consumption, aiming to limit the collection of low-quality data. In literature, only a few works propose sustainable MCS solutions, and none of them scales to city-wide scenarios with realistic pedestrian mobility. When this research started, no studies had so far analyzed the amount of collected data and the associated energy costs of DCFs for large scale MCS campaigns (e.g., thousands of users that move in city-wide scenarios over multiple days). This Thesis fills this gap by proposing sustainable MCS solutions that enable the development of different services in smart cities and presenting novel taxonomies to bring clarity to the vast world of MCS systems. The manuscript also illustrates consistent improvements to a previously developed MCS simulator and discusses real use cases for urban environments.

1.3 List of Contributions

This Ph.D. thesis aims to investigate energy-efficient MCS systems, focusing on how smart cities' complex urban dynamics impact them. The most significant contributions can be summarized as follows:

- A comprehensive study of existing literature presents MCS in a nutshell and outlines the absence of unambiguous terms and definitions of basic concepts.
- A description of the MCS data collection paradigm as a four-layered architecture, divided into application, data, communication, and sensing layers.
- Novel detailed taxonomies based on the layered architecture, which shed light on the current landscape and classify applications, methodologies, and architectures by covering all MCS aspects.
- Additional features in the previously founded simulation environment called CrowdSenSim, which outperforms other existing tools in the MCS scenario. Novel features include real devices' energy consumptions and more easy-to-use realistic pedestrian mobility models for any real-world city street network.
- A novel energy-efficient MCS system and a methodology to compare different data collection frameworks. This approach includes models, real measurements performed in a laboratory, and large scale simulations with CrowdSenSim.
- A learning-based estimation of local businesses' attractiveness based on data crowdsensed from users.

- A study to efficiently deploy edge data centers in urban environments based on human mobility. It shows how moving the intelligence closer to end-users with paradigms like multi-access edge computing (MEC) is a win-win strategy to perform MCS operations quickly.

1.4 Thesis Structure

The manuscript is organized as follows:

- **Chapter 2** presents basic notions that are essential to read the dissertation. The preliminaries include a brief overview of smart cities (Section 2.1) and mobile crowdsensing (MCS) (Section 2.2), basic concepts on energy efficiency to develop sustainable MCS solutions in urban environments (Section 2.3), and a discussion on distributed computing paradigms, such as cloud computing, fog computing, and multi-access edge computing (MEC) (Section 2.4).

- **Chapter 3** presents MCS in a nutshell. It illustrates a primer with a historical analysis (Section 3.1), discusses the main factors contributing to the rise of MCS (Section 3.2), and proposes MCS as a layered architecture (Section 3.3).

- **Chapter 4** proposes a novel comprehensive taxonomy on MCS systems. First, it explains the need for a clear taxonomy that sheds light on unclear and undefined aspects, then illustrates detailed taxonomies on the application (Section 4.1), data (Section 4.2), communication (Section 4.3), and sensing layers (Section 4.4).

- **Chapter 5** presents a background on simulators to assess the performance of MCS activities in urban environments (Section 5.1), illustrates the CrowdSenSim structure and modules (Section 5.2), proposes novel features that consistently improve the previous version of the simulator, such as city-wide scenarios and realistic pedestrian mobility models (Section 5.3).

- **Chapter 6** analyzes the energy efficiency in MCS data collection frameworks (DCFs). First, it presents an overview of related works and DCFs under analysis (Section 6.1). Then, it profiles the energy consumption of DCFs with real experimental measurements (Section 6.2), proposes a novel methodology for large scale analysis (Section 6.3), and evaluates the performance by exploiting CrowdSenSim (Section 6.4).

- **Chapter 7** illustrates a crowdsensed data-driven approach to estimate local businesses' attractiveness. After discussing background and motivation on urban planning (Section 7.1) and a preliminary data analysis

(Section 7.2), it proposes an ML-augmented methodology that exploits pedestrian mobility based on crowdsensed data (Section 7.3). Then, it discusses a data-driven evaluation (Section 7.4).

- **Chapter 8** proposes an approach to deploy edge data centers in urban environments based on citizens' mobility. First, it discusses background and motivation (Section 8.1). Then, it presents models for EDCs and complex urban dynamics (Section 8.2). Finally, it proposes policies to deploy EDCs (Section 8.3) evaluating them (Section 8.4).

- **Chapter 9** concludes the work by discussing future directions, interconnections with other research areas (Section 9.1), and concluding remarks (Section 9.2).

Fig. 1.1 shows the organization of this dissertation.

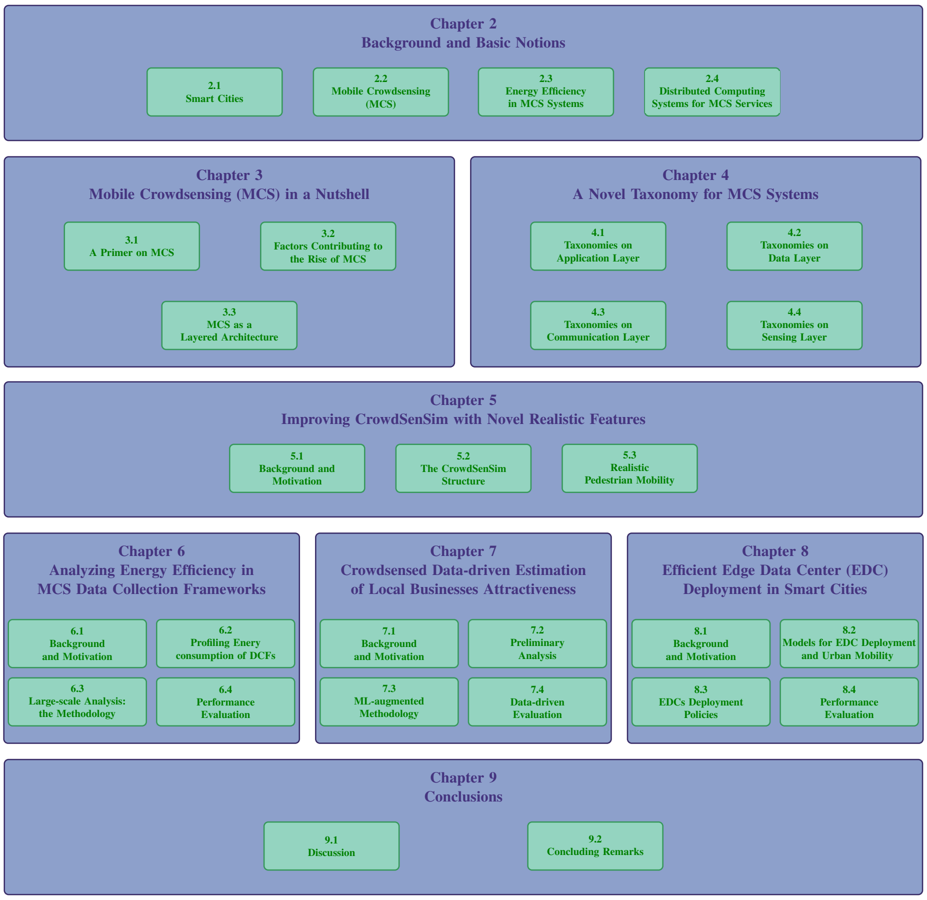


Figure 1.1: Dissertation Structure

Chapter 2

Background and Basic Notions

This chapter presents the background and discusses basic notions to understand the main topics of the thesis better. First, it introduces the concept of smart cities, which plays a central role in the sustainable development of metropolitan areas. Second, it explains how mobile crowdsensing (MCS) represents a promising data collection paradigm to monitor phenomena in this context. Then, it presents energy efficiency as one of the most challenging issues to face. Finally, it discusses the most popular distributed computing paradigms that support MCS services.

2.1 Smart Cities

Over the past century population living in urban areas has experienced unprecedented growth. Only 10% of the worldwide population lived in cities during the 20th century, while today the 50% lives in metropolitan areas. This percentage is estimated to increase consistently in the next three decades, up to reaching 68% of the global population by 2050 [23]. Hence, providing efficient and sustainable solutions plays a fundamental role in supporting urban growth and the lifestyle of citizens[1]. While cities occupy only 2% of the Earth's surface, urban areas contribute to 60% of water use, 80% of world gas emission, and 75% of energy consumption [24]. In this context, it is crucial to monitor resources, their usage, and the underlying infrastructure.

Nowadays, cities face complex challenges to support sustainable and efficient development while ensuring citizens' quality of life. To this end, consistent research efforts are undergoing to provide innovative solutions for public services. Smart cities exploit Information and Communication Technology (ICT) solutions to improve citizens' quality of life by adding value to existing public infrastructure and services. Although there is no

[1] In the rest of the thesis, the terms citizens, crowd, participants, and users will be used interchangeably

widely accepted and precise definition of the smart city concept, it essentially consists of taking advantage of ICT developments to benefit citizens. In other words, a smart city aims to make available services for citizens by exploiting an ICT platform while reducing efforts and costs required from institutions, public administrations, and companies. The Internet of Things (IoT) paradigm represents the candidate solution for deploying sensing infrastructures empowering smart cities' applications in metropolitan areas [25]. The widespread diffusion of IoT devices enables the IoT paradigm's urban deployment, making the Internet more pervasive with connected objects and bringing several benefits to develop sustainable ICT platforms for smart cities. Public services essential for the community (e.g., public transportation, street lamps, waste collection) can quickly become smart when equipped with sensing, computing, storage, and communication capabilities.

Sustainable development requires continually monitoring resources (e.g., gas and water), their usage, and different phenomena that directly and indirectly impact the citizens' quality of life, such as air pollution, waste management, and natural disasters. In this context, sensing represents a crucial role in gathering the massive amount of data required to monitor phenomena, resources, and infrastructures' current status. Including citizens in the loop of sensing represents a win-win solution for smart city applications because it augments existing infrastructure capabilities and permits gathering a large amount of data. To this end, mobile crowdsensing is a novel and promising paradigm that allows collecting data from citizens' smart devices by exploiting ICT platforms' capabilities without deploying new sensing infrastructures and introducing related additional costs.

2.2 Mobile Crowdsensing (MCS)

This Section briefly introduces mobile crowdsensing (MCS). It provides basic notions about how it represents a promising data collection paradigm for smart cities, the most common MCS scenario, and some popular applications. MCS will be discussed more in detail in Chapter 3.

2.2.1 A Promising Paradigm for Sustainable Data Collection

MCS has recently become a popular and appealing paradigm that can significantly improve citizens' quality of life and provide unexpected urban societies perspectives. It is an essential solution for building smart cities of the future that leverages sensing and communication capabilities [21], [22] provided by common smart devices (e.g., smartphones, tablets, and wearables) used in everyday life activities, such as communication, entertainment, healthcare, and business [1], [2].

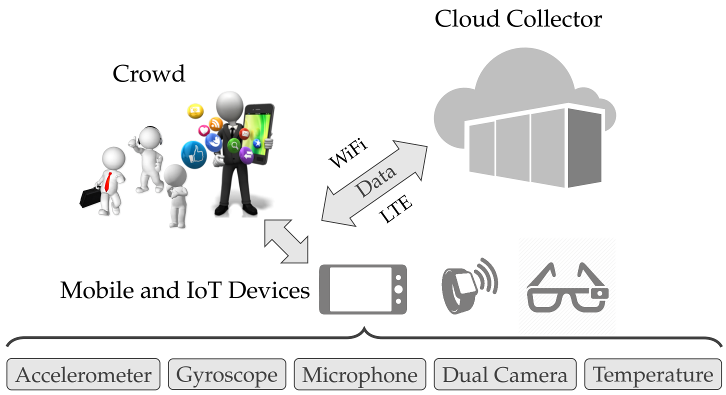


Figure 2.1: The MCS cloud-based scenario

Ganti et al. introduced the term *mobile crowdsensing* in 2011 [26], indicating an evolution and extension to the crowd of the mobile phone sensing paradigm, which can is the forefather of MCS [27], [28]. Mobile phone sensing was adopted when phones did not support current smart devices' computation and communication capabilities. It mainly focused on individual applications (e.g., personal health care, elderly fall detection). A definition from Guo et al. clarifies the difference between these two paradigms [29]: *"MCS is a new sensing paradigm that empowers ordinary citizens to contribute data sensed or generated from their mobile devices, aggregates and fuses the data in the cloud for crowd intelligence extraction and people-centric service delivery".* MCS campaigns need broad participation and data contribution from a crowd of citizens to be effective. An individual may be reluctant to share collected data for privacy concerns or device battery drain, and it is crucial to investigate how to foster participation. A great research effort is undergoing in recent years to investigate incentive mechanisms [30], [31], [32], [33] and ensure privacy [34], [35].

2.2.2 The MCS Scenario

Fig. 2.1 illustrates the most common scenario of an MCS system and its fundamental elements. The most relevant feature that characterizes MCS systems is including humans in the sensing process loop, which has revealed a win-win strategy [40]. Citizens utilize smartphones, tablets, wearable, and IoT devices in everyday life, representing a widespread data source for MCS systems. These smart devices have sensing and communication capabilities that enable sensing and reporting data to a central collector. Unlike traditional fixed sensor networks, these factors represent unlimited possibilities to develop data collection campaigns with different purposes to improve citizens' quality of life. First, smart devices represent sensors and communication nodes recharged continuously and maintained by users without any external intervention. Second, human intelligence and mobility provide higher coverage and a deeper context-awareness than traditional sensor networks with no need for further investments [41]. Accelerometer, gyroscope, magnetometer, camera, and GPS are only a few representative examples of sensors commonly embedded in mobile and IoT devices. Smart devices sense and deliver data to a central database where information is stored and made available to campaign organizers and stakeholders (e.g., academic and governmental institutions, companies, corporate businesses). The central collector is usually placed in the cloud and responsible for data storage, analysis, and processing.

2.2.3 Popular MCS Applications

As previously discussed, urban areas are facing consistent issues while deploying infrastructure services for supporting sustainable development. This Section discusses specific applications and use cases to illustrate how human involvement in monitoring and maintenance can represent a win-win solution. MCS is essential to enable several applications in public transportation, health care, environmental and traffic monitoring, emergency management, and many other domains [28]. To give some representative examples, Creekwatch empowers the monitoring of watershed through crowdsensed data, such as the amount of water in the river bed, trash in the riverbank, and the flow rate [13]. The National Environment Agency of Singapore uses HazeWatch to leverage crowd contribution for air monitoring [11]. Waste management is a crucial topic in sustainable smart cities of the future. Garbage Watch [14] and WasteApp [15] are two applications that employ citizens to monitor the content of bins aiming to improve recycling. Accelerometers embedded in smartphones over moving vehicles can help to detect bridge vibrations [7]. MCS allows traffic management [42], [43] and free parking spot detection, such as ParkSense that uses WiFi scans [44] or ParkGauge that permits to share real-time information and detect driving states with low-consuming sensors [45].

2.3 Energy Efficiency in MCS Systems

Providing solutions for more sustainable and greener development is the most challenging issue we are facing worldwide. In this context, proposing efficient and smart ICT systems can significantly reduce energy consumption. For several reasons, the greener future's scope attracts companies, private citizens, public institutions, and governments. First, energy production consistently affects the environment and requires reducing carbon footprints and gas emissions. Second, and most important for industries, sustainable progress has a relevant economic impact on bills. Also, companies care about green initiatives to gain more visibility for their brand on the market and more attractivity to customers. Nowadays, researchers are making a great effort to investigate energy-efficient approaches in distributed computing systems and communication networks. In particular, data collection paradigms that have a low impact on energy consumption represent a great research focus.

In this context, MCS campaigns require a vast amount of information from smart devices, and battery consumption should be as low as possible not to limit participants from contributing data. Most of the energy consumed by mobile devices consists of sensing and reporting operations, which depend on selected sensors and communication technologies. Usually, the energy spent on delivery has a more significant impact than delivery. In MCS systems, the energy efficiency can be seen as a trade-off between devices' battery drain and key performance indicators of campaigns, such as quality of information, space coverage, amount of data [46], [47]. The MCS paradigm also enables to develop sustainable data collection frameworks to avoid battery waste and encourage the involvement of private citizens [48].

2.4 Distributed Computing Systems for MCS Services

In MCS campaigns, smart devices contribute a considerable amount of data that needs to be stored, but local storage presents minimal capabilities. Delivering data to distributed computing systems for processing and analysis represents a win-win solution that enforces crowd intelligence [16], [17]. Distributed computing systems refer to paradigms that split computational problems into small tasks executed by multiple entities, aiming to improve performance and efficiency. MCS applications nowadays typically exploit cloud computing, multi-access edge computing (MEC), and fog computing among several distributed systems [37], [49], [50].

Cloud computing enables the access of shared storage and computational infrastructure, assuring efficient data management with a ubiquitous approach [18], [51]. Nonetheless, the widespread diffusion of smart devices makes it difficult for the cloud computing paradigm to fulfill the consistent increase of high-performance and low-latency requirements from mobile applications [52], [53]. To this end, the scope of fog computing and MEC paradigms is to move the intelligence closer to end-users, representing a win-win strategy for data collection paradigms like MCS that need to perform operations quickly [54], [55], [56]. Fog computing was proposed as a cloud extension by Cisco [19]. A fog platform typically combines many layers made of a high number of nodes that provide storage, computation, and communication capabilities (e.g., base stations, access points, and gateways) [57]. Its peculiarity is that some layers are close to end-users. The European Telecommunications Standards Institute (ETSI) standardized MEC in early 2017, changing the concept of mobile edge computing into multi-access edge computing to highlight it as the enabler of multiple access technologies and cellular radio [58]. It makes available application-oriented capabilities at a one-hop distance between end-used and core of mobile operators' networks [20]. To illustrate with a few representative examples, EdgeSense is a MEC-based MCS system that exploits a secured peer-to-peer network for environmental monitoring [59]. RMCS is a Robust MCS architecture that integrates MEC resources and deep learning to minimize the transmission latency [60].

Chapter 3

Mobile Crowdsensing (MCS) in a Nutshell

This Chapter discusses MCS in a nutshell, focusing on reasons that made it a prominent paradigm and motivations giving rise to it. Specifically, Section 3.1 overviews related surveys to guide the reader into past research and shows the temporal evolution of pillar works in the area. Then, Section 3.2 provides an overview of the main factors empowering MCS. Finally, Section 3.3 describes MCS as a layered architecture, illustrating each layer in detail.

3.1 A Primer on MCS

This Section presents a primer on MCS by examining and selecting from the large body of literature the most relevant works that have guided research in this field. First, it illustrates surveys related directly and indirectly to MCS. Then, it discusses works that represent *milestones* by considering their temporal evolution.

3.1.1 Related Surveys

MCS is a subject that embraces many different aspects, not necessarily related between them (e.g., sensing and communication equipment). For this reason, when examining related works and research efforts that contributed to MCS, it is essential to consider different topics. This Section explores related works by dividing them into five different main topics: MCS, mobile phone sensing, wireless sensor networks (WSN), user recruitment, and privacy concerns.

Mobile Crowdsensing. Mobile Crowd Sensing and Computing (MCSC) is the first name adopted by Guo et al. [61] to investigate how machine and human intelligence are complementary in sensing and computing operations. Users are network nodes that exploit their smart devices to gather and deliver information to the cloud, which has storage and computing capabilities to empower crowd intelligence. Phuttharak et al. discuss the fundamental characteristics of MCS architectures, focusing on task allocation, user recruitment, data collection, and processing [63]. The most crucial challenges of MCS systems and solutions to effectively use resources are discussed in [64]. Abualsaud et al. present several MCS applications in the context of IoT and smart cities [65]. Xu at al. investigate interactions between MCS and social networks (e.g., Twitter), focusing on public security and location-based services [66]. To successfully accomplish MCS campaigns, it is crucial to assess the Quality of Information (QoI), which still requires investigation efforts. In [67], Restuccia et al. propose a framework to evaluate and improve the QoI of collected data.

Mobile Phone Sensing. Mobile phone sensing represents the ancestor of mobile crowdsensing. This paradigm was popular when mobile phones did not have storage, communication, and computation capabilities as recent advanced smart devices. Unlike MCS, investigation on mobile phone sensing concentrated on personal sensing applications, such as individual well-being or elderly fall detection. Existing literature in this field is vast, including solutions for sensing and methodologies to aggregate data for context awareness [27], [28].

Sensors & WSN. Sensors represent the most fundamental element in MCS and the first step of the sensing process to gather data. Typically smart devices embed sensing equipment, but some applications (e.g., environmental monitoring [10], [68], [69]) requires specific sensors connected wirelessly (e.g., Bluetooth). Ming et al. examine standard sensors embedded in mobile devices and related application [70]. Participants in MCS campaigns represent nodes of mobile and dynamic Wireless Sensor Networks (WSNs). For this reason, the broad literature on WSNs is crucial to empower MCS systems. Several surveys investigate how the progress in sensing, communication, and networking technologies contributed to empowering WSNs significantly [71], [72], [73].

User Recruitment. The extensive participation of citizens and their data contribution is critical for the success of an MCS campaign. User recruitment and incentive mechanisms are crucial to enhance users' willingness for sensing and delivery operations. In [75], [76], the authors propose taxonomies and use cases for incentive strategies. Zhang et al. discuss and compare several research works that present mechanisms to recruit users and stimulate them to contribute data [31].

Privacy Concerns. Privacy is one of the most challenging issues to address when collecting data from citizens. Even out of this dissertation's scope, investigating privacy concerns in MCS is fundamental not to prevent user contribution. The literature in this field is vast and includes many different

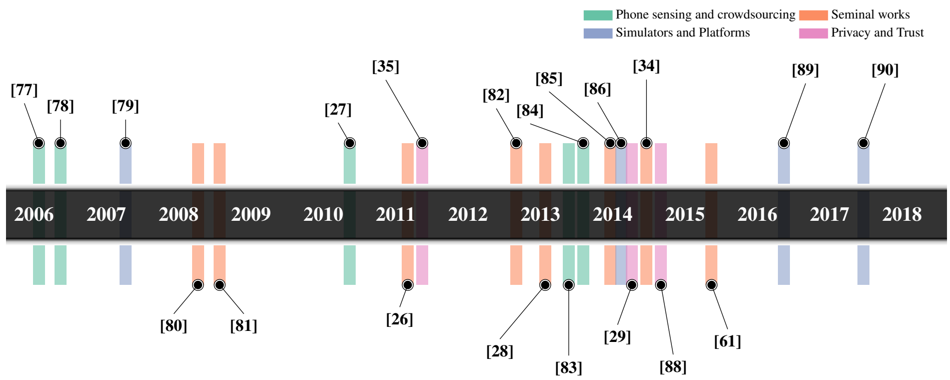


Figure 3.1: Historical evolution of MCS milestones works

aspects and threats, such as tracking locations or disclosing private information in image, audio, and video files. Several studies analyze privacy concerns in user recruitment, task allocation, and selection [34]. Christin et al. provide an overview of MCS applications and related threats to individual privacy, focusing on how existing solutions address them and proposing other countermeasures [35].

3.1.2 Historical Evolution of Milestones Works

This section examines the pillar works that significantly contributed to MCS by presenting their temporal evolution. Fig. 3.1 illustrates the time of publication and divides the research studies into four groups:

- Mobile phone sensing and crowdsourcing;
- Seminal works;
- Simulators and platforms;
- Privacy and trust.

The following paragraphs discuss the milestones works by year to uncover the time evolution of MCS.

2006. The term *crowdsourcing* initially appeared in a study published by Howe [77] to provide a clear definition of this rising paradigm and discuss its first applications and use cases. Burke et al. later proposed the concept of *participatory sensing* as a promising paradigm to exploit citizens and their mobile phones for collecting information [78].

2007. DietSense is one of the first widespread applications for health care monitoring [79]. It aims to exploit mobile phones and their sensors for image browsing, processing, and clustering, focusing mainly on nutrition.

2008. CenceMe is the first application to combine mobile phone sensing with social networks [80]. Gaonkar et al. proposed Micro-blog, the first system to share and query content through mobile devices, provide various sensor data, and enable multimedia blogs [81].

2010. The first survey on mobile phone sensing was proposed by Lane et al. [27]. It presents a comprehensive classification of sensors and applications, discussing scalability concerns from individual to community sensing and data aggregation.

2011. The work presented by Ganti et al. is the first relevant survey that specifically unveils MCS as a novel and promising paradigm, highlighting the crowd potential to monitor phenomena in urban environments [26]. Christin et al. discussed privacy concerns related to participatory sensing systems [35].

2013. Fostering citizens' participation is the focus of one of the first MCS platforms presented in [82]. Khan et al. provide a broad overview of existing mobile phone sensing works and the first classification on different methodologies to involve citizens [28]. Vastardis et al. discuss different architectures of mobile social networks, their characteristics, and future research directions [83]. MOSDEN had a significant impact as a collaborative sensing framework to share information between users and several distributed applications [84].

2014. ParticipAct Living Lab is the first and most relevant large-scale real experiment involving data collection for one year from smart devices of 200 students of the University of Bologna [85]. Kantarci et al. present a reputation-based mechanism to guarantee data integrity, where smart devices can enhance public safety [86]. Guo et. al define the term Mobile Crowdsensing in [29] by clarifying the necessary peculiarities of a system to be defined as an MCS system. Pournajaf et al. presented threats to citizens' privacy when data collection can lead to personal information disclosure, discussing how privacy mechanisms in existing literature protect users [34]. Tanas et al. used the ns-3 network simulator to evaluate MCS systems' performance by combining its specific features (e.g., network nodes' mobility characteristics with different communication technology interfaces) [88].

2015. Guo et al. highlight the combination between machine and human intelligence to empower MCS systems and the central role of humans in the sensing process [61].

2016. Chessa et al. propose a mechanism to enhance MCS systems' performance focusing on socio-technical networking aspects [89].

2017. CrowdSenSim is the first and most popular MCS simulator. It includes independent modules dedicated to urban mobility, communication technolo-

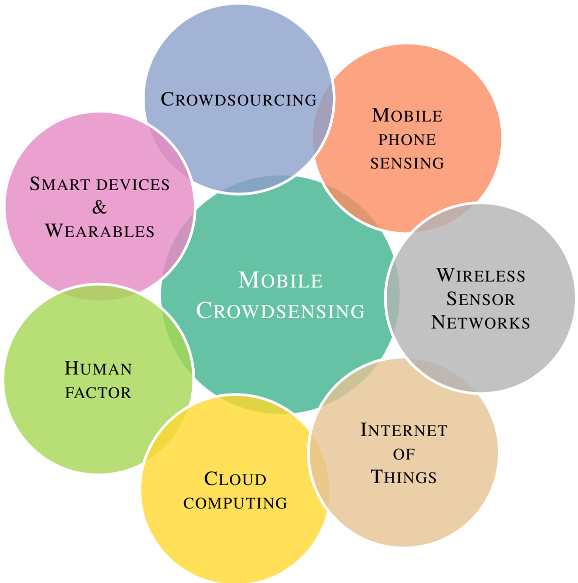


Figure 3.2: Factors contributing to the rise of MCS

gies, sensing equipment, and energy-related measurements according to the needs of different campaigns [90].

3.2 Factors Contributing to the Rise of MCS

Different factors have contributed to the rise of MCS as one of the most promising paradigms for data collection in urban environments. Fig. 3.2 illustrates the most relevant ones, which we discuss in the following paragraphs.

Mobile Phone Sensing. As already discussed, mobile phone sensing is the ancestor of MCS. Unlike MCS, individual applications are the target of phone sensing (e.g., fitness and healthcare [91]. Ubifit is a personal mobile app that stimulates users to monitor their daily activities (e.g., walking, running, or cycling) [92]. Other typical examples are diet monitoring [79], fall elderly exposure [93], [94], [95], [96], transportation mode detection [97], [98], [99], [100], speech recognition [101], and indoor navigation [103], [104], [105].

Mobile smart devices/Wearables. One of the most empowering factors to MCS has been the transition from mobile phones to smart devices. While mobile phones permitted only phone calls and text messages, smart devices embed sensing, communication, and computational capabilities that empower the user experience. Wearables represent a precious source to develop applications for improving sport and physical activities, also involving WSNs [106]. Body sensor networks represent a crucial data source for health care monitoring (e.g., nutrition and medical treatments) [107], [108]. Wearable motion tracking systems and inertial measurement units (IMUs) are representative examples for cost-effective motion tracking with a high impact in human-robot interaction [109].

Crowdsourcing. Crowdsourcing is the data collection paradigm that has driven phone sensing towards crowdsensing by including the crowd in the sensing process and leveraging massive citizens' participation. The concept of crowdsourcing has seen different definitions according to the context [110], [111], [112]. Howe coined the original term of crowdsourcing as the act of a company or an institution in outsourcing tasks formerly performed by employees to an undefined network of people in the form of an open call [77]. This definition opens to tasks accomplished collaboratively, but crowdsourcing is not necessarily only collaborative. Several single citizens joining a campaign and operating individually still match with the definition. The novelty of crowdsourcing has also been in proposing incentives mechanisms since the early days to engage many users [113], [114].

Human factor. The combination of human and machine intelligence enables including humans in the loop of sensing, computing, and communicating processes [61]. The human factor has a significant impact on MCS systems for many different reasons. First, citizens' intelligence and mobility guarantee higher coverage and better context awareness than traditional sensor networks. Second, owners maintain their smart devices by themselves without external intervention and periodically recharge them. Devising effective human-in-the-loop systems is challenging. Learning and predictive approaches represent promising solutions to exploit human behaviors within MCS systems [115], [116].

Cloud computing. Distributing computing paradigms represent a win-win solution to analyze and process data, considering the limited resources of local storage in mobile devices [16], [17]. MCS is one of the most prominent paradigms in cloud-centric IoT systems, where mobile devices offer resources through cloud platforms on a pay-as-you-go basis [37], [49], [50]. Mobile devices contribute to a significant amount of collected data that needs to be stored for analysis and processing. The cloud allows easy access to shared infrastructures and resources with a ubiquitous approach for effective information management [18], [51].

Internet of Things (IoT). The IoT paradigm presents a massive heterogeneity of devices, end systems, and link-layer technologies. Nonetheless, MCS focuses on smart city applications, which narrows down the scope of IoT applicability. In order to preserve the smart cities vision, urban IoT systems should focus on improving the quality of citizens' life while targeting sustainable development and providing added value to the community through the most recent ICT systems [25]. To this end, MCS combines human contribu-

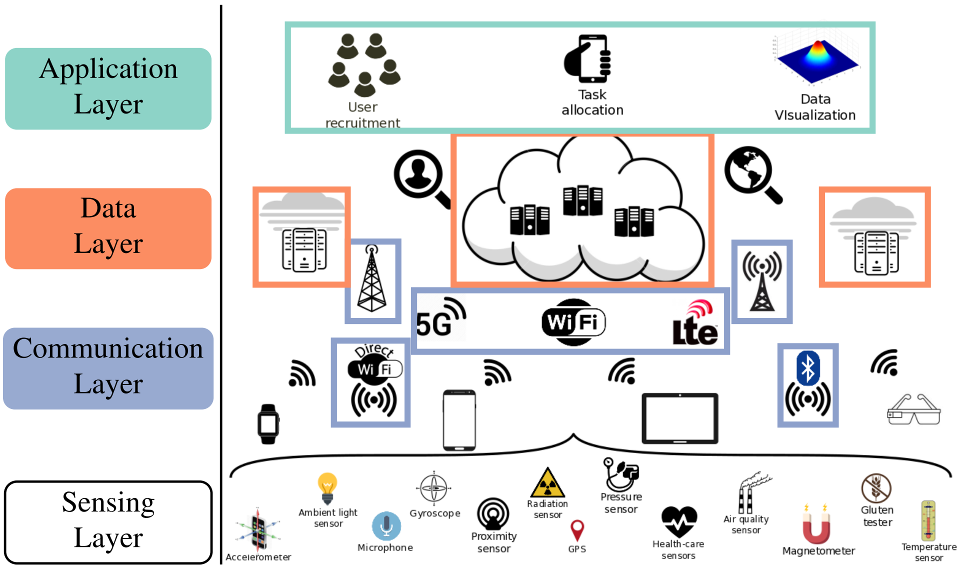


Figure 3.3: Layered architecture of MCS systems

tion and existing sensing infrastructures with no need to further investments.

Wireless Sensor Networks (WSN). WSNs are sensing infrastructures employed to monitor phenomena in urban environments. Sensing nodes, which differ according to the context, in MCS consists of citizens and their smart devices. As WSN nodes continuously increase their sensing, computational, and communication power, several applications can run over the same WSN infrastructure by exploiting virtualization techniques [117]. The most challenging issue in recent WSNs is to provide extensive scalability while maintaining high performance. The Software-Defined Networking (SDN) approach can be applied to address these issues and increase sustainability and efficiency [118]. The increase of small sensors generates a considerable amount of data that cannot be effectively processed and analyzed in WSNs due to their weak communication capability. In this context, combining WSNs with cloud computing represents a promising solution [119].

3.3 MCS as a Layered Architecture

This section introduces a four-layered architecture to illustrate the MCS landscape. A similar proposal is discussed in [62], [63], but the different rationale presented here is to follow the direction of command and control. As shown in Fig. 3.3, the highest layer is the *Application* layer, which concerns everything related to tasks and users. From top to bottom, the second layer is the *Data* layer, which characterizes storage and analytics processes on gathered information. Then, the third layer is the *Communication* layer that comprises delivery techniques and communication technologies. The lowest

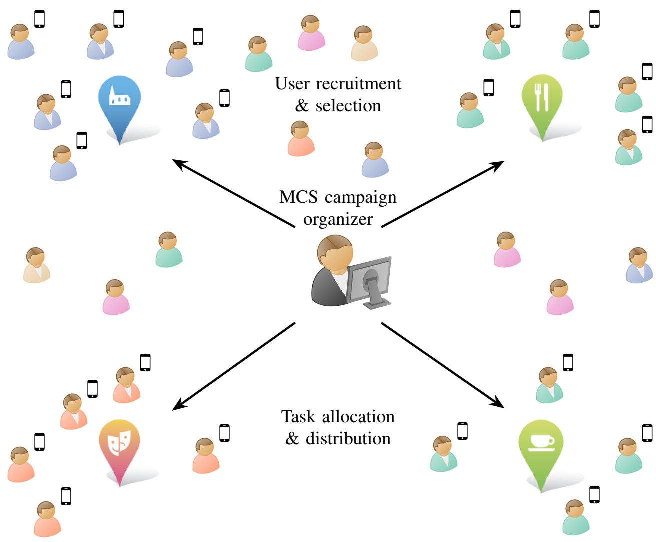


Figure 3.4: Application layer

layer is the *Sensing* layer, which includes sensing processes and modalities. While this section discusses the architecture, Chapter 4 will exploit it to present novel taxonomies for each layer.

Application layer. The application layer comprises all high-level characteristics and methodologies of MCS systems. Fig. 3.4 illustrates generic operations to organize and accomplish MCS campaigns, such as rewarding mechanisms and user recruitment strategies to maximize the number of participants or task selection and allocation mechanisms to optimize contributions and costs users sustain. Sec. 4.1 will discuss taxonomies regarding this layer.

Data layer. The data layer includes the storage, analysis, and processing of gathered information. Typically, as shown in Fig. 3.5, these operations are located in the cloud or closer to end-users by exploiting fog and edge computing. This layer includes not only raw data collected from sensors but also inferred information. Sec. 4.2 will discuss taxonomies on this layer.

Communication layer. The communication layer indicates both methodologies and technologies to report to the central collector information gathered through sensors. As shown in Fig. 3.6, smart devices usually embed different communication technologies and interfaces (e.g., WiFi, 4G/LTE, Bluetooth), and different operations can be performed by exploiting these interfaces (e.g., coding, transmitting, avoid redundant data). Sec. 4.3 will present taxonomies related to the communication layer.

Sensing layer. The sensing layer is the first and most crucial for MCS, consisting of sensors, methodologies, and processes to acquire data. As shown in Fig. 3.7, smart devices exploit embedded or connected sensors to contribute data. Typical built-in sensors have the main scope to support basic

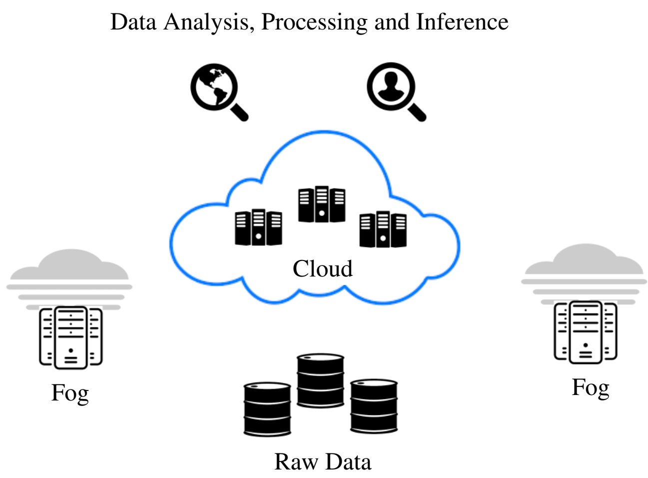


Figure 3.5: Data layer

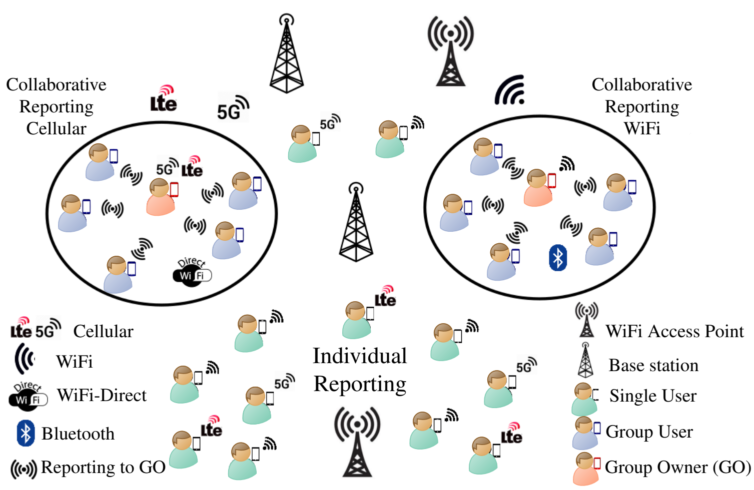


Figure 3.6: Communication layer

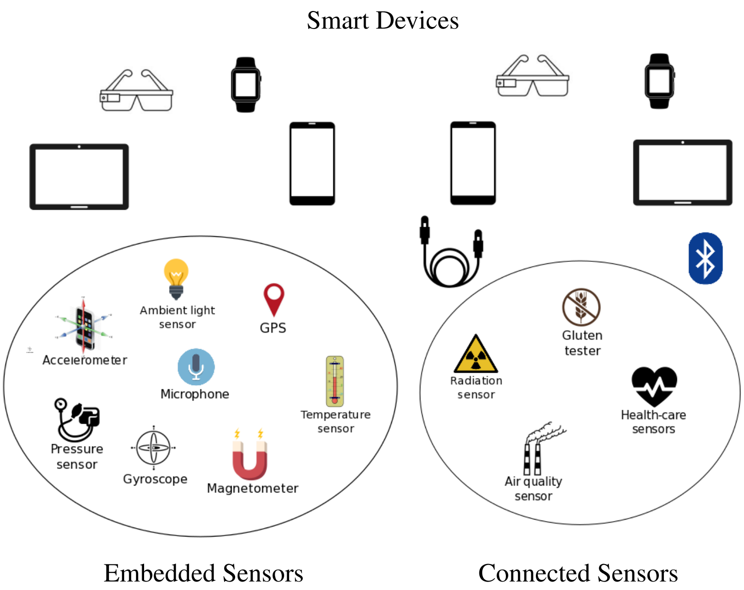


Figure 3.7: Sensing layer

device functionalities (e.g., microphone, light sensor to adjust the display brightness, or accelerometer to orientate the monitor). Many other sensors that are not necessary for primary usage are embedded because widespread between users (e.g., GPS, camera, pressure) and fundamental for acquiring data. Sec. 4.4 will propose taxonomies corresponding to this layer.

Chapter 4

A Novel Taxonomy for MCS Systems

This chapter brings the organization of the vast literature on MCS one step further by proposing novel taxonomies based on the layered architecture discussed in Section 3.3. The main aim of the taxonomy is to classify previous works considering their “technological” layer, i.e., sensing, communication, data processing, and application. While existing literature usually focuses on specific MCS aspects (e.g., incentive mechanisms, task assignment, privacy concerns), this novel perspective gives insights into issues and challenges for each layer and its interconnections. The main focus is to establish consensus on many concepts employed with different meanings and categorize the vast amount of literature by following the proposed criteria. The proposed taxonomies subdivide each architecture layer into two categories, as Fig. 4.1 illustrates. The application layer includes *task* and *user* categories, discussed in Sec. 4.1. The data layer comprises *management* and *processing* groups, explained in Sec. 4.2. The communication layer is divided in *technologies* and *reporting*, presented in Sec. 4.3. Lastly, the sensing layer taxonomies are divided in *sampling process* and *elements* proposed in Sec. 4.4. While this chapter discusses the novel taxonomy in detail, it omits the detailed classification of existing works for space reasons. The interested reader can refer to [120].

4.1 Taxonomies on Application Layer

This Section analyzes the taxonomies of the application layer, mainly composed of *task* and *user* categories. These are discussed with two corresponding taxonomies, as shown in Fig. 4.2.

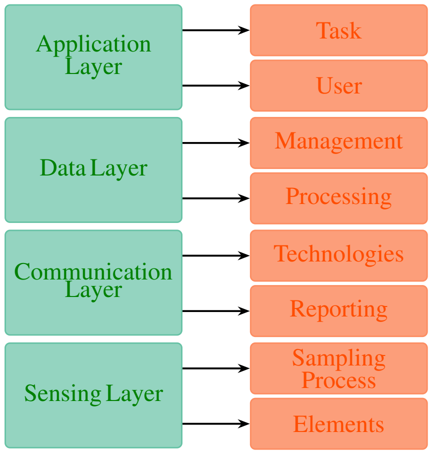


Figure 4.1: General taxonomies on MCS four-layered architecture

4.1.1 Task

The upper part of Fig. 4.2 illustrates task-related taxonomies, which comprise *scheduling*, *assignment*, and *execution* categories.

Scheduling. It illustrates task allocation to participants, which depends on the type of contribution. A *pro-active* behavior requires users to actively contribute data without a pre-assigned task. Contributors can autonomously decide when and where to accomplish the task. Typical examples include taking pictures for emergency management or public safety (e.g., floodings, earthquakes, car accidents), and social networks applications [80], [121]. Participants follow a *reactive* approach when they receive precise tasks and accomplish them accordingly. This policy requires that tasks are preliminarily decided. Typical use cases include monitoring phenomena such as noise and air pollution [69].

Assignment. It describes the process of assigning tasks to participants [34]. When the campaign presents an entity that dispatches tasks among users, the approach is called *central authority* assignment. Common use cases include environmental monitoring, such as measuring pollution [122] or nuclear radiation [123]. Task distribution can also be *decentralized* when users have the authority to forward tasks to other participants. Typical examples are mobile social networks where users are interested in the same events or activities, such as sharing public transport delays [124] or comparing real-time prices of goods [125].

Execution. It refers to the methodology followed in executing tasks. The category *single task* indicates that users need to accomplish only a type of task, such as recording a video or measuring the temperature, or detecting noise level in decibel [126]. On the other side, *multi-tasking* refers to campaigns that require users to perform different types of tasks, typically assigned by

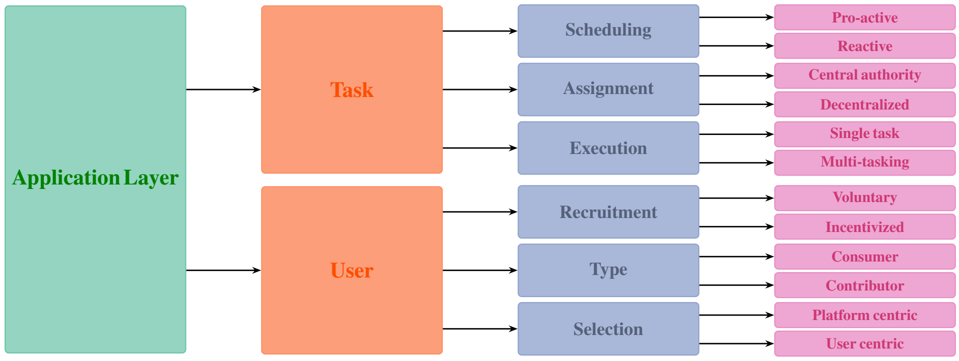


Figure 4.2: Taxonomies on application layer

a central authority. Some representative examples consist of monitoring temperature and air quality or taking a picture associated with a measured noise level.

4.1.2 User

An MCS campaign needs the large participation of users to be successful. To this end, it is crucial to develop recruitment strategies, select contributors, and differentiate between participants. The lower part of Fig. 4.2 describes user-related taxonomies, which include *recruitment*, *selection*, and *type*. Typically, the term *user recruitment* assumes two different meanings in literature. More popularly, it refers to citizens joining a campaign and being contributors. Less commonly, it is related to select users for accomplishing a specific task between all possible participants. The taxonomy proposes to capture this difference by introducing two different categories, namely *user recruitment* and *user selection*. Fig. 4.3 clarifies how the whole user-related process is intended in the taxonomy. Recruitment refers only to the approach of recruiting participants, who can join through incentives or voluntarily. Then, the campaign organizer allocates and distributes tasks to participants. Finally, some users are selected to gather and deliver data to the central collector.

Recruitment. It is *voluntary* when citizens spontaneously join an MCS campaign for personal willingness and interests without receiving incentives from organizers. Representative examples where citizens tend to volunteer are healthcare applications for mapping restaurants with dietary requirements [8], noise monitoring [126], and air pollution [69]. To increase participation, users can also be *incentivized* [127], [128], [129]. Literature is vast in proposing many different strategies to empower participation [30], [34], [130], which can be mainly classified into three groups [31]: money, service, and entertainment. A monetary incentive consists of rewarding participants with money in proportion to the amount and quality of contributed data. Service incentives stimulate users by offering services provided by the

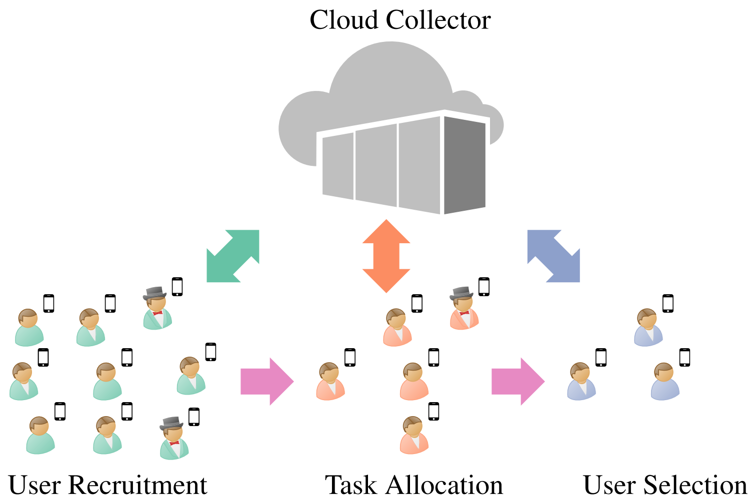


Figure 4.3: User recruitment, task allocation, and user selection process

system and needed from users (e.g., receiving the same type of collected information on a broader scale). The entertainment category consists of assigning tasks as games to accomplish with some competition between users (e.g., through a leader board or levels to be completed).

Selection. It refers to selecting data contribution between all participants and depends on better matching campaign requirements (e.g., spatial or temporal coverage) according to different approaches (e.g., user density in a particular area of interest or their availability). The selection is *user centric* when gathering and delivering data depends only on participant willingness to sense and report to the central collector, which does not send any requests. The *platform centric* approach, instead, indicates that the central authority decides who collects data according to several parameters. This decision is taken following the quality of information, which can be related to different indicators, such as the campaign coverage, the total amount of required data, or the contribution density in a particular region of interest.

Type. It divides users into two different roles within an MCS system. A participant is a *contributor* when collects and delivers data to the central collector with no interest in receiving information. Typically this category is guided by rewards (e.g., money) or willingness to help in accomplishing the campaign target (e.g., helping the scientific community to map noise pollution in cities [131]). The category *consumer* includes users that join an MCS campaign for a personal interest. For instance, celiacs have an interest in receiving data about dietary advice and restaurants [8]. Several MCS campaigns have participants assuming both roles, for instance, when taking pictures to share fuel prices [132].

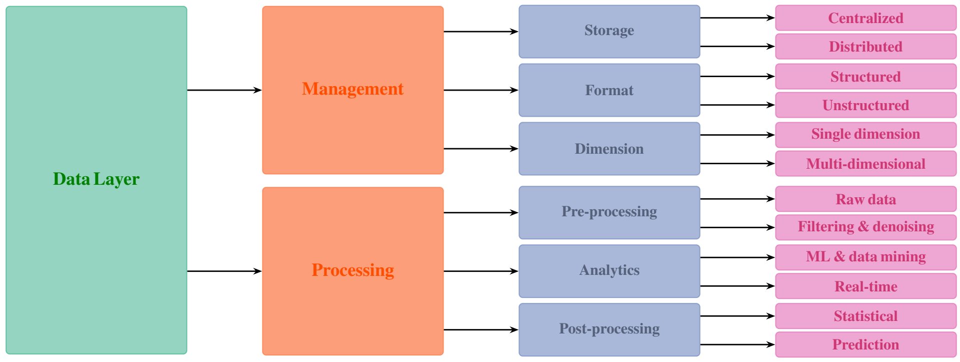


Figure 4.4: Taxonomies on data layer

4.2 Taxonomies on Data Layer

This section presents the taxonomies on data layer. As shown in Fig. 4.4, it is mainly divided in *management* and *processing* categories, which include other subcategories.

4.2.1 Management

Data management comprises *storage*, *format*, and *dimension* categories. The upper part of Fig. 4.4 illustrates subcategories of each group.

Storage. It concerns how to keep and maintain gathered data, and the location used for storage. The *centralized* manner is based on storing data in a single location, typically a database available in the cloud. It is usually employed when consistent processing or analytics is required, such as for emergency situation management [133] and urban monitoring [134]. A *distributed* approach is usually exploited for delay-tolerant applications, such as air quality [69] monitoring and urban planning [135]. Recent distributed computing paradigms, such as fog and multi-access edge computing, empower this approach by making available resources closer to end users [136].

Format. It divides data according to its structure. *Structured* data is organized and clearly defined to be stored, processed, and analyzed. It is usually self-explanatory, such as a specific identifier, location coordinates, and a measured value (e.g., noise or air pollution). *Unstructured* data does not present a specific identifier captured by search functions. Typical examples are video, audio, images, and all other files requiring complex analysis.

Dimension. Dimension indicates the number of collected data types. Data is *single dimension* when users gather only one type of data, such as when using only one sensor (e.g., temperature or pressure). Data is *multi-dimensional* when participants contribute more types by using multiple sensors (e.g., when uploading files in mobile social networks).

4.2.2 Processing

After data storage, processing is the most crucial step in MCS systems. Taxonomies include *pre-processing*, *analytics*, and *post-processing*, as shown in Fig. 4.4.

Pre-processing. It includes all the operations on contributed data before analytics. When no operations are conducted, *raw data* is stored. It allows applying inferring techniques at later stages by working on original unprocessed data. Data can be manipulated through several strategies, but the most common are *filtering and denoising*. They consist of refining information by removing redundant and irrelevant data that also permits to reduce the amount of data to be stored.

Analytics. It is related to the processes required to extract and infer meaningful information from contributed data. The first category is *ML and data mining*, which refers to all non-real-time techniques employed to identify patterns, infer information, or predict future trends. Representative examples undergoing research efforts are indoor navigation systems and urban planning. Analytics is *real-time* when it examines data as soon as delivered to the central collector. This approach requires high computational resources to be effective. Typical examples are traffic monitoring, unmanned vehicles, and emergency management.

Post-processing. It includes all approaches employed after analytics. A *statistical* analysis aims to infer information from quantitative inputs and study the correlation between different factors. The category *prediction* comprises all techniques that focus on determining future outcomes given new inputs. A typical example is to predict arrival time, estimating the traffic at a particular time of day and day of the week.

4.3 Taxonomies on Communication Layer

This Section proposes the taxonomies on the communication layer by analyzing data reporting from mobile devices and their application domains according to different sensing campaigns. Fig. 4.5 illustrates the two main categories of *technologies* and *reporting* and their subcategories.

4.3.1 Technologies

This taxonomy include *infrastructured* and *infrastructure-less*. The upper part of Fig. 4.5 shows also their subcategories.

Infrastructured. This category includes all technologies that exploit an infrastructure to establish a connection and deliver data, such as base stations and access points. It comprises *cellular* data communications and *WLAN*

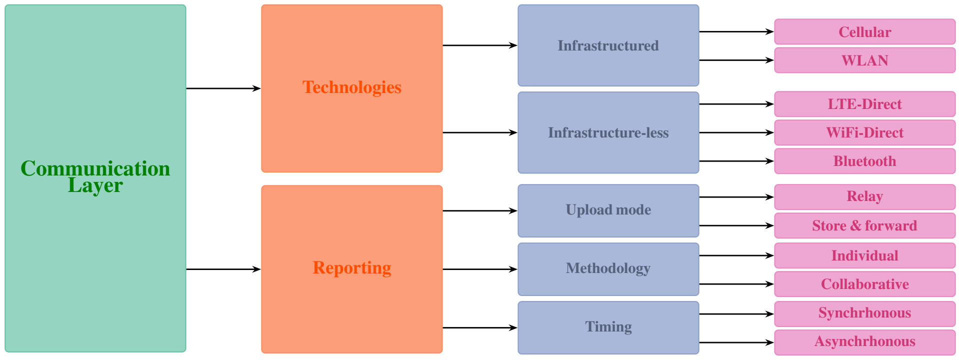


Figure 4.5: Taxonomies on communication layer

interfaces. A sensing campaign typically employs *cellular* connectivity when data collection presents latency bounds that cannot be guaranteed by WiFi connectivity. For instance, real-time monitoring for parking availability cannot employ WiFi [45]. Currently, 4G and LTE systems provide latencies and data rates required from MCS applications, but 5G technology is bringing a relevant contribution to MCS systems by providing network function virtualization and high rates. When a campaign design does not present delay constraints, *WLAN* interfaces allow saving costs without the need for subscription fees. This approach is mostly used when organizers do not ask for specific communication technologies, such as mapping air and noise pollution.

Infrastructure-less. It includes device-to-device (D2D) communications, which require proximity between devices to exchange data but do not need any access point. One of the most employed D2D standards is *WiFi Direct*, which perfectly fits MCS systems characterized by a group owner and group members. A system where participants sense and report data after electing a group owner is presented in [137]. *LTE-Direct* is another emerging paradigm, which presents low energy consumption while discovering devices rapidly in proximity. Another typical solution for MCS systems is *Bluetooth* [138]. Specifically, Bluetooth Low Energy (BLE) is a low power version.

4.3.2 Reporting

The scope of reporting is to highlight how smart devices deliver gathered information to the data collector. It is not related with communication technologies and comprises *upload mode*, *methodology*, and *timing*. Their subcategories are shown in the lower part of Fig. 4.5.

Upload mode. It considers if data reporting is in real-time or delayed according to policies on delay tolerance. Upload is *relay* when reporting happens just after sensing in a real-time manner. Whenever real-time delivery is impossible, samples are discarded to avoid energy waste [139]. Emergency

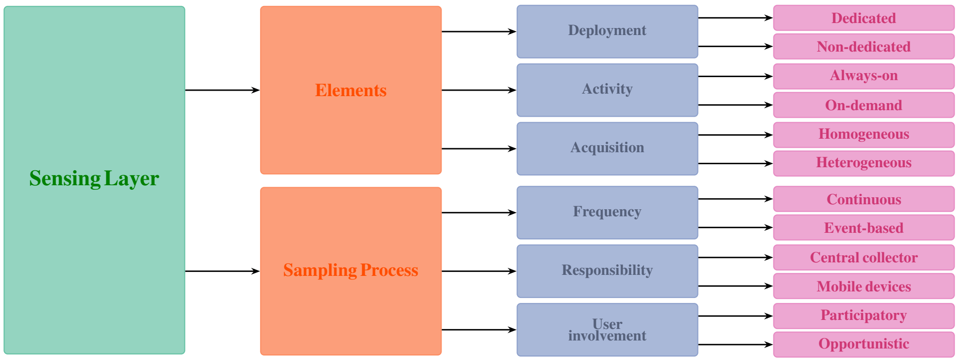


Figure 4.6: Taxonomies on sensing layer

management and traffic monitoring applications (e.g., Waze [140]) are representative examples of this class. When applications permit delay-tolerant data delivery, reporting is *store and forward*. It usually requires to label data with timestamps [141] and includes applications such as urban mobility [124] and gluten sensor to share healthy food in restaurants [8].

Methodology. It discusses if smart devices accomplish sensing tasks individually or acting as peers. The execution methodology is *individual* when participants execute tasks with no interactions with other users. A *collaborative* methodology is when participants exchange information and cooperate as peers to accomplish tasks. In this approach, usually, groups are created by exploiting short-range communication technologies, such as Bluetooth or WiFi-direct [138]. Hierarchical policies may elect a superuser who is responsible for delivering data for the whole group. To this end, different incentives can be assigned within the same group.

Timing. It includes policies that require users to collect data at the same time or not. Execution timing is *synchronous* when sensing campaigns aim to compare phenomena simultaneously, and participants must start and finish sensing in a specific time window. Typically, users can communicate between themselves to better synchronize the sensing phase. Representative examples consist in comparing real-time prices of goods [125], and traffic conditions [142]. Timing is *asynchronous* when sensing processes of participants do not necessarily need to be within the same time interval. Some examples are air pollution [10] and noise [131] mapping.

4.4 Taxonomies on Sensing Layer

This Section analyzes the sensing layer and its taxonomies, which include *elements* and *sampling process*. Fig. 4.6 illustrates sensing layer and all its subcategories. Note that the scope of these taxonomies is not to consider technical aspects of sensors as they are already very well investigated in the literature on smartphones [70], [74] and mobile phone sensing [27], [28].

4.4.1 Elements

Sensing elements comprise *deployment*, *activity*, and *acquisition*. The upper part of Fig. 4.6 shows all subcategories related to elements.

Deployment. It differentiates between *dedicated* and *non-dedicated* sensors [143]. Smart devices embed most sensors for their basic operations (e.g., the light sensor to adjust screen brightness orthe microphone for phone calls) or added functionalities (e.g., GPS for navigation). These sensors are *non-dedicated* because they cover multiple purposes. Some applications can require additional *dedicated* sensors for specific purposes. These sensors are standalone and typically not embedded but connected to smart devices via cable or wireless technologies such as Near Field Communications (NFC) or Bluetooth. Representative examples are the gluten sensor to detect food allergies [8], and other sensors for environmental monitoring, such as dust for air quality [144] or nuclear radiation [123].

Activity. It includes *always-on* and *on-demand* sensors. Smart devices embed several basic *always-on* sensors, which cannot be switched off and furnish essential functionalities to operate. Typical examples are accelerometer and gyroscope for monitoring rotation or light sensor to adjust the screen brightness. They consume a small amount of energy and can be used for many different applications, like activity recognition [145], [146], [147]. Other sensors are *on-demand* because they can be switched on and off according to many different policies (e.g., manually or automatically through context awareness). Usually, they consume more energy than the basic ones and, in normal conditions, are disabled. Camera for taking pictures, microphone to detect noise, and GPS for navigation are typical examples [132].

Acquisition. It includes *homogeneous* or *heterogeneous* data, according to the scope of the sensing campaign and required sensors. Data acquisition is *homogeneous* if it requires only one data type, such as detecting air quality [122] or noise level in dB [131]. When a sensing campaign aims to collect different types of data, the acquisition is *heterogeneous*, and typically it involves multiple sensors (e.g., monitoring traffic condition [140].

4.4.2 Sampling Process

Sampling process analyzes decision-making policies for sensing and includes *frequency*, *responsibility*, and *user involvement*. Their subcategories are shown in the lower part of Fig. 4.6.

Frequency. It investigates how often sampling should take place. Sensing is *continuous* when sampling is executed regularly, and tasks are accomplished independently by context. This process continues until the sensing organizer or the device owner stops it and can be very energy consuming according to the sampling rate. Environmental monitoring is a typical example of continuous sensing, such as air quality or noise detection [69]. When some events or a specific context trigger sensing, the sampling frequency is *event-based*. This is usually energy-consuming, and representative examples are taking pictures in emergency management, healthcare applications [8], or activity recognition [124].

Responsibility. It examines which entity makes sampling decisions. Responsibility is given to *mobile devices* when independently from a central authority, they can sense and deliver data according to different possible policies. This approach can include both a manual intervention from a user or an application running in the background for context-awareness. Sharing real-time costs is a typical example [148]. In the centralized approach, a *central collector* is responsible for making sampling decisions and communicate them to participants. It is beneficial to reach a certain amount of data or space coverage because the central collector knows already received data. Besides, it allows saving a consistent amount of energy.

User involvement. It is a very generic concept in MCS literature, often ambiguous due to the context [78]. This taxonomy aims to clarify the concept of user involvement and provide a unique classification of its categories. The term user involvement is used to specify if the sensing process needs or not an active user intervention, and it is classified as *opportunistic* and *participatory*. When a direct user action is required to sense data, the approach is *participatory*. Participants are responsible for accomplishing tasks and meeting the campaign requirements by actively deciding where and when to sense data through their devices. They can accept or decline tasks before manually performing any operations [29]. Typically, this approach presents a high data quality due to human intelligence. Representative examples are taking pictures for different applications [125], and record audio signals [78], [126]. A direct user intervention often leads to lower user willingness to contribute, and an application should run in the background to gather data. Direct user involvement is not required in the *opportunistic* approach. In this case, users have only to join a campaign, and an application running on background and communicating directly with the MCS platform will be responsible for the whole fully automated sensing process. Unlike the participatory approach, the platform can dynamically change tasks, and smart devices are responsible for context awareness. This approach requires energy-saving policies to preserve devices' duration [150]. A typical example is monitoring road conditions [149].

Chapter 5

Improving CrowdSenSim with Novel Realistic Features

This chapter presents the novel developed features over the simulation environment called CrowdSenSim. First, it overviews existing simulators and provides motivations to develop CrowdSenSim and its additional features. Then, it presents the current structure of the simulator, discussing each module. Finally, it explains novel features and discusses them in detail.

5.1 Background and Motivation

This Section provides background on existing MCS simulators and puts the basis for additional novel features in CrowdSenSim. An MCS campaign needs large data contributions from citizens to be successfully accomplished [151], and it looks unfeasible to develop testbeds for large-scale urban environments. Simulators represent a solution, but they need realistic settings, such as user mobility, applications, sensing, and communication equipment. Before CrowdSenSim, previous simulation tools used with MCS purposes focused on communications or mobility in open spaces, but it was not enough to evaluate realistic MCS campaigns [152]. Network Simulator 3 (NS-3) was used considering mobile users as network nodes [88]. While this approach provides very accurate simulation on communication aspects, it cannot scale to thousands of citizens and campaign duration. CupCarbon is a wireless sensor network (WSN) simulator that generates discrete events in the context of smart cities and IoT environments [154]. It provides the possibility to simulate WSN on real street networks through OpenStreetMap but does not scale to realistic numbers of citizens. A simulator for MCS activities in a city parking scenario is presented in [155]. The authors propose to extend the simulator in more generic scenarios, but the current implementation does not consider data collected from devices' sensors.

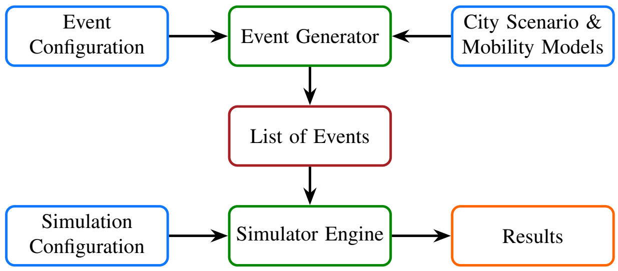


Figure 5.1: Simulator new modular architecture

When CrowdSenSim was released, it covered many of the pitfalls presented by previous simulators [90]. It also provided support to smart city applications, such as solutions for smart lighting [153]. Based on independent modules (e.g., communication, sensing, mobility, MCS inputs), it enabled researchers to investigate different aspects, such as energy consumption, pedestrian movements, and data contribution. Some years have passed since the first release of CrowdSenSim, but it has continuously been improved. This Chapter illustrates the most significant novelties and their impact on simulating MCS campaigns, such as real-world street networks, more realistic pedestrian spatial patterns and mobility models, and real measurements on devices' energy consumptions.

5.2 The CrowdSenSim Structure

This section illustrates the CrowdSenSim architecture, describing the novel features and their impact on MCS.

5.2.1 General Architecture

Figure 5.1 exhibits the new CrowdSenSim modular architecture. Most impactful novelties comprise real energy measurements (will be discussed in details in Chap. 6), real-world street networks, and realistic pedestrian mobility models. CrowdSenSim simulates a certain number of pedestrians (in the order of thousands) that walk on the chosen real-world street network. While walking, users contribute data sensed through devices' sensors and transmitted via cellular or WiFi interfaces, according to campaign specifications. The event generator creates events, such as citizens' arrivals in a defined street network node at a specific moment of the simulation runtime. To create these events, the simulator uses pedestrian mobility models, real urban layouts, positions of antennas, and MCS inputs following campaign specifications defined in a configuration file. After the list of events is created, the engine uses it to simulate users' behaviors according to different events.

5.2.2 City Layout

This module provides real city street networks given by coordinates (latitude and longitude), where simulated pedestrians move and contribute data. The "old" simulator needed a *.txt* file with all the coordinates, and it represented a bottleneck to expand the simulator to worldwide users and researchers. This feature represents a very impactful novelty, as now the simulator automatically downloads the required coordinates from OpenStreetMap (OSM) by using OSMnx, an open-source Python package [156]. Also, CrowdSenSim exploits the AOP algorithm (discussed in 5.3) to provide higher precision to the street network and giving the possibility to decide the granularity of nodes where MCS participants walk.

5.2.3 User Mobility and Event Generation

This module generates spatial and temporal patterns of participants. Pedestrian trajectories have random starting and ending nodes regulated by users' walking periods over the chosen street network. Different realistic mobility models can regulate the walking patterns according to campaign requirements. Mobility can be based on real traces, or random walking, or depending on social interactions and citizens' behaviors that differ city by city, such as following Google Popular Times[1]. Pedestrians have specific walking periods (e.g., 30 mins), and their trajectories are a list of discrete events. Pedestrians jump from one node to another one, following the network topology with a fixed speed, usually distributed between [1 - 1.5]m/s.

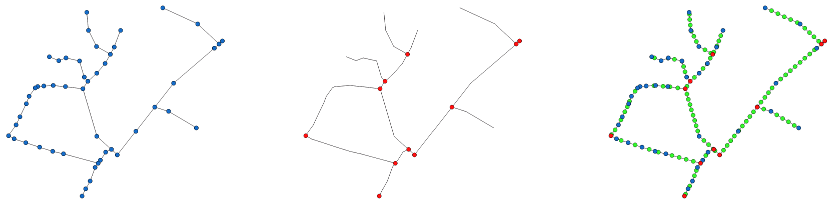
5.2.4 Simulator Engine

The engine is written in C++. It requires a list of events as input and generates users' actions accordingly. Events trigger sensing and communication activities to contribute data to the central collector. It also generates movements on the street network according to the mobility models.

5.3 Realistic Pedestrian Mobility

This section illustrates in detail CrowdSenSim novel features. First, it discusses how real street networks are generated for any selected city. Then, it discusses various realistic pedestrian mobility models, assessing the performance. Note that real energy-related experiments and measures, which is another significant innovation, will be considered in Chapter 6.

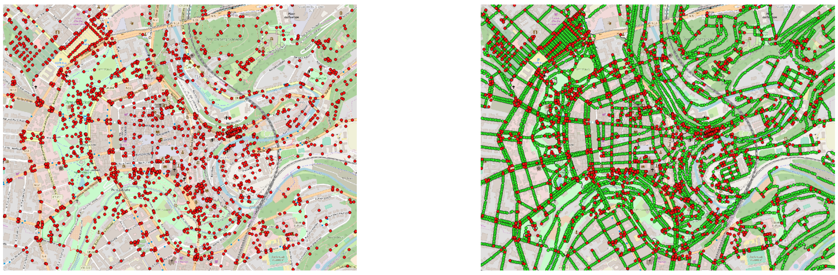
[1] https://support.google.com/business/answer/2721884



(a) Brancolino - Precision of OSM

(b) Brancolino - Precision of OSMnx

(c) Brancolino - Precision of AOP



(d) Luxembourg - Precision of OSMnx

(e) Luxembourg - Precision of AOP

Figure 5.2: Street network granularity for different cities.

5.3.1 Real Worldwide City Street Networks

One of the most impactful features in the history of CrowdSenSim as a software product adopted worldwide is the possibility to perform simulations in any city by automatically downloading street networks without the need to input coordinates. This feature consists of two phases. First, incorporating a Python library called OSMnx developed by Geoff Boeing that has provided a tool to download and simplify OSM coordinates [156]. Then, developing the AOP algorithm to have uniform street networks with custom granularity.

The AOP algorithm. OSM makes available a set of coordinates to generate a graph of any street networks, seen as a set of nodes and links. OSMnx, as previously discussed, enables to download this set of coordinates and simplifies it. The main shortcoming for MCS purposes is that the resulting street network does not have a uniform granularity of nodes. If used to generate pedestrians that walk on the graph, it would unrealistically result in citizens' different densities in different areas. The AOP algorithm was developed to overcome this issue. AOP aims to increase the graph's precision by reaching a fixed uniform target distance among any two adjacent nodes. The target distance can be as low as 1 m. Two different algorithms are proposed to interpolate the location of the nodes appended to reach the target distance. V-AOP reflects the precise range between two nodes in the OSM graph. L-AOP follows a linear approximation of the distance, decreasing the required computation time. A curious reader can find the implementation details of V-AOP and L-AOP in [157].

Fig. 5.2 presents the graphs of street networks from two cities. Brancolino is a village in northern Italy, while Luxembourg city is a mid-size European capital. The figure exhibits the different precisions of graphs obtained with OSM (Fig. 5.2(a)), OSMnx (Fig. 5.2(b) and Fig. 5.2(d)) and AOP (Fig. 5.2(c) and Fig. 5.2(e)). AOP is generic enough to be applied to any case in which the precision of OSM is not sufficient.

5.3.2 Pedestrian Mobility Models

Simulations exploit two different pedestrian arrival models. The first approach is called U-MOB and arrivals are uniformly distributed over simulation runtime. The second approach is called D-MOB and arrivals depend on traces taken from the popular MCS dataset ParticipAct [159].

The U-MOB Mobility Model. U-MOB generates arrivals in a uniformly distributed fashion over simulation runtime. Usually, the period is one hour long to enable comparisons with D-MOB because ParticipAct traces have that granularity. Pedestrians who have a walking period at different hours count as one arrival.

The D-MOB Mobility Model. D-MOB is based on the ParticipAct dataset, collected from an MCS campaign of 170 students of the University of Bologna in the Emilia Romagna region (Italy) [159]. As the ParticipAct dataset is not publicly available, D-MOB is extracted by a profile of the average number of arrivals over a week. In particular, fixed the simulation period, the time is divided in hours and it is estimated the minimum number of users to allocate so that arrivals follow the ParticipAct profile.

The simulator implements two approaches for D-MOB. The Contact-Only-Distribution (COD) approach allocates hourly pedestrians until reaching the average number of contacts provided by the dataset. Once the preliminary phase is finished, it assigns the remaining users proportionally to the hourly number of contacts. The Contact and User-Distribution (CUD) method is hybrid. The first phase behaves the same as COD, but then it allocates the remaining number of users proportionally to the number of hours of the simulation time.

Fig. 5.3 shows an example of COD and CUD arrivals over a 36 hours period with $50,000$ users.

5.3.3 Performance Evaluation

Performance evaluation discusses the scalability of the AOP procedure, human mobility metrics, and the accuracy of D-MOB arrivals methods.

Scalability of AOP. The street network topology and the original accuracy of OSM influence the scalability of the AOP approach. Experiments have a

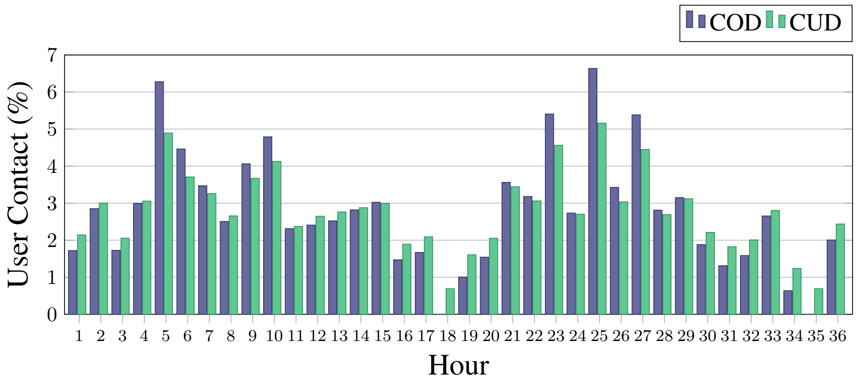


Figure 5.3: Dataset-based arrivals

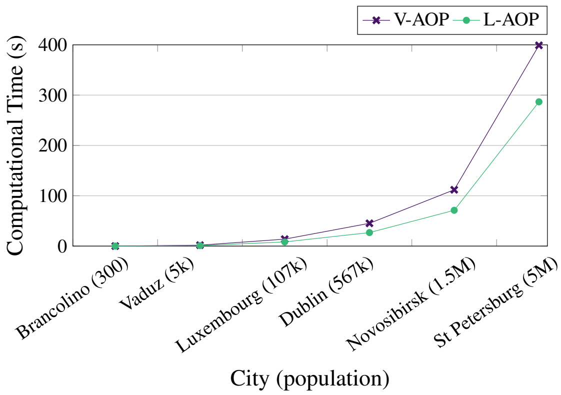
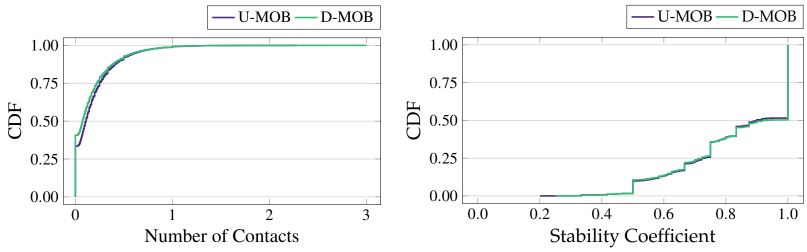


Figure 5.4: Computational time and relative accuracy of V-AOP and L-AOP

target distance of 3 m. Simulations are performed on a Linux Ubuntu 16.04 laptop, with Intel ®Core TM i7-4710HQ 2.50 GHz x8 CPU, and a 7.7 GiB system memory. AOP requires more computational time for bigger cities. Considering two cities of comparable sizes, the one with lower original accuracy will take less time to be processed. We selected locations with increasing area and latitude (for the approximation of the Earth axis), utilizing V-AOP and L-AOP for evaluation purposes. The time needed for computation is proportional to the number of original nodes and the average edge length. To give some examples, AOP requires less execution time for municipalities like Edinburgh (171, 271 initial edges, average length 17.54 m) or Genoa (166, 479 initial edges, average length 15.31 m) than Novosibirsk (133, 556 initial edges, average length 42.69 m). The time difference between the two approaches is negligible for small and medium cities, but it consistently increases with the city's size and the original number of edges. L-AOP requires less time than V-AOP (e.g., in St Petersburg, the difference of computational times is approximately 100 s).

Human mobility metrics. The performance of U-MOB and D-MOB is evalu-



(a) Per-user average contacts (UAC) **(b)** Stability coefficient (SC)

Figure 5.5: Analysis of contact distribution and stability of contacts

ated by introducing two different metrics. Simulations are performed with $50,000$ participants that walk for 2 days. Timeslots are 1 minute long. Two pedestrians have contact if they are at a distance R closer than 50 m. The first metric is the per-User Average Contacts (UAC):

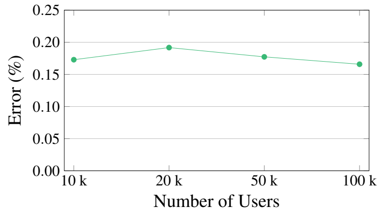
$$\mathrm{UAC}_i = \frac{1}{T_i} \cdot \sum_{j=1}^{T_i} n_{j,i}, \tag{5.1}$$

where T_i is the period (number of timeslots) that the pedestrian i walk across the street network, and $n_{j,i}$ is the number of contacts in timeslot j. The stability coefficient (SC) is the second metric [161]:

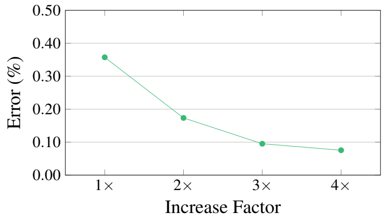
$$\mathrm{SC}_i = \frac{1}{T_i} \cdot \sum_{j=1}^{T_i} \frac{|n_{j+1,i} \setminus n_{j,i}| + |n_{j,i} \setminus n_{j+1,i}|}{|n_{j,i}| + |n_{j+1,i}|}, \tag{5.2}$$

where $n_{j,i}$ is the set of neighbors of participant i in timeslot j, $|n_{j+1,i} \setminus n_{j,i}|$ are the contacts that i loses between timeslots j and $j+1$ and $|n_{j,i} \setminus n_{j+1,i}|$ are the contacts that i acquires between timeslots j and $j+1$. SC represents how often a citizen changes contacts. Fig. 5.5 exploits the cumulative distribution function (CDF) to compare U-MOB and D-MOB for both metrics. Fig. 5.5(a) exhibits that for the UAC metricU-MOB based on uniform arrivals almost overlaps D-MOB obtained with ParticipAct. Fig. 5.5(b) illustrates that 50% of participants stay in contact for two adjacent timeslots. Approximately for 75 % of pedestrians, almost 30 % of neighbors is stable with values of SC lower than 0.7.

Accuracy of the D-MOB arrivals methods. Two different experiments in Luxembourg city are conducted to assess the accuracy of D-MOB. The first experiment presents four scenarios with simulation time set to 12 hours and an increasing number of participants. Fig. 5.6(a) illustrates the relative error comparing COD and CUD for D-MOB. As the error is below the 0.25% the accuracy is high and almost constant for different numbers of participants. The second experiment focuses on assessing the accuracy by jointly increasing the simulation time and participants' number. Fig. 5.6(b) exhibits the



(a) Simulation period of 48 hours



(b) Joint increase of users and simulation period

Figure 5.6: Accuracy of mobility models

obtained results for various increasing factors. Note that an increasing factor of 1× corresponds to a simulation period of 24 hours and $10,000$ pedestrians. Results are accurate, with the greatest error at almost 0.4 %. As expected, the lowest number of participants corresponds to the biggest error.

Chapter 6

Analyzing Energy Efficiency in MCS Data Collection Frameworks

As introduced in previous Chapters, devising energy-efficient data collection frameworks (DCF) is essential to foster users' participation and make a campaign effective. Ensuring devices' duration while sensing and delivering data is one of the most challenging issues to accomplish MCS campaigns. Despite literature to build large-scale MCS applications is vast, the design of efficient DCFs still requires investigation, and only a few works have tried to analyze the performance of DCFs from an experimental perspective. To fill this gap, this chapter proposes a novel experimental methodology for comparing and assessing the performance of different DCFs. It consists of three different phases. First, an Android application is developed to implement different DCFs. Second, a power monitor and Wireshark profile energy- and network-related performance. Third, the CrowdSenSim simulator is fed with the obtained traces to assess large urban environments' performance.

6.1 Background and Motivation

This Section discusses related works and illustrates three popular DCFs under study.

6.1.1 Related Works

DCFs provide mechanisms to accomplish MCS campaigns aiming to save costs (e.g., monetary rewards or energy consumption) while maximizing specific key performance indicators, such as the spatial and temporal coverage or the quality of data while saving costs [47]. In [46] the authors analyzes the energy consumption related to the amount of gathered data and present both on-line and off-line use cases to allocate tasks efficiently. Scheduling multiple MCS tasks to maximize the quality of information and minimize the energy cost is studied in [162]. Zhao et al. present an energy-efficient mechanism that allocates tasks by minimizing the time required for sensing through an NP-hard problem. CARDAP is a distributed DCF based on a fog computing platform that supports efficient data analytics [54]. Another fog computing architecture is presented in [48], where the authors propose multiple criteria to recruit users efficiently. Minimize the user arrival time and sensing duration to save energy is proposed in [163]. In [166] the authors propose a framework that minimizes energy consumption by dividing users into different groups according to their costs and selecting them for reporting. At the time of writing, no studies investigate and compare the performance of DCFs in large scale campaigns to mimic a real MCS deployment. This chapter tackles this challenge into two phases. First, by implementing and assessing the performance of DCFs. Then, by exploiting CrowdSenSim to run simulations in large-scale urban environments.

6.1.2 DCFs under Analysis

The proposed methodology compares three DCFs, which include different approaches with specific properties. These DCFs represent three categories that also include other solutions proposed in the existing literature.

DDF - Deterministic Distributed Framework. DDF is an approach tailored for opportunistic sensing campaigns to enhance energy-efficient contributions [158]. Aiming to save energy consumption for end-users' devices, this DCF exploits the central collector's utility in receiving a type of data from a specific region to decide who should contribute with a distributed approach. Smart devices periodically receive beacons from the central collector and locally decide when performing sensing operations according to different parameters, such as the device's battery level.

PDA - Probabilistic Distributed Algorithm. PDA is an algorithm for opportunistic MCS campaigns that exploits a probabilistic mechanism to minimize data redundancy and reduce mobile devices' energy consumption [168]. To this end, smart devices locally determine when sensing and delivering information. Unlike DDF, this probabilistic method considers independent events and does not include devices' historical behaviors.

PCS - Piggyback Crowdsensing. PCS focuses on regular device activity to minimize energy costs related to sensing and delivery (e.g., phone calls or applications) [169]. This approach significantly reduces the battery drain because it does not require smart devices to use communications interfaces or sensors specifically for the MCS campaign. PCS perfectly fits delay-tolerant applications because it does not consider any priority or feedback from the collector to receive data.

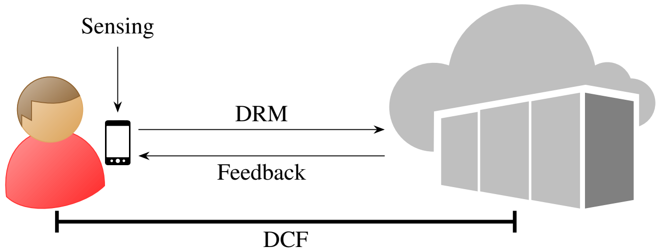


Figure 6.1: MCS data collection end-to-end: role of DCF and DRM

6.2 Profiling Energy Consumption of DCFs

This Section illustrates the experimental methodology proposed to compare and evaluate the DCFs under analysis. First, it describes the development of an Android application to implement the DCFs. Second, it presents experiments conducted in a laboratory using Wireshark for network analyses and a Power Monitor for energy measurements. These experimental results will be used in the next Section to feed CrowdSenSim for evaluating the DCFs in realistic city-wide scenarios.

6.2.1 Data Reporting Mechanisms (DRM)

A DCF comprises different elements, as shown in Figure 6.1 . The fundamental part is the data reporting mechanism (DRM), representing the approach to deliver information. Other elements can include techniques to send information to users about the data utility or task allocation strategies. In the following, the most common DRMs will be presented. Different DCMs consider different DRMs.

Continuous-DRM (CON). CON is based on delivering data as soon as gathered. It needs communication interfaces always-on and implies an energy-consuming process for participants. This mechanism is commonly used in real-time campaigns, such as DDF.

Delayed-DRM (DEL). DEL consists of reporting information when sensing is completed. It does not require sending data continuously, consuming less energy. DEL is typically employed for delay-tolerant campaigns, such as PCS.

Probabilistic-DRM (PRO). PRO is a DRM with characteristics in the middle of CON and DEL. It generates a value for each timeslot and compares it to a threshold to decide if reporting data. PDA exploits PRO.

6.2.2 The Application Architecture in a Nutshell

To implement the proposed DRMs, a custom Android application has been developed. It follows the REST guidelines, such as other popular works in

(a) Architecture of the mobile application **(b)** Power measurements setup

Figure 6.2: Experimental set-up

the MCS domain [170], [173]. To be REST-complain, the application is based on self-descriptive resources linked to each other, with a uniform interface to decouple the architecture using HTTP methods to interact with the cloud. Fig. 6.2(a) shows the client-server architecture and all its components. The public WiFi network EDUROAM provides the connection between the smart device under analysis and the central collector. The programming language used for the development is Java, and the server-side scripting language for web-development is PHP. The application is compatible with Android Oreo 8.0 (API level 26), but Marshmallow 6.0 (API level 23) is the minimum supported version. The central collector is a laptop used for storage and processing, exploiting XAMPP (v7.1.8 - 32bit) with a unique distribution Apache web server and phpMyAdmin to manage the database based on MariaDB. Each DCF impacts the implementation design of DRMs according to its characteristic features. For instance, CON and PRO store the collected information in a buffer, while DEL needs a local database, which guarantees reliability.

6.2.3 Experimental Set-up

Fig. 6.2(b) illustrates the experimental setup to obtain energy profiles of the different DRMs. A power monitor enables energy measurements, similarly to existing literature [174], [175]. The smartphone under analysis is a Wiko Sunny running Android Marshmallow version 6.0 (API Level 23), equipped with a quad-core 1.3 GHz Cortex-A7 processor, and powered by a 1 200 mA, 3.7 V battery [1]. For experimental purposes, the application exploits the accelerometer, proximity sensor, and GPS for sensing. The cloud collector is the laptop, equipped with a dual-core 2.6 GHz Intel i5-4210M, 8 GB of RAM, a 256 GB Crucial SSD as storage, and a Realtek card for WiFi 802.11 b/g/n connectivity. The hypothesis is consistent as the laptop exceeds the

[1]Available at http://it.wikomobile.com/m1330-sunny

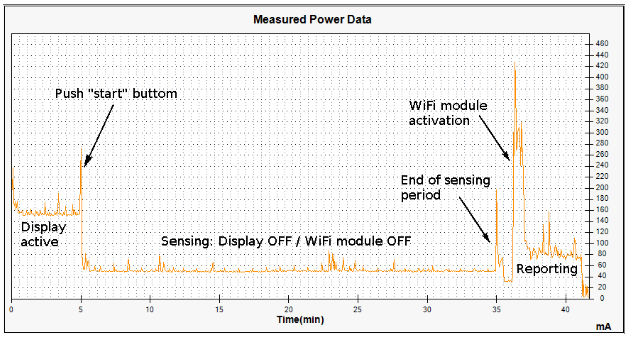
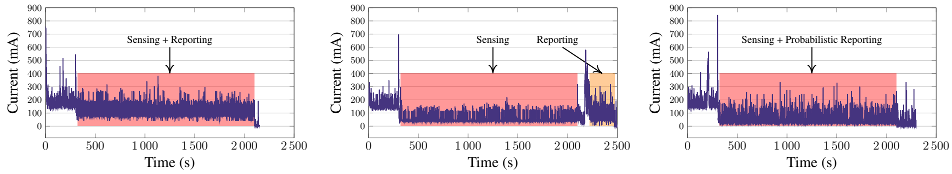


Figure 6.3: Screenshot of power monitor measurements



(a) Continuous DRM (CON) **(b)** Delayed DRM (DEL) **(c)** Probabilistic DRM (PRO)

Figure 6.4: Real energy measurements performed with Power Monitor for different data reporting mechanisms (DRMs)

smartphone performance. The power monitor hardware is by Monsoon[2]. Unlike existing works [176], [177], the power monitor directly powers the smartphone to substitute the internal battery in the equivalent circuit and directly retrieve measures.

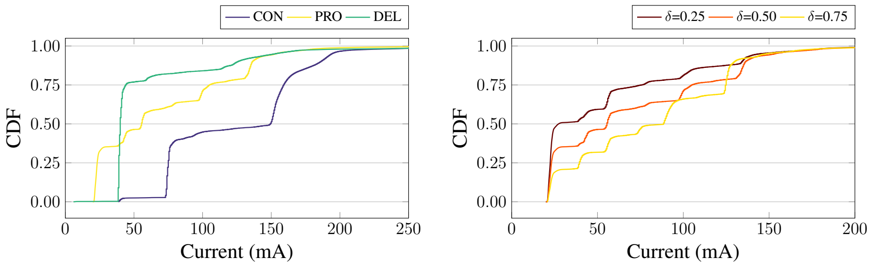
6.2.4 Experimental Results

This Section discusses results obtained from energy- and network-related experiments conducted by implementing the DRMs. Different DRMs have different times and costs for reporting. For instance, DEL transmits only in specific timeslots, while CON has always-on interfaces. For this reason, the implemented campaign considers the same sensing timeslots for each DRM (e.g., 30 minutes). As a consequence, all DRMs collect the same amount of information to deliver with different delivery times. For instance, DEL senses for 30 minutes and delivers collected data in 6 minutes.

Fig. 6.3 shows the profile of DEL as a representative example. In particular, it highlights different activities at different times (e.g., data reporting after sensing is finished).

Fig. 6.4 illustrates the instantaneous current drain of CON, DEL, and PRO when delivering information. In CON, the graph is bursty because the smartphone transmits readings as soon as taken. As shown in Fig. 6.4(a),

[2]Available at http://www.msoon.com/LabEquipment/PowerMonitor/



(a) Data Reporting Mechanisms

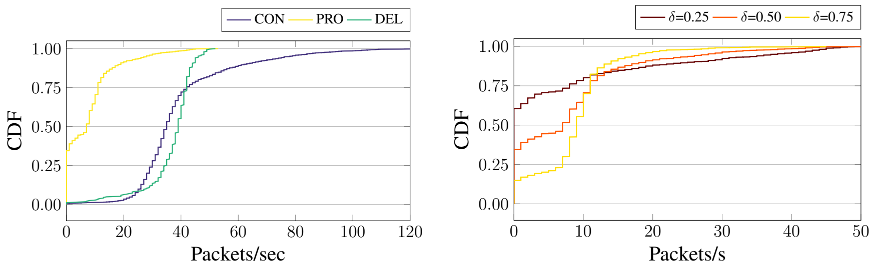
(b) Different thresholds for probabilistic reporting

Figure 6.5: CDF of energy spent for different reporting approaches

the reporting time is short but corresponds to a higher current drain to maintain the interface active. Unlike CON, the delayed approach (DEL) has a shorter delivery time because it happens after sensing is finished, allowing to maintain the interface active for a small time window. As Fig. 6.4(b) shows, the interface activation creates a peak around the 2150 second, but the current drain is on average lower than CON. PRO presents a performance in the middle of the two previous approaches, as shown in Fig. 6.4(c). Even if the network interface stays active for a more extended period than CON, it corresponds to a lower current drain on average.

Fig. 6.5 shows the CDF corresponding to the battery consumption. Specifically, Fig. 6.5(a) illustrates the various DRMs and Fig. 6.5(b) highlights the impact of the threshold δ for data tranmission in PRO. note that each timeslot has a duration of 40 seconds. The battery drain in CON is lower than 75 mA for a consistent amount of time. Unlike CON, DEL presents on average higher instantaneous peak values. While CON achieves peak values above 150 mA for 50% of the delivery period, DEL exhibits values above 40 mA. Such behavior is expected from the theoretical results on WiFi energy consumption as one of the substantial components to the total energy budget depends on the traffic load [178]. On the one side, in DEL the interface is active for shorter but transmitting higher bursts of packets each time. On the other side, CON sends a few packets per time for longer. PRO shows an intermediate behavior between the two previous approaches. In PRO, a higher value of the threshold δ translates to a higher probability to deliver data. This corresponds to different distributions of the instantaneous current peak values, as shown in Fig. 6.5(b). The peak values are below 75 for 75% of the delivery period when $\delta = 0.75$, becoming 62.5% of the delivery period when $\delta = 0.5$, and 48% when $\delta = 0.25$.

Fig. 6.6 exhibits the CDF of the packet rate transmission measured with Wireshark. Fig. 6.6(a) compares the DRMs that highlight different distributions of packet transmission rates. PRO reaches rates as high as 10 packets/s for 75% of the delivery period. Both CON and DEL achieve rates as high as 40 packets/s for 75% of the delivery period. Note that CON has higher



(a) Data Reporting Mechanisms

(b) Different thresholds for probabilistic reporting

Figure 6.6: Distribution of transmission rates

variability than PRO and DEL, which converge to a maximum rate. The technical implementation gives such behavior. While CON delivers smaller amounts of data requiring the central authority and smart device to interact often, DEL reports a bigger and unique file. Similarly to the profiles obtained for the energy in Fig. 6.5(b), Fig. 6.6(b) exhibits that the transmission rate also changes when δ increases. Note that the variation is consistent only for low rates. For instance, rates up to 5 packets/s occur for 20%, 40% and 70% of the delivery period for $\delta = 0.25$, $\delta = 0.5$ and $\delta = 0.75$ respectively.

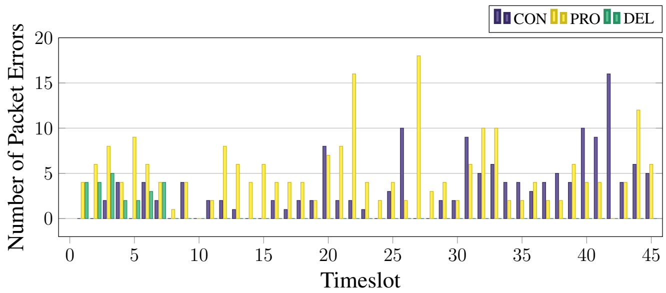
Fig. 6.7 illustrates the distribution of packet errors considering timeslots of 40 seconds. Fig. 6.7(a) compares the considered DRMs. On the one side, DEL packet errors have a concentrated distribution due to the shorter delivery period. On the other side, CON and PRO present high variability in the packet error distribution. Such behavior is given by the public network and the realistic environment. Interestingly, note that PRO exhibits a higher number of losses due to the well-known inefficiency of the 802.11 protocol for its scheduling strategy allocating single resources to single nodes [179]. Consequently, it favors CON rather than PRO mechanism. To compare PRO with CON and DEL, the threshold δ is set to 0.5. However, additional measurements are conducted with different thresholds ($\delta = [0.25, 0.5, 0.75]$), as shown in Fig. 6.7(b).

6.3 Large-scale Analysis: the Methodology

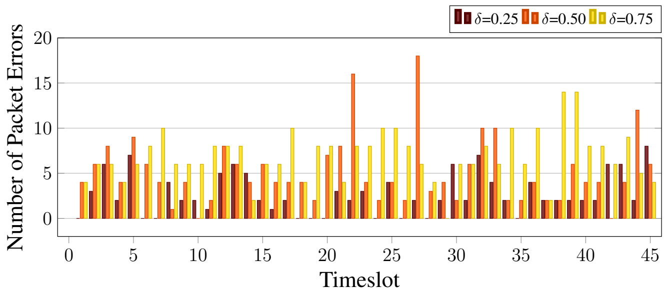
This Section presents the methodology employed to perform the large-scale evaluation.

6.3.1 Feeding CrowdSenSim

Energy measurements obtained by experiments are exploited to feed the CrowdSenSim simulator. During simulation runtime, participants gather and deliver information following the implemented DCFs. CrowdSenSim



(a) Data Reporting Mechanisms



(b) Different thresholds for probabilistic reporting

Figure 6.7: Distribution of packet errors

calculates the amount of contributed data and the energy consumption for each user [158]. Data generation exploits popular sensors embedded in smart devices, as already discussed in Subsection 6.2.3. Simulation results are obtained at the system and individual level, allowing different analyses on DRMs and DCFs, which are implemented as previously discussed. Citizens' mobility follows a profile extracted from the ParticipAct dataset [159], which provides information on the user contact per-hour. To assess the energy consumption for a realistic MCS campaign, CrowdSenSim exploits the experimental energy- and network-related measurements. The battery drain of smart devices is calculated proportionally to the contribution time. The energy consumption profiles are taken from traces of 30 minutes, as shown in Fig. 6.5(a).

6.3.2 Performance Metrics

Different performance metrics are considered to compare the proposed DCFs.

Amount of contributed data. The primary purpose of a DCF is to gather enough data to monitor phenomena or capture events. Consequently, it is fundamental to assess and quantify data according to space and temporal coverage. Different key performance indicators can be proposed to evaluate an MCS campaign, but this study considers only the amount of gathered information as the main target to analyze the proposed DCFs.

Energy consumption. It quantifies the costs participants sustain in terms of battery drain, measured in mAh.

Fairness. MCS campaigns should collect data aiming to ensure the quality of information and fair treatment to the users. In other words, participants that contribute more data and sustain higher costs should be rewarded better than others. To evaluate fairness between participants in different DCFs, the Jain Fairness Index is exploited [180]. It measures the equity when distributing a set of limited resources according to specific policies. For instance, participants that walk for longer periods than others are expected to gather more information. The data contribution fairness index (F_{D}) is defined as follows:

$$F_{\text{D}} = \frac{\left(\sum_{i=1}^{N} d_i\right)^2}{N \cdot \sum_{i=1}^{N} d_i^2}, \tag{6.1}$$

where:

$$d_i = \frac{D_i}{D_i^M}. \tag{6.2}$$

D_i is the amount of contributed information from user i and D_i^M is the maximum amount of data a participant could gather in the corresponding time. F_{D} assumes values equal to 1 when participants contribute data proportionally to their walking time. However, this index exhibits a significant shortcoming if not combined with other factors. Indeed, F_{D} does not differentiate between devices with different initial battery levels. To this end, an additional index is needed. In particular, a smartphone with a lower battery level is expected to gather less data than others with a higher battery level. The battery fairness index (F_{B}) is defined as:

$$F_{\text{B}} = \frac{\left(\sum_{i=1}^{N} b_i\right)^2}{N \cdot \sum_{i=1}^{N} b_i^2}, \tag{6.3}$$

where:

$$b_i = \frac{B_i}{B_i^T}. \tag{6.4}$$

B_i and B_i^T are the i device's battery levels in mAh. Specifically, B_i is the *battery drain* experienced while contributing and B_i^T is the total battery level when the device starts its sensing. The crowdsensing fairness index (F_{CS}) considers both battery drain and contributed data:

$$F_{\text{CS}} = \sigma \cdot F_D + (1 - \sigma) \cdot F_B, \tag{6.5}$$

where σ is a coefficient between $[0, 1]$ to balance F_D and F_B.

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

Figure 6.8: Cities considered for the evaluation and street networks.

6.4 Performance Evaluation

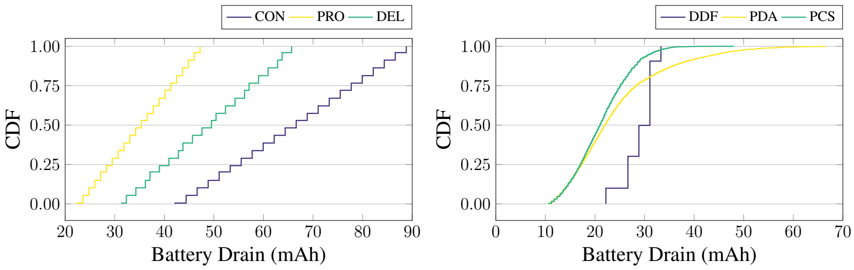
This Section discusses the performance evaluation by presenting the simulation setup and the obtained results.

6.4.1 Simulation Setting

Fig. 6.8 illustrates the selected cities to simulate the MCS campaign: Luxembourg City, Turin (Italy), and Washington DC (USA). The rationale behind the choice is twofold. The first motivation is about the different sizes. The center of Luxembourg City occupies a surface of 51.47 km^2 with a population of 114 303 inhabitants as of the end of 2017. The city center of Turin covers an area of 130.17 km^2 and has a population of 883 601 inhabitants as of the beginning of 2016. The city center of Washington DC occupies approximately a surface of 158.1 km^2 with a resident population of 672 228 inhabitants as of the end of 2015. Urban morphology is the second motivation, as it determines the street network topology. Luxembourg City presents the typical north European pattern with a high density of crossroads. Washington DC has a street network with large roads and many parallel long streets. Turin is among the two previous typologies for its roman grid morphology. As already discussed, the user arrival pattern is based on realistic mobility traces and the simulation period is 12 consecutive hours in one day. The PartecipAct dataset provides the user contacts per-hour. Pedestrians walk with speed uniformly distributed between $[1, 1.5]$ m/s in a uniformly distributed period between $[20, 40]$ minutes. The number of participants is 10 000 unless differently stated. The full battery capacity is generated by randomly picking from a list of popular smartphones, including 3300 mAh (Samsung Galaxy J7), 2800 mAh (LG G5), 2550 mAh (Samsung Galaxy S6), and 2200 mAh (Huawei P8 Lite). When devices start contributing, the initial battery percentage is uniformly distributed between $[10 - 90]$%.

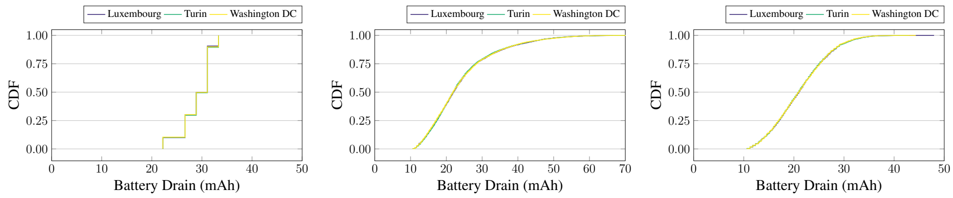
6.4.2 Simulation Results

Performed simulations assess energy consumption, amount of collected data, and fairness.



(a) Data Reporting Mechanisms **(b)** Data Collection Frameworks

Figure 6.9: CDF average battery drain per user on large scale



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

Figure 6.10: CDF of battery drain per user for DCFs in different cities

Energy Consumption. Fig. 6.9 shows the CDF of the battery drain per-user for the presented DRMs and the DCFs in Luxembourg City. Note that the difference between DRMs and DCFs is consistent. The motivation is that DRMs lack crucial features of DCFs, like feedback from the central authority or criteria to stop data collection. Fig. 6.9(a) exhibits the CDF of battery drain for the DRMs, which vary by values range and slope. The steps outline groups of participants that have stopped contribution after reaching a certain battery drain. As expected, PRO and DEL are less energy consuming than CON. However, Fig. 6.9(b) highlights that implementing costly DRM like CON in a DCF with a feature to stop data contribution like DDF is beneficial. Indeed, all devices consume at most 33 mAh. Unlike DDF, with PCS and PDA the percentage of users that spend more than 33 mAh is respectively 3% and 16%. Nonetheless, a consistent number of users consume a small amount of battery with PCS. In PDA most devices spend a higher amount of energy than PCS due to a higher amount of collected data. This aspect will be shown inFig. 6.13 and in Fig. 6.14

Fig. 6.10 shows the similarity of battery drain among various cities, highlighting that the street pattern and the size of the city have a minor impact on the energy performance of the DCFs. DDF presents a CDF as a step function, where each step indicates a group of users that contributed a similar amount of information. Therefore, they stopped because their device reached a threshold due to the battery drain or the amount of collected data.

Fig. 6.11 shows the amount of contributed data and the related DCF

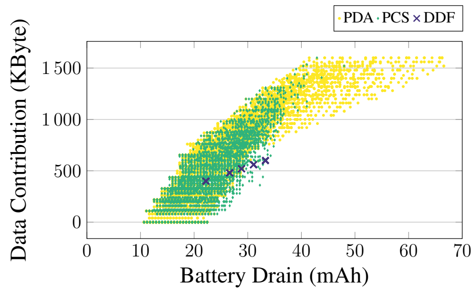
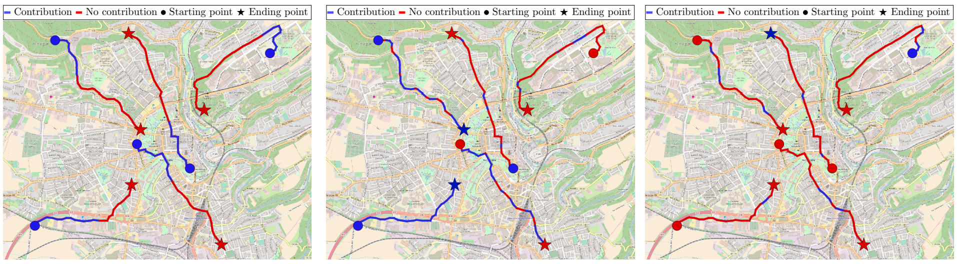


Figure 6.11: Amount of collected data and the associated battery drain in Luxembourg City



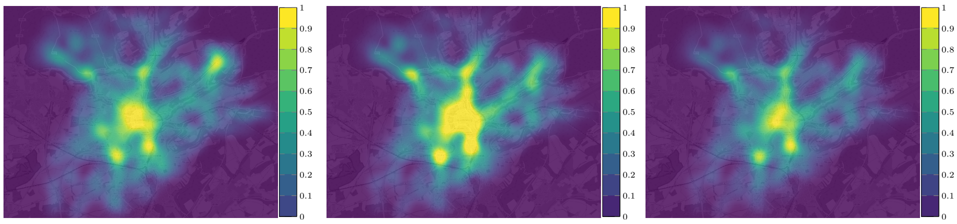
(a) DDF **(b)** PDA **(c)** PCS

Figure 6.12: User trajectories with the associated data contribution in Luxembourg City

energy consumption. Marks describe the battery drain that a group of users consumed for a specific amount of data. Note that DDF shows a few marks because the stopping policy indirectly regulates the battery drain. Hence, devices exhibit similar behaviors. Unlike DDF, PCS and PDA present much more significant variability due to their DRMs: users have different battery drains to contribute a certain amount of information. The higher the contribution, the higher the variability. From a concrete perspective, obtained results highlight that incentive mechanisms based only on the amount of contribution lack to fairly compensate participants because of the technical implementation of delivery methodologies.

Amount of Collected Data. Fig. 6.12 illustrates the trajectories of 5 pedestrians in Luxembourg City, aiming to highlight periods of active contributions and differences between DCFs. DDF enables participants to deliver data until devices reach the threshold and stop. PDA exhibits periodic delivery due to the probability and the collector feedback. In PCS, the contribution is minimal and depends on placing calls or using an application.

Fig. 6.13 illustrates the spatial coverage of collected information after a simulation runtime in Luxembourg. The heatmap is normalized between 0 and 1, where 1 represents 100 MB of data. PDA exhibits a high spatial distribution because it gathers data until the collector reaches a specific value



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

Figure 6.13: Heatmaps of Luxembourg city with different DCFs

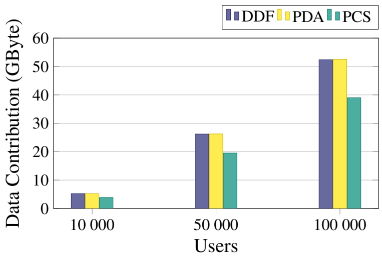
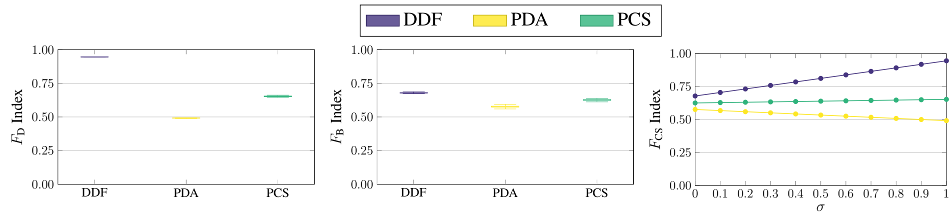


Figure 6.14: Data contribution for considered DCFs in Luxembourg City

to lower the probability threshold. DDF presents a lower amount of data in the center due to the stopping mechanism, as shown in Fig. 6.9b. PCS obtains the lowest amount of data because it only depends on phone usage.

Fig. 6.14 illustrates the contribution in Luxembourg City, comparing the DCFs for a different number of participants. DDF reaches a significant amount of data due to continuous reporting even if users stop for saving energy. PDA accomplishes a comparable amount of information. Differently, PCS obtains the lowest amount of data and would most likely fail to monitor the region of interests accurately. Again, the motivation lies in the DRM.

Fairness. Fig. 6.15 shows the different fairness indexes for each proposed DCF. 100 rounds of simulations are performed to obtain results. Fig. 6.15(a) illustrates the data contribution fairness index (F_D) as boxplots. DDF reaches the highest values of data contribution fairness for two reasons. First, participants that walk for more extended periods fairly contribute more data. Second, the central authority sends feedback to regulate users' contributions. PCS is less fair than DDF because it depends on smart devices' activities and is not proportional to the walking time. PDA has the lowest values because a few users can deliver a consistent amount of information if the collector needs data, while many participants may not contribute at all if the collector does not need additional information. Fig. 6.15(b) illustrates the battery fairness index (F_B). DDF has much lower values than in the previous index due to the stop mechanism. PDA and PCS are similar and below DDF. Even if they do not include any stop mechanism, the limited walking period provides a similar effect. Fig. 6.15(c) exhibits the MCS Index (F_{CS}) given by the



(a) Data Fairness Index **(b)** Battery Fairness Index **(c)** MCS Fairness Index

Figure 6.15: Fairness Indexes

combination of the previous two indexes. It reflects the properties of the DCFs. DDF shows a linear increase between data and battery fairness, while PDA presents a linear decrease. PCS does not include any feedback from the collector and presents the most uniform pattern.

Chapter 7

Crowdsensed Data-driven Estimation of Local Businesses Attractiveness

Public institutions, private companies, and urban planners have relied for decades on traditional strategies and experience-based methodologies to investigate cities' complex dynamics and tackle urbanization issues. Nowadays, data-driven approaches unleash unprecedented possibilities to improve citizens' quality of life in wide urban environments **rajiv-citydatafusion** Smart devices carried by citizens are a massive data source under the MCS paradigm. This Chapter illustrates how to leverage mobile devices' data to enhance the traditional approaches used for urban decisions with novel techniques. In this context, machine learning (ML) techniques enable to obtain highly accurate estimations of categories of local businesses (LBs) (e.g., shops, bars) and their attractiveness during different times of day (e.g., working or shopping hours).

7.1 Background and Motivation

Private and public owners of LBs (e.g., institutions, companies, or individuals) make several decisions to offer customers competitive services while maximizing profit. The most critical choices include an LB location according to its typology, setting the number of employees required per hour, prices, and opening hours. Practical solutions to increase the success of LBs need an understanding of cities' complex dynamics based on the citizens' mobility and spatial distribution [181]. Comprehending real-time urban mobility and estimating its variability according to special events are typical examples that could help companies decide required staff and municipalities to regulate traffic.

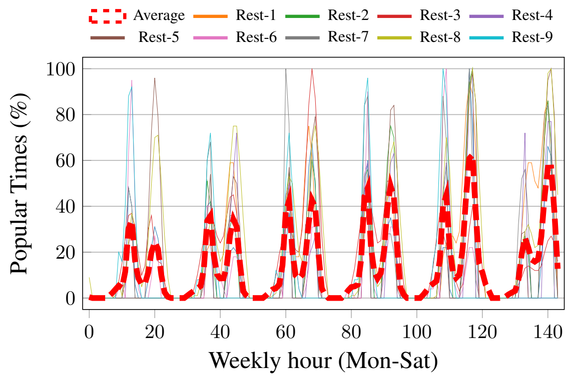
7.1.1 Related Works

Surveys catching travel behaviors or traces collected from mobile devices are standard methods to understand citizens' mobility and investigate the popularity of LBs [183], [186], [187]. Unfortunately, these strategies are inclined to consistent shortcomings, such as technical constraints (e.g., weak network accessibility), low location accuracy, user misconduct, and datasets not publicly available [184]. In this context, data-driven solutions are fundamental to improve existing approaches for understanding citizens' spatial patterns. In particular, MCS enables citizens to contribute data for different purposes, e.g., directly, differentiate residents from visitors, and recognize special events [190]. Feeding ML algorithms with information collected from MCS campaigns represents a win-win solution in several fields [191], such as detecting accident risks [192], and predicting traffic [193]. A work similar to the proposed approach aims to predict the temporal dynamics of *newly established* LBs by using a Foursquare dataset [195]. The novelty of the proposed study consists of considering existing LBs and exploiting novel Google Popular Times datasets.

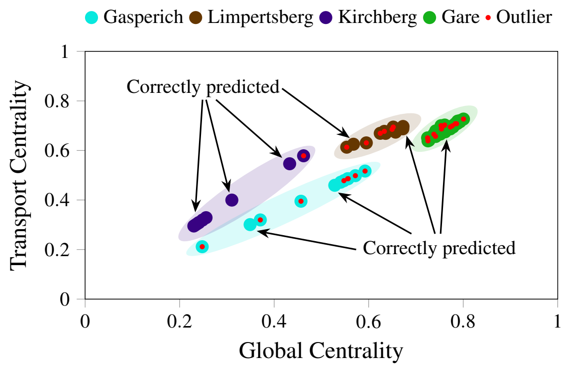
7.1.2 Crowdsensed Datasets

The data-driven approach presented in this Chapter overcomes the flaws of traditional experience-based methodologies and brings one step further the research in urban computing. Citizens share data every day by using different MCS applications (e.g., Waze, OpenStreetMap) and location-based social networks (LBSN) (e.g., Twitter or Foursquare). Large datasets are available from those contributions, allowing investigation of users' spatial patterns, travel behaviors, accessibility of urban areas. In particular, Google aggregates and anonymize data passively crowdsensed from Google Maps users[1]. These datasets enable the analysis of LBs' popularity by providing a massive amount of information, such as popular times per hour, waiting time to access the service, and the average service duration. The proposed approach utilizes Google Popular Times to obtain highly accurate estimations of LBs category and attractiveness by exploiting ML algorithms already employed for urban applications [185]. The objective is twofold. First, it analyzes the importance of different features that influence the LBs' popularity. Second, extracted features feed ML techniques to estimate the category and attractiveness of LBs. For instance, bars and restaurants are typically close in an area, while LBs like post offices or pharmacies are distributed over a city. Public transport is an example of a factor that consistently impact LBs popularity due to the reachability.

[1] https://support.google.com/business/answer/2721884



(a) Average popularity of restaurants in Ville Haute



(b) Centrality and similarity

Figure 7.1: Data aggregated from Luxembourg districts for restaurants

7.2 Preliminary Analysis

This Section highlights the shortcomings of common urban metrics in estimating LBs attractiveness accurately and paves the path to data-driven approaches. The objective focuses on LBs' popularity, their centrality on a street network, and similarity in a neighborhood. Fig. 7.1 shows results on data crowdsensed in Luxembourg City.

Weekly popularity. Google Popular Times describes the LB's weekly temporal behavior using an array of hourly values normalized in a week between [0 : 100]. Specifically, 0 represents closing times, 1 the minimum hourly number of customers in a week, and 100 the maximum. Using normalized values allows investigating the temporal profile and the most significant factors (e.g., pubs have more visits in the evenings).

Fig. 7.1(a) illustrates the behavior of 9 restaurants and their weekly average (Monday-Saturday) in the city center of Luxembourg (Ville Haute), which is characterized by offices, banks, shops, and touristic places. The maximum values of popularity are around 12, 20, 36, 44, etc., which represent lunch (noon) and dinner (8 PM) times every day. It is easy to understand the district lifestyle by comparing the temporal profiles. The highest values correspond to lunch and dinner times on weekdays, while only dinner time on Saturday because offices are closed. The most crowded day is Friday because both citizens and tourists are around the city.

The use of normalized values permits to analyze the trend of LBs during a week and its influencing factors (e.g., LBs that have more success at weekends in touristic areas or at lunchtime in business districts). This hides the degree of success of a single LB (e.g., having more customers than others), which is however not the purpose of this work.

Centrality and similarity. LBs' popularity depends on their closeness and accessibility to public transport. The centrality indicates the importance of a single node in a network and can measure the popularity. In particular, the *closeness centrality* represents the sum of all shortest paths between a node and all other nodes within the street network. This works calculates the *global-centrality* and the *transport-centrality*. The *global-centrality* defines the proximity of a LB with the other ones:

$$C_B(k) = \frac{N_B - 1}{\sum_{i \neq k} d_{ki}}, \tag{7.1}$$

where k is the k-th node, N_B is the total number of LBs and d_{ki} is the distance between a couple of nodes. The *transport-centrality* calculates the closeness of an LB with all transport facilities:

$$C_T(k) = \frac{N_T}{\sum_{i \neq k} d_{ki}}, \tag{7.2}$$

where N_T is the number of transport facilities (e.g., underground stations, bus stops) and d_{ki} is their distance with the LB. The *similarity* compares the popularity of a LB with all other LBs in a district. It is measured considering the symmetric index of Jensen-Shannon divergence (JSD), which presents better performance than the asymmetric Kullback-Leibler divergence(KLD) [195]. The similarity of two LBs i and j is:

$$J(D_i, D_j) = H\Big(\frac{D_i + D_j}{2}\Big) - \frac{H(D_i) + H(D_j)}{2}, \tag{7.3}$$

where H is the Shannon entropy, D is the temporal profile of a LB, and J represents the divergence of two temporal profiles. The similarity is between $[0 - 1]$, where 0 is the maximum and 1 is the maximum divergence.

Fig. 7.1(b) associates similarity and centrality metrics in 4 districts of Luxembourg city. Dots of the same colors have a temporal demand closer to their district. Red dots indicate LBs with temporal demands closer to other districts (outliers). Except for the Kirchberg district, which is isolated from other districts, most of the LBs are marked as outliers. Hence, this study highlights that traditional urban metrics fail to evaluate LBs' popularity and districts' association. The proposed ML-based approach will outperform traditional metrics.

7.3 ML-augmented Methodology

This section presents the ML-based approach fed with crowdsensed data. Features are extracted and selected from Google Popular Times, aiming to augment the output accuracy after the training. This procedure is omitted for space reasons.

7.3.1 Machine Learning Techniques

Considered ML techniques for multi-classification problems are Support Vector Machines (SVMs) and MultiLayer Perceptron (MLP) neural network. The selection fits well the peculiarity of the scenario under study, with an intermediate number of training samples M and few features N. *Support Vector Machines (SVMs)* map samples into output categories using kernel methods to divide a hyperplane with an optimal boundary. It requires a fine-tuning of the regularization parameter C, which regulates the correct classification of training points and the smooth decision boundary. High values lead to a hyperplane with a small margin and a high-accurate classification. Low values correspond to a higher tolerance to errors and simple decision functions, smoothing the training samples' classification. The chosen kernel method is gaussian and needs to fix the standard deviation γ, representing the influence of a single training point on other samples, which depends on its distance from the boundary. *Multilayer Perceptron (MLP)* is a feedforward artificial neural network that exploits several hidden layers to map an input vector to an output one. All layers have different nodes linked layer by layer with various weights and form a fully combined topology. The sum of all weighted inputs gives the output of each node.

7.3.2 Estimating LBs Category and Attractiveness

Two multi-classification problems are considered to estimate the LBs category and attractiveness. This paragraph discusses the extracted features and the output classes.

Extracted features. Selected features are divided between *intrinsic* and *extrinsic*. Intrinsic features do not vary consistently over time and relate to geographical properties and owners' decisions (e.g., service offered, location, closing hours). In particular, this study considers *global-centrality*, *transport-centrality*, *opening hours*, and *category*. Centralities were previously discussed. *Opening hours* is a vector of 144 hourly values (Mon-Sat) indicating when the place is open. The category indicates the provided service. Extrinsic features fluctuate continuously with time and depend on customers' interactions and behaviors (e.g., waiting and staying time). They are *popular times*, *average time of visit*, and *average waiting time*. Popular Times were already presented. Average staying and waiting time represent respectively the duration of visiting and queueing in minutes.

Output classes. The type of services offered by different LBs define their categories, classified in *public*, *store, health*, *restaurant*, and *bar*. *Public* represents offices and agencies, such as banks and institutions. *Store* indicates all types of shops and sellers (e.g., markets, clothing). *Health* includes private and public LBs associated with the healthcare system, such as doctors, dentists, hospitals. *Restaurant* comprises places that serve food where customers can sit. *Bar* represents LBs that offer mostly drinks and related services. The categories of LBs attractiveness are *working*, *nightlife*, *weekend*, *business hours* (Bus. H.), and *shopping hours* (Shop. H.). *Working* represents places with crowded hours during working breaks, such as lunchtimes during the week. It typically includes bars, cafes, shopping centers. *Nightlife* indicates rush hours in evenings and nights along all week and can comprise pubs, restaurants, clubs. *Weekend* shows high popularity only at weekends (e.g., isolated shopping malls or touristic places). *Business hours* includes peak hours from early morning to afternoon, typical of public offices and institutions. *Shopping hours* indicates the high popularity of places where citizens buy goods, with a uniform distribution during daytime all week.

7.4 Data-driven Evaluation

This Section focuses on evaluation settings, performance metrics, and results.

7.4.1 Evaluation Set-up

This study analyzes LBs Popular Times from Luxemburg city and Munich (Germany) collected between July 21st and July 30th, 2018. The selection of these two cities is given by having different properties, such as size, morphology, street network. Downloaded data consists of 1 084 LBs for Luxembourg city and 3 784 for Munich, split in 80%, 10%, and 10% for the phases of training, cross-validation, and test. Evaluation is conducted using an open-source Python-based library called Scikit-learn. For LBs category estimation input classes are *average opening hours*, *time spent*, *global-*, and *transport-centrality*. For SVMs, the hyperparameters are fixes as $\gamma = 2^{-12}$ and $C = 2^8$. Fig. 7.3a illustrates the rationale behind parameters selection. One hidden layer with 13 nodes characterizes the MLP technique, chosen after an exhaustive search through a grid-search algorithm. *Opening hours*, *category*, *district*, *popular times*, *global-*, and *transport-centrality* are the features used to predict LBs attractiveness. Hyperparameters chosen for SVM are $C = 2^6$ and $\gamma = 2^{-10}$, while MLP presents 2 hidden layers with 8 nodes per layer. The rationale to select parameters is shown in Fig. 7.3b).

7.4.2 Performance Metrics

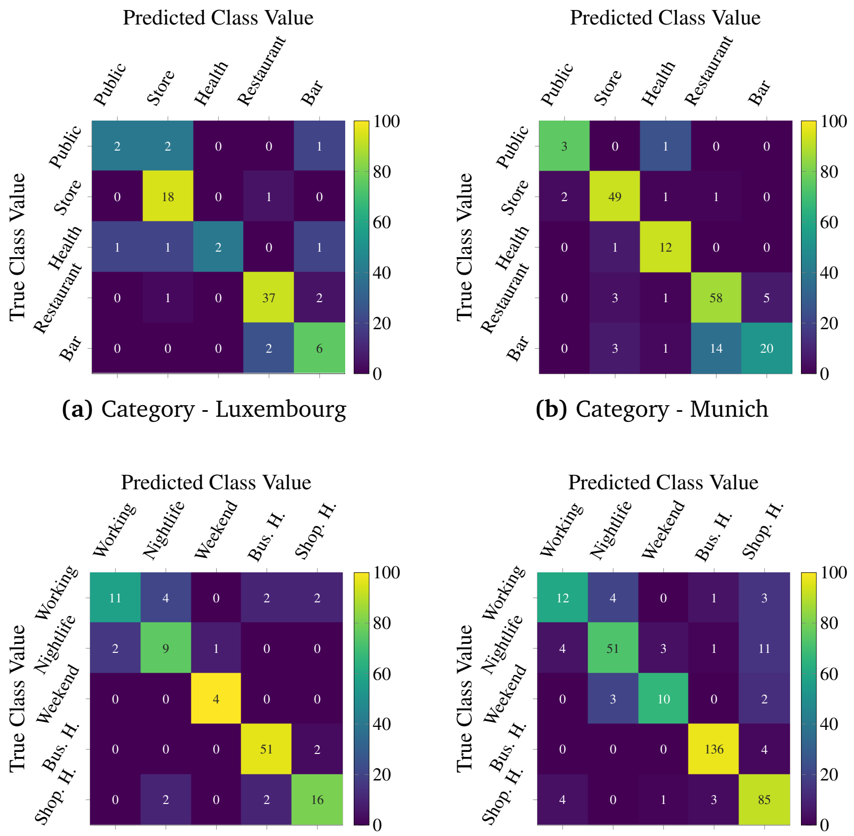
The metrics considered to evaluate performance are precision, recall, F1 score (per-class), and accuracy (average of all classes). *True positive* (tp) and *true negative* (tn) indicate a correct prediction, while *false positive* (fp) and *false negative* (fn) a wrong one. The *precision* shows the model's potential not to predict another true class as the current class and is given by the ratio between correct positive predictions and the total predicted positive observations ($tp/(tp + fp)$) The *recall* is the model ability to get all the occurrences of a class and consists of the ratio between correct predictions on positive events and all the observations in class under study ($tp/(tp + fn)$) The *F1 score* reveals when false positives and negatives have different costs, investigating inaccurate predictions and providing the weighted average of precision and recall. The *accuracy* determines the classifier's performance and is optimally used for symmetric classes (e.g., when wrong estimations have equal weights). It is given by the ratio of exact estimations over the total observations.

7.4.3 Results

Table 7.1 illustrates accurate evaluation results for precision, recall, F1 score, and accuracy on the chosen classes with MLP and SVM techniques in Luxembourg City and Munich. The evaluation of LBs categories shows a higher accuracy for Luxembourg using both approaches. On the contrary, LBs attractiveness has higher accuracy in Munich. In general, SVMs achieve higher values of accuracy compared to MLP. Precision is low for bar and public because they have properties shared with other categories, while it is high for restaurant, health, and store because they are more peculiar than others. Interestingly, precision presents distinct values in the two cities because they have different characteristics, mostly related to visitors' and citizens' lifestyles. For instance, the precision is lower in Munich because it is an international and large city with extensive opening hours (e.g., pubs until late at night), while Luxembourg has more defined closing times. The attractiveness presents a higher precision for the class working in Luxembourg because LBs crowded at job breaks are not popular at different times. Instead, Munich does not allow to satisfactorily predict the class working because those places are also popular with other types of customers at different times. Shopping and business hours exhibit better results in Munich because they tend to be concentrated for the well-organized urban plan (e.g., shopping malls). For the same motivations, the proposed approach presents high values of recall

Table 7.1: Statistics for LB category and attractiveness prediction

		Precision				Recall				F1 score				Accuracy			
		SVM		MLP		SVM		MLP		SVM		MLP		SVM		MLP	
		Lux	Mun	Lux	Mun	Lux	Mun	Lux	Mun	Lux	Mun	Lux	Mun	Lux	Mun	Lux	Mun
Category	Public	0.67	0.60	1.00	0.67	0.40	0.75	0.40	0.50	0.50	0.67	0.57	0.57				
	Store	0.82	0.88	0.81	0.87	0.95	0.92	0.89	0.91	0.88	0.90	0.85	0.89				
	Health	1.00	0.75	0.75	0.67	0.40	0.92	0.60	0.92	0.57	0.83	0.67	0.77	0.84	0.81	0.81	0.76
	Restaurant	0.93	0.79	0.90	0.77	0.93	0.87	0.88	0.79	0.93	0.83	0.89	0.78				
	Bar	0.60	0.80	0.45	0.60	0.75	0.53	0.62	0.47	0.67	0.63	0.53	0.53				
	Average	0.85	0.81	0.83	0.75	0.84	0.81	0.81	0.76	0.84	0.80	0.81	0.75				
Attractiveness	Working	0.85	0.60	1.00	0.57	0.58	0.60	0.40	0.40	0.69	0.60	0.57	0.47				
	Nightlife	0.60	0.88	0.81	0.86	0.75	0.73	0.89	0.77	0.67	0.80	0.85	0.81				
	Weekend	0.80	0.71	0.75	0.79	1.00	0.67	0.60	0.73	0.89	0.69	0.67	0.76	0.84	0.87	0.80	0.88
	Business hours	0.93	0.96	0.90	0.96	0.96	0.97	0.88	0.97	0.94	0.97	0.89	0.97				
	Shopping hours	0.80	0.81	0.45	0.81	0.80	0.91	0.62	0.92	0.80	0.86	0.53	0.86				
	Average	0.85	0.87	0.83	0.87	0.84	0.87	0.81	0.87	0.84	0.87	0.81	0.87				



(a) Category - Luxembourg

(b) Category - Munich

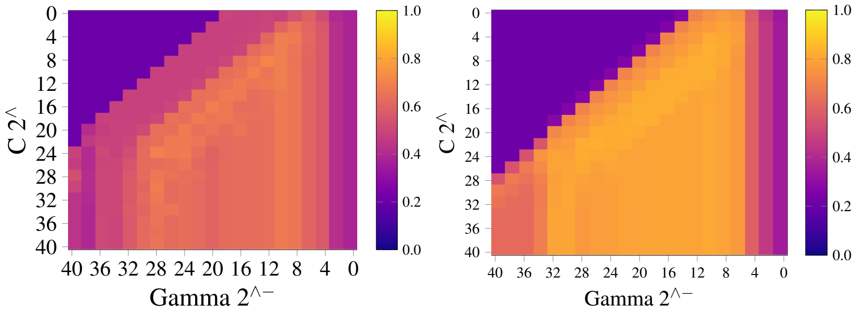
(c) Attractiveness - Luxembourg

(d) Attractiveness - Munich

Figure 7.2: Confusion matrices for LBs category and attractiveness prediction with SVM technique.

index in both cities for stores and restaurants as categories, as well as shopping and business hours as popularity. Results on the incorrect predictions given by the F1 scores show similar evidence.

Fig. 7.2 presents confusion matrices with single observations on every true or predicted class. They allow comparing different behaviors of LBs and cities. Each value shows how many observations of a predicted class for true inputs. Colors highlight the recall, which considers the percentage of correct estimations over the total. Columns exhibit predicted class values and their sum shows the total observations of each class. The diagonal presents the number of correct predictions. Differently from Table 7.1, confusion matrices allow analyzing single wrong occurrences. *Restaurant* and *store* categories exhibit higher recall values in both cities and ML approaches because they have peculiar features like opening times. Instead, *bar* and *public* have a low recall with wrong occurrences in *restaurant* and *store* for their similar behaviors. This is clearly shown in Fig. 7.2(a) and Fig. 7.2(b), where 2 bars over 8 are classified as restaurants in Luxembourg city, while



(a) Category **(b)** Attractiveness

Figure 7.3: Analysis of F1 score for SVM parameter selection in Munich.

this happens on 14 LBs over 34 in Munich. Interestingly, *health* presents divergent results in cities for the different number of places considered in the datasets. Specifically, a smaller dataset provides a lower precision. For similar reasons, Fig. 7.2(c) and Fig. 7.2(d) exhibit that *nightlife* and *working* have the highest number of wrong occurrences. *Business hours* presents the highest number of correct occurrences because peak hours are uniform along all week in both cities. *Weekend* is easier to predict in Luxembourg because it is not a touristic city and the difference between weekdays and weekends is consistent.

Fig. 7.3 illustrates the selection of parameters for SVM using the Munich dataset. It exploits F1 score as it represents a balance of recall and precision. In particular, the parameters are $C = 2^8$ and $\gamma = 2^{-12}$ for category, $C = 2^6$ and $\gamma = 2^{-10}$ for attractiveness.

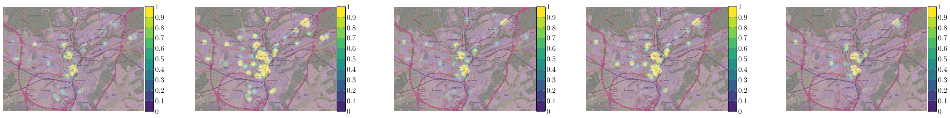
Chapter 8

Efficient Edge Data Center (EDC) Deployment in Smart Cities

Multi-access Edge Computing (MEC) is a rapidly emerging paradigm that proposes to ensure high-bandwidth and low-latency performance by deploying computational and storage resources close to end-users. To this end, this Chapter brings the research in EDCs placement one step forward by investigating citizens' mobility in vast urban environments. The final aim is to minimize network outages and increase service availability.

8.1 Background and Motivation

The European Telecommunications Standards Institute (ETSI) introduced the MEC paradigm [196] to deploy computing services and network intelligence close to end-users, aiming to improve the performance of high-bandwidth and low-latency applications. MEC offers the possibility to operate with different mobile networks (e.g., LTE, 4G, 5G) as it is independent of the network evolution. The *edge*, or MEC host, is a data center deployed close to the base station (BS) that enables storage and computing resources for applications, aggregates radio network functionalities [197], and improves the performance of applications [198]. The deployment of MEC resources for smart cities is highly promising and requires investigation efforts. A vision work proposes to deploy edge resources by exploiting existing infrastructure, such as cellular BSs, street lamps, and routers [199] but it does not consider the complex urban dynamics of a city and related traffic workloads. Other works have analyzed the impact of citizens' activities on the traffic volume [200] or how to offload computational workload to MEC host in proximity [201], [202], [203]. A recent proposal is the human-driven edge computing (HEC)



(a) H: 8:00 **(b)** H: 12:00 **(c)** H: 15:00 **(d)** H: 17:00 **(e)** H: 20:00

Figure 8.1: Traffic generation in Luxembourg City at different hours of a working day

paradigm, which proposes to combine MEC and MCS [204]. A promising edge placement driven by user allocation is discussed in [205], but it does not include mobility within the network and extensive scalability.

This Chapter proposes a novel approach to deploy EDCs in smart cities efficiently by considering human mobility. Multiple urban dynamics regulate spatial patterns of citizens, places they visit, and time spent in each place [185], [206]. In this context, crowdsensed data can provide information to infer and predict citizens' mobility with the final goal to estimate traffic workload and propose EDCs deployments that minimize outages. In particular, Google Popular Times[1] empower the reproduction of realistic spatial patterns of citizens. As assumptions, cellular connectivity ensures network access, users generate LTE traffic, and EDCs are deployed only among base stations (BSs) to utilize existing infrastructure. Fig. 8.1 exhibits heatmaps with the estimated traffic workload of BSs in Luxembourg City. Spatial patterns of citizens are generated exploiting an average weekday of Google Popular Times. Interestingly, areas including BSs with high computing demands are the city center at lunch and dinner times, the railway stations at commuting times (e.g., early morning and mid-afternoon), and the university campus (from H: 8:00 to H: 20:00). Note that traffic workloads and associated computing demands of BSs have a consistent variability during the day. Hence, this opens to investigate policies for deploying EDCs in urban environments.

8.2 Models for EDC Deployment and Urban Mobility

This Section presents the problem formulation to deploy EDCs in cities and proposes models that consider MEC aspects and complex urban dynamics.

8.2.1 Problem Formulation

Given a set of base stations $\mathcal{B} = \{b_1, \ldots, b_{N_B}\}$ at a certain location and a subset of EDCs $\mathcal{E} = \{e_1, \ldots, e_{N_E}\}$ to be deployed among the BSs, the final aim is to choose the N_E BSs to host EDCs. Re-using existing infrastructure instead of

[1] https://support.google.com/business/answer/2721884

creating new sites is key to save costs [199]. The *latency outage probability* O is the considered performance metric to evaluate the system [207]:

$$O = Pr\{L \geq D_{max}\}, \tag{8.1}$$

where D_{max} is the maximum admissible delay and L is the end-user latency. Differently from [207], O indicates a Round-Trip-Time (RTT) latency to catch at the user-side when an EDC does not accomplish a request. It can happen when an EDC is overloaded and declines the request or when a user does not receive a reply within the acceptable delay.

The problem is to choose a subset E of N_E BSs among the total N_B to deploy EDCs with the final aim to minimize O:

$$\min_{E} O. \tag{8.2}$$

8.2.2 MEC Model

A considered urban environment is split into a set of regions $\mathcal{R} = \{r_1, ..., r_l\}$. Each EDC is assigned to different BSs in a specific region and is responsible for processing, following the cloud-RAN paradigm [208]. EDCs include N_s servers with service rate μ. A task is declined when the total service rate cannot perform it in due time. Service migration between EDCs is not considered. Mobile devices generate traffic workload according to different applications [205] and always transmit to the closest BS. Poisson processes with arrival rate λ_i model tasks sent from each user u_i. *Processing* and *network delays* (e.g., application, propagation, queuing, and routing) [202] contribute to the total latency L:

$$L^i = D_p^{i,k} + D_c^k + D_p^{k,i}, \tag{8.3}$$

where $D_p^{i,k}$ is the *network delay* from device i to EDC k, D_c^k is the *processing delay* at EDC k, and $D_p^{k,i}$ is the *network delay* from EDC k to device i. A $M/M/N_s$ queue with N_s servers is used to model EDCs. D_c^k describes the required time for an EDC to accomplish a task and is measured according to [205]:

$$D_c^k = f_Q\left(\phi_k \cdot \sum_{u_i \in \mathcal{U}_l} \lambda_i\right) + 1/\mu. \tag{8.4}$$

$f_Q(\lambda)$ is the waiting time to access the service in average, and ϕ_k represents the component of workload accepted in an EDC:

$$\phi_k = \begin{cases} 1, & \text{if } \lambda_{max} > \lambda(k); \\ \frac{\lambda_{max}}{\lambda(k)}, & \text{otherwise.} \end{cases} \tag{8.5}$$

$f_Q(\lambda)$ receives the task arrival rate $\lambda(k) = \sum_{u_i} \lambda_i$ at EDC e_k and computes the average queuing time:

$$f_Q(\lambda) = \frac{C\left(N_s, \frac{\lambda}{\mu}\right)}{N_s\mu - \lambda}. \tag{8.6}$$

C is calculated by using the Erlang's formula [209]:

$$C(N_s, \rho) = \frac{\left(\frac{(N_s\rho)^{N_s}}{N_s!}\right)\left(\frac{1}{1-\rho}\right)}{\sum_{k=0}^{N_s-1} \frac{(N_s\rho)^k}{k!} + \left(\frac{(N_s\rho)^{N_s}}{N_s!}\right)\left(\frac{1}{1-\rho}\right)}, \tag{8.7}$$

where $\rho = \lambda/\mu$.

8.2.3 Citizens Mobility Model

Citizens' mobility and social interactions impact the traffic workload and required computing resources. This study generates users' spatial patterns and their temporal distribution by considering local businesses' (LBs) popularity downloaded from Google Popular Times. They provide hourly values between 0 and 100, which are normalized according to the weekly minimum and maximum number of customers. The real number of clients in each LB is not provided.

To overcome the limitation, this study considers a random value N_L^t, which represents the maximum number of customers in a specific local business L of type t. N_L^t is randomly chosen between 0 and N_{max}^t, where N_{max}^t is the maximum number of clients set for each type of local business. Popular Times datasets enable then to compute the temporal variability of customers in each LB as follows:

$$D_{L,h} = P_{h,L} \cdot N_L^t. \tag{8.8}$$

$D_{L,h}$ is the number of customers at local business L within the period h and $P_{h,L}$ is the popularity from datasets. The total demand $A_{d,h}$ for each region d is calculated combining all LBs' demands within the area:

$$A_{d,h} = \sum_{l \in L_d} D_{l,h}, \tag{8.9}$$

where L_d is the subset of LBs in d.

8.3 EDCs Deployment Policies

This work proposes two algorithms to deploy EDCs and assign BSs to them, called *distributed deployment algorithm (DDA)* and *mobility-aware deployment algorithm (MDA)*. DDA places EDCs as centroids of clusters made by BSs at comparable distances. MDA exploits users' mobility to compute the expected workload for each BS, which is used as a weight to place EDCs efficiently.

8.3.1 Distributed Deployment Algorithm (DDA)

DDA clusters BSs and deploys EDCS following the k-medoids algorithm, a variant of the more popular k-means. While k-means chooses as centroids also points not included as inputs, k-medoids select only between inputs. It allows selecting an EDC among the given BSs. This policy's main weakness is to have some EDCs over-utilized with significant delays and outages, while others result under-utilized. For this reason, this work introduces a novel approach that distributes EDCs and allocates servers in each EDC according to the traffic workload and associated computational requests.

8.3.2 Mobility-aware Deployment Algorithm (MDA)

Aiming to increase the system performance in terms of outage probability, MDA assigns EDCs according to the computational demand. To this end, it considers complex urban dynamics, such as citizens' mobility, behaviors, and social interactions. MDA also uses the k-medoids algorithm, but it allocates EDCs by calculating a cost associated with the traffic workload and related computational requests for each BS.

8.3.3 Allocation of Servers among EDCs

Selecting the number of servers in each EDC is another approach to improve system performance and reduce costs. The problem is: fixed the total number of servers, how they can be assigned to different EDCs. This work proposes two different approaches to investigate this aspect. The fixed number of servers (FNS) policy simply divides all the servers between EDCs. The proportional number of servers (PNS) approach, instead, distributes servers according to EDCs computational demand.

8.4 Performance Evaluation

This Section presents simulation set-up and obtained results.

8.4.1 Simulation Set-up

Table 8.1 shows the main parameters set for performing simulations. The chosen scenario is Luxembourg city and its mobile network infrastructure of 141 BSs[2], downloaded as a set of coordinates with latitude and longitude[3]. Mobile devices transmit traffic workload for BSs characterized by an arrival rate λ_i between $[0 - 2.99]$ [205]. Each server has a service rate $\mu = 100$. An

[2]https://data.public.lu/fr/datasets/cadastre-gsm/

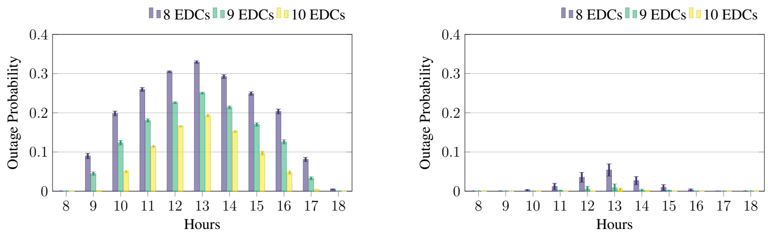
[3]https://map.geoportail.lu/

Table 8.1: Setup Parameters

SYMBOL	VALUE	DESCRIPTION
N_u	100 000	Number of users
N_b	141	Number of BSs
N_e	8	Number of edge data centers
N_l	1083	Number of total LBs
N_t	13	Number of LBs typologies
N_s	10	Number of servers in each EDC
λ_i	$0 < \lambda_i < 2.99$	Task arrival rate for user i
μ	100	Server service rate

(a) DDA **(b)** MDA

Figure 8.2: Distributed (DDA) and Mobility-aware (MDA) Deployments



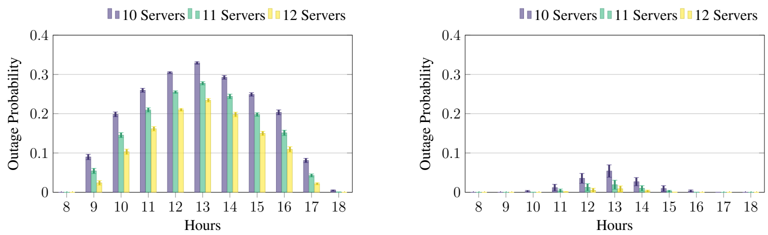
(a) DDA **(b)** MDA

Figure 8.3: Total Outage Probability in a working day with a different number of EDCs (number of servers per EDC fixed to 10)

extended and adapted version of CrowdSenSim generates citizens' mobility by considering datasets from Google Popular Times with 1 083 LBs of 13 types (e.g., bars, restaurants, banks). User arrivals across the city street network depend on spatial patterns weighted with the hourly LBs popularity for a simulation time of 24 hours fixed as the average of weekdays.

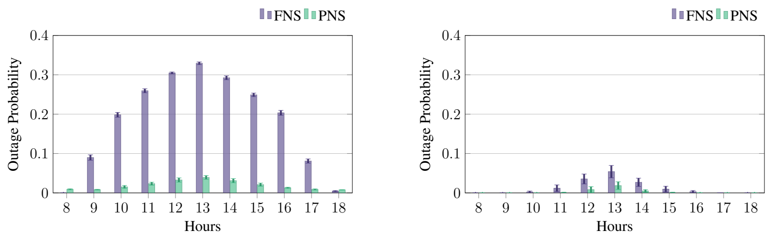
8.4.2 Simulation Results

Fig. 8.2 shows the allocation of 8 EDCs in Luxembourg city with DDA and MDA policies. Circles denote BSs, stars indicate EDCs, and different clusters have different colors. The figure exhibits that the approaches present consis-



(a) DDA **(b)** MDA

Figure 8.4: Total Outage Probability in a working day with a different number of servers (number of EDCs deployed in city fixed to 8)



(a) DDA **(b)** MDA

Figure 8.5: Total Outage Probability in a working day with FNS and PNS approaches

tently distinct deployments. Fig. 8.2(a) illustrates how DDA places EDCs to have BSs at similar ranges. Fig. 8.2(b) exhibits how EDCs tend to be closer to the center and in the north-eastern part of the city, which represents the business district (Kirchberg) and is particularly crowded during the week.

Fig. 8.3 illustrates the hourly outage probability with 10 servers per EDC and different numbers of EDCs. Comparing Fig. 8.3(a) and Fig. 8.3(b), it is clear how MDA significantly improves DDA performance. Note that for DDA the outage probability consistently decreases while increasing the number of EDCs, while MDA does not show significant improvements between 9 and 10 EDCs.

Fig. 8.4 illustrates the two algorithms with 8 EDCs and changing the number of servers for each EDC. In DDA, it is interesting to note that increasing the number of servers produces a minor effect than increasing EDCs. MDA presents proportionally better results and outperforms again DDA.

Fig. 8.5 shows the influence of server assignment between 8 EDCs. For FNS, the number of servers is fixed to 10 per EDC. As presumed, PNS presents better performance than FNS, specifically during peak hours and for DDA approach. The motivation is to assign more servers where the computational demand is higher, decreasing the outages. These result highlight how crucial is to design efficient deployments of edge resources in large-scale urban environments.

Chapter 9

Conclusion

9.1 Discussion

Since the first works on the MCS paradigm appeared, a decade has passed and researchers have investigated many aspects (e.g., task assignment, participant recruitment, and incentive mechanisms) for several applications (e.g., environmental monitoring, healthcare, public transports). The incredible evolution of ICT systems has changed the MCS scenario, enabling incredible opportunities far from imagination until a few years ago. Novel communication technologies are laying the foundation of future MCS solutions, such as 5G and MEC. A connected society is changing citizens' behavior and their social interactions, also under the impact of smart mobility and related services (e.g., car-sharing, electric scooters, food delivery). All these novelties enable an incredible pervasiveness and an unlimited set of different approaches for organizers of MCS campaigns. MCS still has many challenges to face in a world that aims to be greener and more sustainable, but it has revealed a win-win strategy to enhance existing infrastructure. In this context, MCS has demonstrated the capacity to improve context-awareness and coverage, with citizens being the platform to unify urban environments and infrastructure. For instance, taxi drivers contribute data from their devices to support civil engineering in monitoring the Harvard Bridge's vibrations (Boston, US) [7]. Asfault [211] and Safestreet [9] enable monitoring road conditions for safer driving. Smart devices can also detect emergency situations [214], such as earthquakes [212], [213], or floodings [13]. Healthcare is another field in which MCS is having a great impact, e.g., for detecting allergies and share information on healthy food [8].

9.2 Concluding Remarks

MCS is currently a well-consolidated data collection paradigm for smart city services. This dissertation presents an overview of MCS systems and different contributions to cover aspects that were not investigated yet when this study started. First of all, a comprehensive survey consolidates the MCS foundation and terminology, proposing a four-layered architecture. The architecture enables classifying all stack levels, from application to sensing layer, passing through data and communication. For each architecture layer, detailed taxonomies clarify and characterize the most relevant aspects. The thesis presents novel features developed over the popular MCS simulator CrowdSenSim, which already outperformed other existing tools. These novelties include real energy measurements that support simulated data contributions, easy-to-use mobility patterns in real-world city street networks, and realistic pedestrian mobility models. Preventing energy waste is one of the first steps to provide efficient and sustainable solutions. This dissertation proposes a novel methodology to assess energy-efficient MCS data collection frameworks. It includes models, real experimental measures performed in a laboratory, and simulations conducted in realistic urban environments. Urban planning typically relies on traditional methodologies but nowadays is crucial to provide data-driven solutions. To this end, the thesis proposes a study on learning-based estimation of local businesses' attractiveness. Finally, this manuscript also focuses on deploying edge data centers (EDCs) in cities efficiently. To empower MCS systems and crowd intelligence, it is fundamental to move computing resources closer to end-users by exploiting novel promising paradigms like multi-access edge (MEC) computing. It requires to place edge resources according to computational demand. This work proposes policies to efficiently deploy EDCs by analyzing complex urban dynamics, such as citizens' mobility and social interactions.

Appendix A

List of Publications

A.1 Chapters

1. C. Fiandrino, A. Capponi, D. Kliazovich, and P. Bouvry, “Chapter 32 - Crowdsensing architectures for smart cities”, in Micro and Nano Technologies, Nanosensors for Smart Cities, Elsevier, 2020, Pages 527-542, ISBN 9780128198704, https://doi.org/10.1016/B978-0-12-819870-4.00030-X.

A.2 Journals

1. P. Vitello, A. Capponi, C. Fiandrino, G. Cantelmo, and D. Kliazovich, “Mobility-Driven and Energy-Efficient Deployment of Edge Data Centers in Urban Environments”, in IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING. DOI: 10.1109/TSUSC.2021.3056621.

2. F. Montori, L. Bedogni, C. Fiandrino, A. Capponi, L. Bononi, “Performance evaluation of hybrid crowdsensing systems with stateful CrowdSenSim 2.0 simulator”, Computer Communications, Volume 161, 2020, Pages 225-237, https://doi.org/10.1016/j.comcom.2020.07.021.

3. A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, “A Survey on Mobile Crowdsensing Systems: Challenges, Solutions, and Opportunities,” in IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2419-2465, thirdquarter 2019, ISSN: 1553-877X. DOI: 10.1109/COMST.2019.2914030.

4. M. Tomasoni, A. Capponi, C. Fiandrino, D. Kliazovich, F. Granelli, and P. Bouvry, “Why energy matters? Profiling energy consumption of mobile crowdsensing data collection frameworks,” in Pervasive and Mobile Computing, vol. 51, pp. 193 – 208, 2018, ISSN 1574-1192, https://doi.org/10.1016/j.pmcj.2018.10.002.

5. 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”, in IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING, vol. 2, no. 1, pp. 3-16, 1 Jan.-March 2017, ISSN: 2377-3782. DOI: 10.1109/TSUSC.2017.2666043.

6. C. Fiandrino, A. Capponi, G. Cacciatore, D. Kliazovich, P. Bouvry, B. Kantarci, F. Granelli, and S. Giordano, "CrowdSenSim: a Simulation Platform for Mobile Crowdsensing in Realistic Urban Environments," in IEEE Access, vol. 5, pp. 3490-3503, 2017. DOI: 10.1109/ACCESS.2017.2671678.

A.3 Conferences

1. P. Vitello, A. Capponi, C. Fiandrino, G. Cantelm,o and D. Kliazovich, “The Impact of Human Mobility on Edge Data Center Deployment in Urban Environments,” 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9013618. **Best Paper Award**

2. A. Capponi, P. Vitello, C. Fiandrino, G. Cantelmo, D. Kliazovich, U. Sorger, and P. Bouvry, "Crowdsensed Data Learning-Driven Prediction of Local Businesses Attractiveness in Smart Cities," 2019 IEEE Symposium on Computers and Communications (ISCC), Barcelona, Spain, 2019, pp. 1-6, doi: 10.1109/ISCC47284.2019.8969771.

3. M. Tomasoni, A. Capponi, C. Fiandrino, D. Kliazovich, F. Granelli and P. Bouvry, “Profiling Energy Efficiency of Mobile Crowdsensing Data Collection Frameworks for Smart City Applications,” in 6th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud), Bamberg, 2018, pp. 1-8, ISSN: 2573-7562. DOI: 10.1109/MobileCloud.2018.00009.

4. P. Vitello, A. Capponi, C. Fiandrino, et al. "Collaborative Data Delivery for Smart City-Oriented Mobile Crowdsensing Systems," 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 2018, pp. 1-6, doi: 10.1109/GLOCOM.2018.8648047.

5. A. Capponi, "Energy-Efficient Data Acquisition in Mobile Crowdsensing Systems," 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Chania, 2018, pp. 14-16, doi: 10.1109/WoWMoM.2018.8449764.

6. P. Vitello, A. Capponi, C. Fiandrino, P. Giaccone, D. Kliazovich and P. Bouvry, "High-Precision Design of Pedestrian Mobility for Smart City

Simulators," IEEE International Conference on Communications (ICC), Kansas City, MO, 2018, pp. 1-6, doi: 10.1109/ICC.2018.8422599.

7. 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”, 8th IEEE International Conference on Cloud Computing Technology and Science (CloudCom), 2016, pp. 456-459, doi: 10.1109/CloudCom.2016.0077.

A.4 Workshops

1. A. Capponi, C. Fiandrino, D. Kliazovich and P. Bouvry, "Energy efficient data collection in opportunistic mobile crowdsensing architectures for smart cities," 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Atlanta, GA, 2017, pp. 307-312, doi: 10.1109/INFCOMW.2017.8116394.

Bibliography

[1] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, “A first look at traffic on smartphones”, in *Proceedings of the 10th acm sigcomm conference on internet measurement*, ser. IMC, Melbourne, Australia, 2010, pp. 281–287, ISBN: 978-1-4503-0483-2. DOI: 10.1145/1879141.1879176 (cit. on pp. 13, 19).

[2] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, “A survey of mobile cloud computing: Architecture, applications, and approaches”, *Wireless communications and mobile computing*, vol. 13, no. 18, pp. 1587–1611, 2013, ISSN: 1530-8677. DOI: 10.1002/wcm.1203 (cit. on pp. 13, 19).

[3] *Gartner says global smartphone sales declined 20% in first quarter of 2020 due to covid-19 impact*, Gartner, 2020. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2020-06-01-gartner-says-global-smartphone-sales-declined-20--in- (cit. on p. 13).

[4] *Gartner says worldwide wearable device sales to grow 26 percent in 2019*, Gartner, 2018. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow- (cit. on p. 13).

[5] *Wearable technology statistics and trends 2018*, smartinsights, 2017. [Online]. Available: https://www.smartinsights.com/digital-marketing-strategy/wearables-statistics-2017/ (cit. on p. 13).

[6] MarketsandMarkets, *Crowd analytics market worth $1,142.5 million by 2021*, 2018. [Online]. Available: https://www.marketsandmarkets.com/PressReleases/crowd-analytics.asp (cit. on p. 13).

[7] T. J. Matarazzo, P. Santi, S. N. Pakzad, K. Carter, C. Ratti, B. Moaveni, C. Osgood, and N. Jacob, “Crowdsensing framework for monitoring bridge vibrations using moving smartphones”, *Proceedings of the ieee*, vol. 106, no. 4, pp. 577–593, 2018, ISSN: 0018-9219. DOI: 10.1109/JPROC.2018.2808759 (cit. on pp. 13, 21, 83).

[8] 6sensorlabs. (2015). Gluten sensor. [Online] (cit. on pp. 13, 36, 37, 41–43, 83).

[9] V. Singh, D. Chander, U. Chhaparia, and B. Raman, “SafeStreet: An automated road anomaly detection and early-warning system using mobile crowdsensing”, in *Ieee international conference on communication systems networks (comsnets)*, 2018, pp. 549–552. DOI: 10.1109/COMSNETS.2018.8328270 (cit. on pp. 14, 83).

[10] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, “Participatory air pollution monitoring using smartphones”, in *Mobile sensing: From smartphones and wearables to big data*, ACM, 2012 (cit. on pp. 14, 25, 41).

[11] 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*, 2013, pp. 56–64. DOI: 10.1109/LCNW.2013.6758498 (cit. on pp. 14, 21).

[12] L. Liu, W. Liu, Y. Zheng, H. Ma, and C. Zhang, “Third-Eye: A mobilephone-enabled crowdsensing system for air quality monitoring”, *Proceedings of the acm on interactive, mobile, wearable and ubiquitous technologies*, vol. 2, no. 1, pp. 1–26, Mar. 2018, ISSN: 2474-9567. DOI: 10.1145/3191752 (cit. on p. 14).

[13] 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 acm conference on human factors in computing systems*, ser. CHI, 2011, pp. 2125–2134. DOI: 10.1145/1978942.1979251 (cit. on pp. 14, 21, 83).

[14] S. Reddy and V. Samanta, *Urban Sensing: Garbage Watch*, UCLA Center for Embedded Networked Sensing, 2011 (cit. on pp. 14, 21).

[15] D. Bonino, M. T. D. Alizo, C. Pastrone, and M. Spirito, “WasteApp: Smarter waste recycling for smart citizens”, in *International multidisciplinary conference on computer and energy science (splitech)*, 2016, pp. 1–6. DOI: 10.1109/SpliTech.2016.7555951 (cit. on pp. 14, 21).

[16] R. Fakoor, M. Raj, A. Nazi, M. Di Francesco, and S. K. Das, “An integrated cloud-based framework for mobile phone sensing”, in *Proceedings of the acm workshop on mobile cloud computing*, 2012, pp. 47–52 (cit. on pp. 14, 22, 29).

[17] D. Huang, T. Xing, and H. Wu, “Mobile cloud computing service models: A user-centric approach”, *Ieee network*, vol. 27, no. 5, pp. 6–11, 2013, ISSN: 0890-8044. DOI: 10.1109/MNET.2013.6616109 (cit. on pp. 14, 22, 29).

[18] C. Doukas and I. Maglogiannis, “Bringing IoT and cloud computing towards pervasive Healthcare”, in *International conference on innovative mobile and internet services in ubiquitous computing*, 2012, pp. 922–926. DOI: 10.1109/IMIS.2012.26 (cit. on pp. 14, 22, 29).

[19] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog computing and its role in the Internet of Things”, in *Proceedings of the acm workshop on mobile cloud computing*, ser. MCC, 2012, pp. 13–16, ISBN: 978-1-4503-1519-7. DOI: 10.1145/2342509.2342513 (cit. on pp. 14, 22).

[20] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, “On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration”, *Ieee communications surveys tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017. DOI: 10.1109/COMST.2017.2705720 (cit. on pp. 14, 23).

[21] O. Alvear, C. T. Calafate, J.-C. Cano, and P. Manzoni, “Crowdsensing in smart cities: Overview, platforms, and environment sensing issues”, *Sensors*, vol. 18, no. 2, 2018, ISSN: 1424-8220. DOI: 10.3390/s18020460 (cit. on pp. 14, 19).

[22] H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, and A. Oliveira, "Smart cities and the future Internet: Towards cooperation frameworks for open innovation", in *The future internet*, ser. Lecture Notes in Computer Science, vol. 6656, Springer Berlin Heidelberg, 2011, pp. 431–446, ISBN: 978-3-642-20897-3. DOI: 10.1007/978-3-642-20898-0_31 (cit. on pp. 14, 19).

[23] UN, *68% of the world population projected to live in urban areas by 2050, says un*, 2018. [Online]. Available: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html (cit. on p. 18).

[24] 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, 2008, pp. 756–760. DOI: 10.1126/science.1150195. eprint: http://www.sciencemag.org/content/319/5864/756.full.pdf (cit. on p. 18).

[25] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities", *Ieee internet of things journal*, vol. 1, no. 1, pp. 22–32, 2014, ISSN: 2327-4662. DOI: 10.1109/JIOT.2014.2306328 (cit. on pp. 19, 29).

[26] R. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges", *Ieee communications magazine*, vol. 49, no. 11, pp. 32–39, 2011, ISSN: 0163-6804. DOI: 10.1109/MCOM.2011.6069707 (cit. on pp. 20, 27).

[27] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, "A survey of mobile phone sensing", *Ieee communications magazine*, vol. 48, no. 9, pp. 140–150, 2010, ISSN: 0163-6804. DOI: 10.1109/MCOM.2010.5560598 (cit. on pp. 20, 25, 27, 41).

[28] 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, ISSN: 1553-877X. DOI: 10.1109/SURV.2012.031412.00077 (cit. on pp. 20, 21, 25, 27, 41).

[29] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing", in *Ieee international conference on pervasive computing and communications workshops (percom workshops)*, 2014, pp. 593–598. DOI: 10.1109/PerComW.2014.6815273 (cit. on pp. 20, 27, 43).

[30] J. Sun, "Incentive mechanisms for mobile crowd sensing: Current states and challenges of work", *Corr*, vol. abs/1310.8364, 2013. [Online]. Available: http://arxiv.org/abs/1310.8364 (cit. on pp. 20, 36).

[31] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, and X. Mao, "Incentives for mobile crowd sensing: A survey", *Ieee communications surveys tutorials*, vol. 18, no. 1, pp. 54–67, 2016, ISSN: 1553-877X. DOI: 10.1109/COMST.2015.2415528 (cit. on pp. 20, 25, 36).

[32] L. Jaimes, I. Vergara-Laurens, and A. Raij, "A survey of incentive techniques for mobile crowd sensing", *Ieee internet of things journal*, vol. 2, no. 5, pp. 370–380, 2015, ISSN: 2327-4662. DOI: 10.1109/JIOT.2015.2409151 (cit. on p. 20).

[33] K. Ota, M. Dong, J. Gui, and A. Liu, “Quoin: Incentive mechanisms for crowd sensing networks”, *Ieee network*, vol. 32, no. 2, pp. 114–119, 2018, ISSN: 0890-8044. DOI: 10.1109/MNET.2017.1500151 (cit. on p. 20).

[34] L. Pournajaf, L. Xiong, D. A. Garcia-Ulloa, and V. Sunderam, “A survey on privacy in mobile crowd sensing task management”, *Technical report tr-2014-002, department of mathematics and computer science, emory university*, 2014 (cit. on pp. 20, 26, 27, 35, 36).

[35] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, “A survey on privacy in mobile participatory sensing applications”, *Journal of systems and software*, vol. 84, no. 11, pp. 1928–1946, 2011 (cit. on pp. 20, 26, 27).

[36] A. B. Zaslavsky, C. Perera, and D. Georgakopoulos, “Sensing as a Service and Big Data”, *Corr*, vol. abs/1301.0159, 2013. [Online]. Available: http://arxiv.org/abs/1301.0159.

[37] X. Sheng, J. Tang, X. Xiao, and G. Xue, “Sensing as a Service: Challenges, solutions and future directions”, *Ieee sensors journal*, vol. 13, no. 10, pp. 3733–3741, 2013 (cit. on pp. 22, 29).

[38] 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)*, 2016, pp. 456–459. DOI: 10.1109/CloudCom.2016.0077.

[39] 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, 2016, ISSN: 1063-6692. DOI: 10.1109/TNET.2015.2418191.

[40] 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, 2017, ISSN: 0733-8716. DOI: 10.1109/JSAC.2017.2695738 (cit. on p. 20).

[41] 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, 2014, ISSN: 1530-437X. DOI: 10.1109/JSEN.2014.2344023 (cit. on p. 21).

[42] K. Farkas, G. Feher, A. Benczur, and C. Sidlo, “Crowdsending based public transport information service in smart cities”, *Ieee communications magazine*, vol. 53, no. 8, pp. 158–165, 2015, ISSN: 0163-6804. DOI: 10.1109/MCOM.2015.7180523 (cit. on p. 21).

[43] M. Moreno, A. Skarmeta, and A. Jara, “How to intelligently make sense of real data of smart cities”, in *International conference on recent advances in internet of things (riot)*, 2015, pp. 1–6. DOI: 10.1109/RIOT.2015.7104899 (cit. on p. 21).

[44] S. Nawaz, C. Efstratiou, and C. Mascolo, “ParkSense: A smartphone based sensing system for on-street parking”, in *Proceedings of the acm conference on mobile computing and networking*, ser. MobiCom, 2013, pp. 75–86, ISBN: 978-1-4503-1999-7. DOI: 10.1145/2500423.2500438 (cit. on p. 21).

[45] J. Cherian, J. Luo, H. Guo, S. S. Ho, and R. Wisbrun, “ParkGauge: Gauging the occupancy of parking garages with crowdsensed parking characteristics”, in *Ieee international conference on mobile data management (mdm)*, vol. 1, 2016, pp. 92–101. DOI: 10.1109/MDM.2016.26 (cit. on pp. 21, 40).

[46] 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, 2016, ISSN: 0733-8716. DOI: 10.1109/JSAC.2016.2621359 (cit. on pp. 22, 52).

[47] 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, 2017, ISSN: 0733-8716. DOI: 10.1109/JSAC.2017.2659258 (cit. on pp. 22, 52).

[48] 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, 2017, ISSN: 2377-3782. DOI: 10.1109/TSUSC.2017.2702060 (cit. on pp. 22, 53).

[49] B. Kantarci and H. T. Mouftah, “Sensing services in cloud-centric Internet of Things: A survey, taxonomy and challenges”, in *Ieee international conference on communication workshop (iccw)*, 2015, pp. 1865–1870. DOI: 10.1109/ICCW.2015.7247452 (cit. on pp. 22, 29).

[50] R. D. Lauro, F. Lucarelli, and R. Montella, “SIaaS - sensing instrument as a service using cloud computing to turn physical instrument into ubiquitous service”, in *Ieee 10th international symposium on parallel and distributed processing with applications*, 2012, pp. 861–862. DOI: 10.1109/ISPA.2012.135 (cit. on pp. 22, 29).

[51] M. Boukhechba, A. R. Daros, K. Fua, P. I. Chow, B. A. Teachman, and L. E. Barnes, “DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones”, in *Ieee/acm conference on connected health: Applications, systems and engineering technologies (chase)*, 2018 (cit. on pp. 22, 29).

[52] P. Hu, S. Dhelim, H. Ning, and T. Qiu, “Survey on fog computing: Architecture, key technologies, applications and open issues”, *Journal of network and computer applications*, vol. 98, pp. 27 –42, 2017, ISSN: 1084-8045. DOI: https://doi.org/10.1016/j.jnca.2017.09.002 (cit. on p. 22).

[53] C. Fiandrino, N. Allio, D. Kliazovich, P. Giaccone, and P. Bouvry, “Profiling performance of application partitioning for wearable devices in mobile cloud and fog computing”, *Ieee access*, vol. 7, pp. 12 156–12 166, 2019, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2892508 (cit. on p. 22).

[54] 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, 2015, ISSN: 2329-924X (cit. on pp. 22, 53).

[55] 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, 2017. DOI: 10.1109/TSUSC.2017.2702060 (cit. on p. 22).

[56] P. Bellavista, S. Chessa, L. Foschini, L. Gioia, and M. Girolami, "Human-enabled edge computing: Exploiting the crowd as a dynamic extension of mobile edge computing", *Ieee communications magazine*, vol. 56, no. 1, pp. 145–155, 2018, ISSN: 0163-6804. DOI: 10.1109/MCOM.2017.1700385 (cit. on p. 22).

[57] R. Roman, J. Lopez, and M. Mambo, "Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges", *Future generation computer systems*, vol. 78, pp. 680 –698, 2018, ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2016.11.009 (cit. on p. 22).

[58] P. Porambage, J. Okwuibe, M. Liyanage, M. Ylianttila, and T. Taleb, "Survey on multi-access edge computing for Internet of Things realization", *Ieee communications surveys tutorials*, pp. 1–1, 2018. DOI: 10.1109/COMST.2018.2849509 (cit. on p. 23).

[59] S. Yang, J. Bian, L. Wang, H. Zhu, Y. Fu, and H. Xiong, "Edgesense: Edge-mediated spatial-temporal crowdsensing", *Ieee access*, pp. 1–1, 2018, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2870298 (cit. on p. 23).

[60] Z. Zhou, H. Liao, B. Gu, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "Robust mobile crowd sensing: When deep learning meets edge computing", *Ieee network*, vol. 32, no. 4, pp. 54–60, 2018, ISSN: 0890-8044. DOI: 10.1109/MNET.2018.1700442 (cit. on p. 23).

[61] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm", *Acm comput. surv.*, vol. 48, no. 1, pp. 1–31, Aug. 2015, ISSN: 0360-0300. DOI: 10.1145/2794400 (cit. on pp. 24, 27, 29).

[62] B. Guo, Q. Han, H. Chen, L. Shangguan, Z. Zhou, and Z. Yu, "The emergence of visual crowdsensing: Challenges and opportunities", *Ieee communications surveys tutorials*, vol. 19, no. 4, pp. 2526–2543, 2017. DOI: 10.1109/COMST.2017.2726686 (cit. on p. 30).

[63] J. Phuttharak and S. W. Loke, "A review of mobile crowdsourcing architectures and challenges: Towards crowd-empowered Internet-of-Things", *Ieee access*, pp. 1–22, 2018, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2885353 (cit. on pp. 25, 30).

[64] J. Liu, H. Shen, H. S. Narman, W. Chung, and Z. Lin, "A survey of mobile crowdsensing techniques: A critical component for the Internet of Things", *Acm transactions on cyber-physical systems*, vol. 2, no. 3, pp. 1–26, 2018, ISSN: 2378-962X. DOI: 10.1145/3185504 (cit. on p. 25).

[65] K. Abualsaud, T. M. Elfouly, T. Khattab, E. Yaacoub, L. S. Ismail, M. H. Ahmed, and M. Guizani, "A survey on mobile crowd-sensing and its applications in the IoT era", *Ieee access*, vol. 7, pp. 3855–3881, 2019, ISSN: 2169-3536 (cit. on p. 25).

[66] Z. Xu, L. Mei, K.-K. R. Choo, Z. Lv, C. Hu, X. Luo, and Y. Liu, "Mobile crowd sensing of human-like intelligence using social sensors: A survey", *Neurocomputing*, pp. –, 2017, ISSN: 0925-2312. DOI: `https://doi.org/10.1016/j.neucom.2017.01.127` (cit. on p. 25).

[67] F. Restuccia, N. Ghosh, S. Bhattacharjee, S. K. Das, and T. Melodia, "Quality of information in mobile crowdsensing: Survey and research challenges", *Acm transactions on sensor networks (tosn)*, vol. 13, no. 4, pp. 1–43, 2017, ISSN: 1550-4859. DOI: `10.1145/3139256` (cit. on p. 25).

[68] A. Overeem, J. R Robinson, H. Leijnse, G.-J. Steeneveld, B. P Horn, and R. Uijlenhoet, "Crowdsourcing urban air temperatures from smartphone battery temperatures", *Geophysical research letters*, vol. 40, no. 15, pp. 4081–4085, 2013. DOI: `10.1002/grl.50786` (cit. on p. 25).

[69] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common sense: Participatory urban sensing using a network of handheld air quality monitors", in *Proc. of the acm conference on embedded networked sensor systems*, 2009, pp. 349–350 (cit. on pp. 25, 35, 36, 38, 43).

[70] M. Liu, "A study of mobile sensing using smartphones", *International journal of distributed sensor networks*, vol. 9, no. 3, p. 272 916, 2013. DOI: `10.1155/2013/272916` (cit. on pp. 25, 41).

[71] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks", *Ieee communications magazine*, vol. 40, no. 8, pp. 102–114, Aug. 2002, ISSN: 0163-6804. DOI: `10.1109/MCOM.2002.1024422` (cit. on p. 25).

[72] P. Rawat, K. D. Singh, H. Chaouchi, and J. M. Bonnin, "Wireless sensor networks: A survey on recent developments and potential synergies", *The journal of supercomputing*, vol. 68, no. 1, pp. 1–48, 2014 (cit. on p. 25).

[73] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey", *Computer networks*, vol. 52, no. 12, pp. 2292 –2330, 2008. DOI: `https://doi.org/10.1016/j.comnet.2008.04.002` (cit. on p. 25).

[74] C. Toth and G. Jóźków, "Remote sensing platforms and sensors: A survey", *Isprs journal of photogrammetry and remote sensing*, vol. 115, no. Supplement C, pp. 22 –36, 2016, ISSN: 0924-2716. DOI: `https://doi.org/10.1016/j.isprsjprs.2015.10.004` (cit. on p. 41).

[75] H. Gao, C. H. Liu, W. Wang, J. Zhao, Z. Song, X. Su, J. Crowcroft, and K. K. Leung, "A survey of incentive mechanisms for participatory sensing", *Ieee communications surveys tutorials*, vol. 17, no. 2, pp. 918–943, 2015, ISSN: 1553-877X. DOI: `10.1109/COMST.2014.2387836` (cit. on p. 25).

[76] F. Restuccia, S. K. Das, and J. Payton, "Incentive mechanisms for participatory sensing: Survey and research challenges", *Acm transactions on sensor networks (tosn)*, vol. 12, no. 2, pp. 1–40, 2016. DOI: `10.1145/2888398` (cit. on p. 25).

[77] J. Howe, *The rise of crowdsourcing*, Wired, 2006. [Online]. Available: `https://www.wired.com/2006/06/crowds/` (cit. on pp. 26, 29).

[78] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, “Participatory sensing”, *Center for embedded network sensing*, 2006 (cit. on pp. 26, 43).

[79] S. Reddy, A. Parker, J. Hyman, J. Burke, D. Estrin, and M. Hansen, “Image browsing, processing, and clustering for participatory sensing: Lessons from a DietSense prototype”, in *Proceedings of the acm workshop on embedded networked sensors*, 2007, pp. 13–17 (cit. on pp. 26, 28).

[80] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, “Sensing meets mobile social networks: The design, implementation and evaluation of the cenceme application”, in *Proceedings of the acm conference on embedded network sensor systems*, 2008, pp. 337–350 (cit. on pp. 27, 35).

[81] S. Gaonkar, J. Li, R. R. Choudhury, L. Cox, and A. Schmidt, “Micro-blog: Sharing and querying content through mobile phones and social participation”, in *Proc. of the acm international conference on mobile systems, applications, and services*, 2008, pp. 174–186 (cit. on p. 27).

[82] G. Cardone, L. Foschini, P. Bellavista, A. Corradi, C. Borcea, M. Talasila, and R. Curtmola, “Fostering participaction in smart cities: A geo-social crowdsensing platform”, *Ieee communications magazine*, vol. 51, no. 6, pp. 112–119, 2013, ISSN: 0163-6804. DOI: 10.1109/MCOM.2013.6525603 (cit. on p. 27).

[83] N. Vastardis and K. Yang, “Mobile social networks: Architectures, social properties, and key research challenges”, *Ieee communications surveys & tutorials*, vol. 15, no. 3, pp. 1355–1371, 2013 (cit. on p. 27).

[84] P. Jayaraman, C. Perera, D. Georgakopoulos, and A. Zaslavsky, “Efficient opportunistic sensing using mobile collaborative platform MOSDEN”, in *International conference conference on collaborative computing: Networking, applications and worksharing (collaboratecom)*, 2013, pp. 77–86 (cit. on p. 27).

[85] G. Cardone, A. Cirri, A. Corradi, and L. Foschini, “The ParticipAct mobile crowd sensing living lab: The testbed for smart cities”, *Ieee communications magazine*, vol. 52, no. 10, pp. 78–85, 2014, ISSN: 0163-6804. DOI: 10.1109/MCOM.2014.6917406 (cit. on p. 27).

[86] 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, 2014, ISSN: 2327-4662. DOI: 10.1109/JIOT.2014.2337886 (cit. on p. 27).

[87] H. Ma, D. Zhao, and P. Yuan, “Opportunities in mobile crowd sensing”, *Ieee communications magazine*, vol. 52, no. 8, pp. 29–35, 2014, ISSN: 0163-6804. DOI: 10.1109/MCOM.2014.6871666.

[88] C. Tanas and J. Herrera-Joancomartí, “Crowdsensing simulation using ns-3”, *Citizen in sensor networks: Second international workshop, citisens*, J. Nin and D. Villatoro, Eds., pp. 47–58, 2014, Springer International Publishing (cit. on pp. 27, 44).

[89] S. Chessa, A. Corradi, L. Foschini, and M. Girolami, "Empowering mobile crowdsensing through social and ad hoc networking", *Ieee communications magazine*, vol. 54, no. 7, pp. 108–114, 2016, ISSN: 0163-6804. DOI: 10.1109/MCOM.2016.7509387 (cit. on p. 27).

[90] 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, 2017, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2671678 (cit. on pp. 28, 45).

[91] *Runtastic*. [Online]. Available: https://www.runtastic.com (cit. on p. 28).

[92] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, *et al.*, "Activity sensing in the wild: A field trial of UbiFit garden", in *Proc. of the acm conference on human factors in computing systems*, ser. SIGCHI, 2008, pp. 1797–1806 (cit. on p. 28).

[93] J. Dai, X. Bai, Z. Yang, Z. Shen, and D. Xuan, "PerFallD: A pervasive fall detection system using mobile phones", in *Ieee international conference on pervasive computing and communications workshops (percom workshops)*, 2010, pp. 292–297 (cit. on p. 28).

[94] Y. He, Y. Li, and S.-O. Bao, "Fall detection by built-in tri-accelerometer of smartphone", in *Ieee embs international conference on biomedical and health informatics (bhi)*, 2012, pp. 184–187 (cit. on p. 28).

[95] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers", *Acm sigkdd explorations newsletter*, vol. 12, no. 2, pp. 74–82, 2011 (cit. on p. 28).

[96] M. A. Habib, M. S. Mohktar, S. B. Kamaruzzaman, K. S. Lim, T. M. Pin, and F. Ibrahim, "Smartphone-based solutions for fall detection and prevention: Challenges and open issues", *Sensors*, vol. 14, no. 4, pp. 7181–7208, 2014 (cit. on p. 28).

[97] S. Wang, C. Chen, and J. Ma, "Accelerometer based transportation mode recognition on mobile phones", in *Asia-pacific conference on wearable computing systems (apwcs)*, IEEE, 2010, pp. 44–46 (cit. on p. 28).

[98] S. Reddy, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Determining transportation mode on mobile phones", in *Ieee international symposium on wearable computers (iswc)*, 2008, pp. 25–28 (cit. on p. 28).

[99] S. Hemminki, P. Nurmi, and S. Tarkoma, "Accelerometer-based transportation mode detection on smartphones", in *Proc. of the acm conference on embedded networked sensor systems*, ser. SenSys, 2013, p. 13 (cit. on p. 28).

[100] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using mobile phones to determine transportation modes", *Acm transactions on sensor networks (tosn)*, vol. 6, no. 2, p. 13, 2010 (cit. on p. 28).

[101] Y. Michalevsky, D. Boneh, and G. Nakibly, "Gyrophone: Recognizing speech from gyroscope signals", in *Proc. 23rd usenix security symposium*, 2014 (cit. on p. 28).

[102] D. Johnson, M. M. Trivedi, *et al.*, "Driving style recognition using a smartphone as a sensor platform", in *Ieee international conference on intelligent transportation systems (itsc)*, 2011, pp. 1609–1615.

[103] H. Ye, T. Gu, X. Tao, and J. Lu, "B-Loc: Scalable floor localization using barometer on smartphone", in *Ieee international conference on mobile ad hoc and sensor systems (mass)*, 2014, pp. 127–135 (cit. on p. 28).

[104] S. Vanini and S. Giordano, "Adaptive context-agnostic floor transition detection on smart mobile devices", in *Ieee international conference on pervasive computing and communications workshops (percom workshops)*, 2013, pp. 2–7 (cit. on p. 28).

[105] K. Komeda, M. Mochizuki, and N. Nishiko, "User activity recognition method based on atmospheric pressure sensing", in *Proceedings of the acm international conference on pervasive and ubiquitous computing*, 2014, pp. 737–746 (cit. on p. 28).

[106] P. Castillejo, J.-F. Martínez, L. López, and G. Rubio, "An Internet of Things approach for managing smart services provided by wearable devices", *International journal of distributed sensor networks*, vol. 9, no. 2, p. 190 813, 2013. DOI: 10.1155/2013/190813 (cit. on p. 28).

[107] X. Lai, Q. Liu, X. Wei, W. Wang, G. Zhou, and G. Han, "A survey of body sensor networks", *Sensors*, vol. 13, no. 5, pp. 5406–5447, 2013. DOI: 10.3390/s130505406 (cit. on p. 29).

[108] M. Ghamari, B. Janko, R. S. Sherratt, W. Harwin, R. Piechockic, and C. Soltanpur, "A survey on wireless body area networks for eHealthcare systems in residential environments", *Sensors*, vol. 16, no. 6, 2016. DOI: 10.3390/s16060831 (cit. on p. 29).

[109] A. Filippeschi, N. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker, "Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion", *Sensors*, vol. 17, no. 6, 2017. DOI: 10.3390/s17061257 (cit. on p. 29).

[110] E. Estellés-Arolas and F. González-Ladrón-De-Guevara, "Towards an integrated crowdsourcing definition", *Journal of information science*, vol. 38, no. 2, pp. 189–200, 2012. DOI: 10.1177/0165551512437638 (cit. on p. 29).

[111] D. C. Brabham, *Crowdsourcing as a model for problem solving: An introduction and cases*, 2008 (cit. on p. 29).

[112] J. M. Leimeister, M. Huber, U. Bretschneider, and H. Krcmar, "Leveraging crowdsourcing: Activation-supporting components for IT-based ideas competition", *Journal of management information systems*, vol. 26, no. 1, pp. 197–224, 2009. DOI: 10.2753/MIS0742-1222260108 (cit. on p. 29).

[113] A. Kittur, E. H. Chi, and B. Suh, "Crowdsourcing user studies with Mechanical Turk", in *Proceedings of the acm conference on human factors in computing systems*, ser. SIGCHI, 2008, pp. 453–456, ISBN: 978-1-60558-011-1. DOI: 10.1145/1357054.1357127 (cit. on p. 29).

[114] J. Ross, L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson, "Who are the crowdworkers?: Shifting demographics in mechanical turk", in *Proceedings of the acm conference on human factors in computing systems*, ser. SIGCHI EA, 2010, pp. 2863–2872. DOI: 10.1145/1753846.1753873 (cit. on p. 29).

[115] 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, 2017, ISSN: 0733-8716. DOI: 10.1109/JSAC.2017.2695738 (cit. on p. 29).

[116] E. Kamar, S. Hacker, and E. Horvitz, "Combining human and machine intelligence in large-scale crowdsourcing", in *International conference on autonomous agents and multiagent systems*, ser. AAMAS, vol. 1, International Foundation for Autonomous Agents and Multiagent Systems, 2012, pp. 467–474 (cit. on p. 29).

[117] I. Khan, F. Belqasmi, R. Glitho, N. Crespi, M. Morrow, and P. Polakos, "Wireless sensor network virtualization: A survey", *Ieee communications surveys tutorials*, vol. 18, no. 1, pp. 553–576, 2016, ISSN: 1553-877X. DOI: 10.1109/COMST.2015.2412971 (cit. on p. 30).

[118] H. I. Kobo, A. M. Abu-Mahfouz, and G. P. Hancke, "A survey on software-defined wireless sensor networks: Challenges and design requirements", *Ieee access*, vol. 5, pp. 1872–1899, 2017, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2666200 (cit. on p. 30).

[119] I. Ali, A. Gani, I. Ahmedy, I. Yaqoob, S. Khan, and M. H. Anisi, "Data collection in smart communities using sensor cloud: Recent advances, taxonomy, and future research directions", *Ieee communications magazine*, vol. 56, no. 7, pp. 192–197, 2018 (cit. on p. 30).

[120] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, "A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities", *Ieee communications surveys tutorials*, vol. 21, no. 3, pp. 2419–2465, 2019 (cit. on p. 34).

[121] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, "EmotionSense: A mobile phones based adaptive platform for experimental social psychology research", in *Proceedings of the acm international conference on ubiquitous computing*, 2010, pp. 281–290 (cit. on p. 35).

[122] D. A. Galvan, A. Komjathy, M. P. Hickey, P. Stephens, J. Snively, Y Tony Song, M. D. Butala, and A. J. Mannucci, "Ionospheric signatures of tohoku-oki tsunami of march 11, 2011: Model comparisons near the epicenter", *Radio science*, vol. 47, no. 4, 2012 (cit. on pp. 35, 42).

[123] Y. Ishigaki, Y. Matsumoto, R. Ichimiya, and K. Tanaka, "Development of mobile radiation monitoring system utilizing smartphone and its field tests in fukushima", *Ieee sensors journal*, vol. 13, no. 10, pp. 3520–3526, 2013 (cit. on pp. 35, 42).

[124] P. Zhou, Y. Zheng, and M. Li, “How long to wait? predicting bus arrival time with mobile phone based participatory sensing”, in *Proc. of the acm international conference on mobile systems, applications, and services*, 2012, pp. 379–392 (cit. on pp. 35, 41, 43).

[125] L. Deng and L. P. Cox, “Livecompare: Grocery bargain hunting through participatory sensing”, in *Proc. of the acm workshop on mobile computing systems and applications*, 2009, p. 4 (cit. on pp. 35, 41, 43).

[126] N. Maisonneuve, M. Stevens, M. E. Niessen, and L. Steels, “NoiseTube: Measuring and mapping noise pollution with mobile phones”, in *Information technologies in environmental engineering*, Springer, 2009, pp. 215–228 (cit. on pp. 35, 36, 43).

[127] D. Yang, G. Xue, X. Fang, and J. Tang, “Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing”, in *Proceedings of the acm international conference on mobile computing and networking*, 2012, pp. 173–184 (cit. on p. 36).

[128] T. Luo, H. P. Tan, and L. Xia, “Profit-maximizing incentive for participatory sensing”, in *Ieee conference on computer communications (infocom)*, 2014, pp. 127–135. DOI: 10.1109/INFOCOM.2014.6847932 (cit. on p. 36).

[129] I. Koutsopoulos, “Optimal incentive-driven design of participatory sensing systems”, in *Ieee conference on computer communications (infocom)*, 2013, pp. 1402–1410. DOI: 10.1109/INFCOM.2013.6566934 (cit. on p. 36).

[130] 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, 2017, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2762238 (cit. on p. 36).

[131] I. Schweizer, R. Bärtl, A. Schulz, F. Probst, and M. Mühläuser, “Noisemap - real-time participatory noise maps”, in *International workshop on sensing applications on mobile phones*, 2011, pp. 1–5 (cit. on pp. 37, 41, 42).

[132] Y. F. Dong, S Kanhere, C. T. Chou, and N. Bulusu, “Automatic collection of fuel prices from a network of mobile cameras”, in *Distributed computing in sensor systems*, Springer, 2008, pp. 140–156 (cit. on pp. 37, 42).

[133] S. Dashti, J. D. Bray, J. Reilly, S. Glaser, A. Bayen, and E. Mari, “Evaluating the reliability of phones as seismic monitoring instruments”, *Earthquake spectra*, vol. 30, no. 2, pp. 721–742, 2014 (cit. on p. 38).

[134] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, “Real-time urban monitoring using cell phones: A case study in Rome”, *Ieee transactions on intelligent transportation systems*, vol. 12, no. 1, pp. 141–151, 2011, ISSN: 1524-9050. DOI: 10.1109/TITS.2010.2074196 (cit. on p. 38).

[135] L. Summers and S. D. Johnson, “Does the configuration of the street network influence where outdoor serious violence takes place? using space syntax to test crime pattern theory”, *Journal of quantitative criminology*, vol. 33, no. 2, pp. 397–420, 2017, ISSN: 1573-7799. DOI: 10.1007/s10940-016-9306-9 (cit. on p. 38).

[136] M. Marjanović, A. Antonić, and I. P. Žarko, "Edge computing architecture for mobile crowdsensing", *Ieee access*, vol. 6, pp. 10 662–10 674, 2018 (cit. on p. 38).

[137] P. Vitello, A. Capponi, C. Fiandrino, P. Giaccone, D. Kliazovich, U. Sorger, and P. Bouvry, "Collaborative data delivery for smart city-oriented mobile crowdsensing systems", in *Ieee global communications conference (globecom)*, 2018, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8648047 (cit. on p. 40).

[138] J. Weppner and P. Lukowicz, "Bluetooth based collaborative crowd density estimation with mobile phones", in *Ieee international conference on pervasive computing and communications (percom)*, 2013, pp. 193–200. DOI: 10.1109/PerCom.2013.6526732 (cit. on pp. 40, 41).

[139] H. To, L. Fan, L. Tran, and C. Shahabi, "Real-time task assignment in hyperlocal spatial crowdsourcing under budget constraints", in *Ieee international conference on pervasive computing and communications (percom)*, 2016, pp. 1–8. DOI: 10.1109/PERCOM.2016.7456507 (cit. on p. 40).

[140] *Waze: Get the best route, every day, with real–time help from other drivers.* https://www.waze.com/, 2006 (cit. on pp. 41, 42).

[141] M. Karaliopoulos, O. Telelis, and I. Koutsopoulos, "User recruitment for mobile crowdsensing over opportunistic networks", in *Ieee conference on computer communications (infocom)*, 2015, pp. 2254–2262. DOI: 10.1109/INFOCOM.2015.7218612 (cit. on p. 41).

[142] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones", in *Proceedings of the acm conference on embedded network sensor systems*, 2008, pp. 323–336 (cit. on p. 41).

[143] H. Habibzadeh, Z. Qin, T. Soyata, and B. Kantarci, "Large-scale distributed dedicated- and non-dedicated smart city sensing systems", *Ieee sensors journal*, vol. 17, no. 23, pp. 7649–7658, 2017, ISSN: 1530-437X. DOI: 10.1109/JSEN.2017.2725638 (cit. on p. 42).

[144] M. Budde, P. Barbera, R. El Masri, T. Riedel, and M. Beigl, "Retrofitting smartphones to be used as particulate matter dosimeters", in *Proc. of the acm international symposium on wearable computers*, ser. ISWC, 2013, pp. 139–140. DOI: 10.1145/2493988.2494342 (cit. on p. 42).

[145] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data", in *Aaai*, vol. 5, 2005, pp. 1541–1546 (cit. on p. 42).

[146] A. Bujari, B. Licar, and C. E. Palazzi, "Movement pattern recognition through smartphone's accelerometer", in *Ieee consumer communications and networking conference (ccnc)*, 2012, pp. 502–506 (cit. on p. 42).

[147] I. Shafer and M. L. Chang, "Movement detection for power-efficient smartphone WLAN localization", in *Proceedings of the acm international conference on modeling, analysis, and simulation of wireless and mobile systems*, 2010, pp. 81–90 (cit. on p. 42).

[148] S. Sehgal, S. S. Kanhere, and C. T. Chou, "Mobishop: Using mobile phones for sharing consumer pricing information", in *Demo session of the intl. conference on distributed computing in sensor systems*, 2008 (cit. on p. 43).

[149] V. Freschi, S. Delpriori, E. Lattanzi, and A. Bogliolo, “Bootstrap based uncertainty propagation for data quality estimation in crowdsensing systems”, *Ieee access*, vol. 5, pp. 1146–1155, 2017. DOI: 10.1109/ACCESS.2017.2651942 (cit. on p. 43).

[150] M. Tomasoni, A. Capponi, C. Fiandrino, D. Kliazovich, F. Granelli, and P. Bouvry, “Why energy matters? profiling energy consumption of mobile crowdsensing data collection frameworks”, *Pervasive and mobile computing*, vol. 51, pp. 193 –208, 2018, ISSN: 1574-1192. DOI: https://doi.org/10.1016/j.pmcj.2018.10.002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574119217305965 (cit. on p. 43).

[151] Y. Xiao, P. Simoens, P. Pillai, K. Ha, and M. Satyanarayanan, “Lowering the barriers to large-scale mobile crowdsensing”, in *Proc. of the acm workshop on mobile computing systems and applications*, ser. HotMobile, 2013, pp. 1–6, ISBN: 978-1-4503-1421-3. DOI: 10.1145/2444776.2444789 (cit. on p. 44).

[152] K. Farkas and I. Lendák, “Evaluation of simulation engines for crowdsensing activities”, in *International conference & workshop mechatronics in practice and education*, ser. MECHEDU, 2015, pp. 126–131 (cit. on p. 44).

[153] 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)*, 2017, pp. 1–6. DOI: 10.1109/ICC.2017.7996886 (cit. on p. 45).

[154] K. Mehdi, M. Lounis, A. Bounceur, and T. Kechadi, “CupCarbon: A multi-agent and discrete event wireless sensor network design and simulation tool”, in *International icst conference on simulation tools and techniques*, ser. SIMUTools, Lisbon, Portugal, 2014, pp. 126–131, ISBN: 978-1-63190-007-5 (cit. on p. 44).

[155] K. Farkas and I. Lendák, “Simulation environment for investigating crowd-sensing based urban parking”, in *International conference on models and technologies for intelligent transportation systems (mt-its)*, 2015, pp. 320–327. DOI: 10.1109/MTITS.2015.7223274 (cit. on p. 44).

[156] 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. DOI: 10.1016/j.compenvurbsys.2017.05.004 (cit. on pp. 46, 47).

[157] P. Vitello, A. Capponi, C. Fiandrino, P. Giaccone, D. Kliazovich, and P. Bouvry, “High-precision design of pedestrian mobility for smart city simulators”, in *Ieee international conference on communications (icc)*, 2018, pp. 1–6. DOI: 10.1109/ICC.2018.8422599 (cit. on p. 48).

[158] 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, 2017, ISSN: 2377-3782. DOI: 10.1109/TSUSC.2017.2666043 (cit. on pp. 53, 59).

[159] 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, ISSN: 1574-1192. DOI: `https://doi.org/10.1016/j.pmcj.2016.09.005` (cit. on pp. 48, 59).

[160] 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, 2017, ISSN: 2169-3536. DOI: `10.1109/ACCESS.2017.2671678`.

[161] M. R. Brust, M. Zurad, L. Hentges, L. Gomes, G. Danoy, and P. Bouvry, "Target tracking optimization of UAV swarms based on dual-pheromone clustering", in *3rd ieee international conference on cybernetics (cybconf)*, 2017, pp. 1–8 (cit. on p. 50).

[162] 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, no. Supplement C, pp. 100 –109, 2017, ISSN: 1389-1286. DOI: `https://doi.org/10.1016/j.comnet.2016.11.020` (cit. on p. 53).

[163] 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)*, 2014, pp. 1366–1374. DOI: `10.1109/INFOCOM.2014.6848070` (cit. on p. 53).

[164] C. Liu, B. Zhang, X. Su, J. Ma, W. Wang, and K. Leung, "Energy-aware participant selection for smartphone-enabled mobile crowd sensing", *Ieee systems journal*, vol. PP, no. 99, pp. 1–12, 2015, ISSN: 1932-8184. DOI: `10.1109/JSYST.2015.2430362`.

[165] W. Sherchan, P. P. Jayaraman, S. Krishnaswamy, A. Zaslavsky, S. Loke, and A. Sinha, "Using on-the-move mining for mobile crowdsensing", in *Ieee 13th international conference on mobile data management*, 2012, pp. 115–124.

[166] 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, 2015, ISSN: 2168-2216. DOI: `10.1109/TSMC.2015.2418283` (cit. on p. 53).

[167] J. Peng, Y. Zhu, Q. Zhao, H. Zhu, J. Cao, G. Xue, and B. Li, "Fair energy-efficient sensing task allocation in participatory sensing with smartphones", *The computer journal*, vol. 60, no. 6, pp. 850–865, 2017.

[168] F. Montori, L. Bedogni, and L. Bononi, "Distributed data collection control in opportunistic mobile crowdsensing", in *Proc. of the acm workshop on experiences with the design and implementation of smart objects*, ser. SMARTOBJECTS, Snowbird, Utah, USA, 2017, pp. 19–24, ISBN: 978-1-4503-5141-6. DOI: `10.1145/3127502.3127509` (cit. on p. 53).

[169] 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, Roma, Italy, 2013, pp. 1–14, ISBN: 978-1-4503-2027-6. DOI: 10.1145/2517351.2517372 (cit. on p. 53).

[170] R. Rouvoy, *APISENSE: Mobile crowd-sensing made easy!*, Keynote of ECAAS'16, Jun. 2016. [Online]. Available: https://hal.inria.fr/hal-01332594 (cit. on p. 55).

[171] F. Paganelli, S. Turchi, and D. Giuli, “A web of things framework for RESTful applications and its experimentation in a smart city”, *Ieee systems journal*, vol. 10, no. 4, pp. 1412–1423, 2016, ISSN: 1932-8184. DOI: 10.1109/JSYST.2014.2354835.

[172] R. Mulero, A. Almeida, G. Azkune, L. Mainetti, V. Mighali, L. Patrono, P. Rametta, and I. Sergi, “An AAL system based on IoT technologies and linked open data for elderly monitoring in smart cities”, in *2nd international multidisciplinary conference on computer and energy science (splitech)*, 2017, pp. 1–6.

[173] F. Kalim, J. P. Jeong, and M. U. Ilyas, “CRATER: A crowd sensing application to estimate road conditions”, *Ieee access*, vol. 4, pp. 8317–8326, 2016. DOI: 10.1109/ACCESS.2016.2607719 (cit. on p. 55).

[174] J. W. Sung and S. J. Han, “Data bundling for energy efficient communication of wearable devices”, in *Ieee 40th conference on local computer networks (lcn)*, 2015, pp. 321–328. DOI: 10.1109/LCN.2015.7366326 (cit. on p. 55).

[175] R. LiKamWa, Z. Wang, A. Carroll, F. X. Lin, and L. Zhong, “Draining our glass: An energy and heat characterization of google glass”, in *Proc. of 5th asia-pacific workshop on systems*, ser. APSys, Beijing, China: ACM, 2014, pp. 1–7, ISBN: 978-1-4503-3024-4. DOI: 10.1145/2637166.2637230 (cit. on p. 55).

[176] J. C. V. Bedregal, E. G. C. Gutierrez, and R. E. A. M., “Optimizing energy consumption per application in mobile devices”, in *International conference on information society*, 2013, pp. 106–110 (cit. on p. 56).

[177] G. P. Perrucci, F. H. P. Fitzek, and J. Widmer, “Survey on energy consumption entities on the smartphone platform”, in *Ieee 73rd vehicular technology conference (vtc spring)*, 2011, pp. 1–6. DOI: 10.1109/VETECS.2011.5956528 (cit. on p. 56).

[178] 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 (cit. on p. 57).

[179] K. Tan, J. Fang, Y. Zhang, S. Chen, L. Shi, J. Zhang, and Y. Zhang, “Fine-grained channel access in wireless lan”, in *Proc. of the acm sigcomm*, New Delhi, India: ACM, 2010, pp. 147–158, ISBN: 978-1-4503-0201-2. DOI: 10.1145/1851182.1851202 (cit. on p. 58).

[180] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems", *Corr*, vol. cs.NI/9809099, 1998 (cit. on p. 60).

[181] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications", *Acm transactions on intelligent systems and technology*, vol. 5, no. 3, pp. 1–55, 2014, ISSN: 2157-6904 (cit. on p. 66).

[182] A. Bucchiarone and A. Cicchetti, "A model-driven solution to support smart mobility planning", in *Proc. of acm/ieee models*, 2018, pp. 123–132.

[183] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira Jr, and C. Ratti, "Understanding individual mobility patterns from urban sensing data: A mobile phone trace example", *Transportation research part c: Emerging technologies*, vol. 26, pp. 301–313, 2013 (cit. on p. 67).

[184] X. Wang, Z. Zhou, Z. Yang, Y. Liu, and C. Peng, "Spatio-temporal analysis and prediction of cellular traffic in metropolis", in *Ieee 25th international conference on network protocols (icnp)*, 2017, pp. 1–10. DOI: 10.1109/ICNP.2017.8117559 (cit. on p. 67).

[185] Y. Zhou, B. P. L. Lau, C. Yuen, B. Tunçer, and E. Wilhelm, "Understand urban human mobility through crowdsensed data", *Corr*, vol. abs/1805.00628, 2018 (cit. on pp. 67, 77).

[186] C. Ratti, D. Frenchman, R. M. Pulselli, and S. Williams, "Mobile landscapes: Using location data from cell phones for urban analysis", *Environment and planning b: Planning and design*, vol. 33, no. 5, pp. 727–748, 2006. DOI: 10.1068/b32047 (cit. on p. 67).

[187] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in rome", *Ieee trans. on intelligent transportation systems*, vol. 12, no. 1, pp. 141–151, 2011, ISSN: 1524-9050. DOI: 10.1109/TITS.2010.2074196 (cit. on p. 67).

[188] D. Hristova, L. M. Aiello, and D. Quercia, "The new urban success: How culture pays", *Frontiers in physics*, vol. 6, p. 27, 2018, ISSN: 2296-424X. DOI: 10.3389/fphy.2018.00027.

[189] C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang, "The promises of big data and small data for travel behavior (aka human mobility) analysis", *Transportation research part c: Emerging technologies*, vol. 68, pp. 285–299, 2016.

[190] M. S. Kaiser, K. T. Lwin, M. Mahmud, D. Hajializadeh, T. Chaipimonplin, A. Sarhan, and M. A. Hossain, "Advances in crowd analysis for urban applications through urban event detection", *Ieee trans. on intelligent transportation systems*, pp. 1–21, 2017, ISSN: 1524-9050. DOI: 10.1109/TITS.2017.2771746 (cit. on p. 67).

[191] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey", *Corr*, vol. abs/1803.04311, 2018. arXiv: 1803.04311. [Online]. Available: http://arxiv.org/abs/1803.04311 (cit. on p. 67).

[192] Z. Liu, Z. Li, K. Wu, and M. Li, "Urban traffic prediction from mobility data using deep learning", *Ieee network*, vol. 32, no. 4, pp. 40–46, 2018, ISSN: 0890-8044. DOI: 10.1109/MNET.2018.1700411 (cit. on p. 67).

[193] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, "Traffic flow prediction with big data: A deep learning approach", *Ieee transactions on intelligent transportation systems*, vol. 16, no. 2, pp. 865–873, 2015, ISSN: 1524-9050. DOI: 10.1109/TITS.2014.2345663 (cit. on p. 67).

[194] A. Dubey, N. Naik, D. Parikh, R. Raskar, and C. A. Hidalgo, "Deep learning the city: Quantifying urban perception at a global scale", in *Computer vision – eccv*, Springer International Publishing, 2016, pp. 196–212, ISBN: 978-3-319-46448-0.

[195] K. D'Silva, A. Noulas, M. Musolesi, C. Mascolo, and M. Sklar, "Predicting the temporal activity patterns of new venues", *Epj data science*, vol. 7, no. 1, p. 13, 2018 (cit. on pp. 67, 69).

[196] F. Giust, G. Verin, K. Antevski, J. Chou, Y. Fang, W. Featherstone, and et al., *MEC deployments in 4G and evolution towards 5G*, ETSI White Paper, 2018 (cit. on p. 76).

[197] M. Satyanarayanan, "The emergence of edge computing", *Computer*, vol. 50, no. 1, pp. 30–39, 2017, ISSN: 0018-9162. DOI: 10.1109/MC.2017.9 (cit. on p. 76).

[198] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, "A comprehensive survey on fog computing: State-of-the-art and research challenges", *Ieee communications surveys tutorials*, vol. 20, no. 1, pp. 416–464, 2018, ISSN: 1553-877X. DOI: 10.1109/COMST.2017.2771153 (cit. on p. 76).

[199] J. Gedeon, J. Krisztinkovics, C. Meurisch, M. Stein, L. Wang, and M. Mühlhäuser, "A multi-cloudlet infrastructure for future smart cities: An empirical study", in *Proc. of acm edgesys*, Munich, Germany, 2018, pp. 19–24, ISBN: 978-1-4503-5837-8. DOI: 10.1145/3213344.3213348 (cit. on pp. 76, 78).

[200] C. Marquez, M. Gramaglia, M. Fiore, A. Banchs, C. Ziemlicki, and Z. Smoreda, "Not all apps are created equal: Analysis of spatiotemporal heterogeneity in nationwide mobile service usage", in *Proc. of acm conext*, 2017, pp. 180–186 (cit. on p. 76).

[201] P. Liu, D. Willis, and S. Banerjee, "ParaDrop: Enabling lightweight multi-tenancy at the network's extreme edge", in *Proc. of ieee/acm sec*, 2016, pp. 1–13. DOI: 10.1109/SEC.2016.39 (cit. on p. 76).

[202] M. C. Filippou, D. Sabella, and V. Riccobene, "Flexible MEC service consumption through edge host zoning in 5G networks", *Corr*, 2019 (cit. on pp. 76, 78).

[203] R. Morabito, V. Cozzolino, A. Y. Ding, N. Beijar, and J. Ott, "Consolidate IoT edge computing with lightweight virtualization", *Ieee network*, vol. 32, no. 1, pp. 102–111, 2018, ISSN: 0890-8044. DOI: 10.1109/MNET.2018.1700175 (cit. on p. 76).

[204] P. Bellavista, S. Chessa, L. Foschini, L. Gioia, and M. Girolami, "Human-enabled edge computing: Exploiting the crowd as a dynamic extension of mobile edge computing", *Ieee communications magazine*, vol. 56, no. 1, pp. 145–155, 2018, ISSN: 0163-6804. DOI: 10.1109/MCOM.2017.1700385 (cit. on p. 77).

[205] M. Jia, J. Cao, and W. Liang, "Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks", *Ieee transactions on cloud computing*, vol. 5, no. 4, pp. 725–737, 2017, ISSN: 2168-7161. DOI: 10.1109/TCC.2015.2449834 (cit. on pp. 77, 78, 80).

[206] K. Jayarajah, A. Tan, and A. Misra, "Understanding the interdependency of land use and mobility for urban planning", in *Proc. of acm ubicomp*, Singapore, Singapore, 2018, pp. 1079–1087, ISBN: 978-1-4503-5966-5. DOI: 10.1145/3267305.3274163 (cit. on p. 77).

[207] J. Liang, Z. Chen, C. Li, B. Xia, and N. Liu, "Delay outage probability of multi-relay selection for mobile relay edge computing systems", *Corr*, 2019 (cit. on p. 78).

[208] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5G mobile crosshaul networks", *Ieee communications surveys tutorials*, vol. 21, no. 1, pp. 146–172, 2019, ISSN: 1553-877X. DOI: 10.1109/COMST.2018.2868805 (cit. on p. 78).

[209] L. Kleinrock, *Queueing systems, volume 1: Theory*. Hoboken, NJ, US, 1975 (cit. on p. 79).

[210] E. Ozer and M. Q. Feng, "Direction-sensitive smart monitoring of structures using heterogeneous smartphone sensor data and coordinate system transformation", *Smart materials and structures*, vol. 26, no. 4, p. 045 026, 2017.

[211] V. M. Souza, R. Giusti, and A. J. Batista, "Asfault: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning", *Pervasive and mobile computing*, vol. 51, pp. 121 –137, 2018, ISSN: 1574-1192. DOI: https://doi.org/10.1016/j.pmcj.2018.10.008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574119218300518 (cit. on p. 83).

[212] J. Reilly, S. Dashti, M. Ervasti, J. D. Bray, S. D. Glaser, and A. M. Bayen, "Mobile phones as seismologic sensors: Automating data extraction for the ishake system", *Ieee transactions on automation science and engineering*, vol. 10, no. 2, pp. 242–251, 2013, ISSN: 1545-5955. DOI: 10.1109/TASE.2013.2245121 (cit. on p. 83).

[213] M. Faulkner, R. Clayton, T. Heaton, K. M. Chandy, M. Kohler, J. Bunn, R. Guy, A. Liu, M. Olson, M. Cheng, and A. Krause, "Community sense and response systems: Your phone as quake detector", *Acm commun.*, vol. 57, no. 7, pp. 66–75, 2014. DOI: 10.1145/2622633 (cit. on p. 83).

[214] T. Ludwig, C. Reuter, T. Siebigteroth, and V. Pipek, "CrowdMonitor: Mobile crowd sensing for assessing physical and digital activities of citizens during emergencies", in *Proc. of the acm conference on human factors in computing systems*, ser. SIGCHI, 2015, pp. 4083–4092. DOI: 10.1145/2702123.2702265 (cit. on p. 83).