Towards efficient mobile crowdsensing assignment and uploading schemes

Rim Ben Messaoud

To cite this version:

Rim Ben Messaoud. Towards efficient mobile crowdsensing assignment and uploading schemes. Mobile Computing. Université Paris-Est, 2017. English. NNT : 2017PESC1031 . tel-01753150

HAL Id: tel-01753150
https://pastel.hal.science/tel-01753150
Submitted on 29 Mar 2018

UNIVERSITÉ PARIS-EST

École Doctorale MSTIC

Mathématiques et Sciences et Technologies de l'Information et de la Communication

Thèse de Doctorat

Discipline: Informatique

Présentée par

Rim Ben Messaoud

Towards Efficient Mobile Crowdsensing Assignment and Uploading Schemes

---

Dirigée par

Yacine Ghamri-Doudane

Soutenance le 05 Juillet 2017 devant le jury composé de:

Mme. Nathalie Mitton	Chargé de Recherche, INRIA Lille	Rapporteur
M. Hervé Rivano	Chargé de Recherche, INRIA Lyon	Rapporteur
M. Marco Fiore	Chargé de Recherche, CNR IEIIT	Examinateur
M. Patrice Raveneau	Maître de Conférences, Université de La Rochelle	Examinateur
M. Rami Langar	Professeur, Université Paris Est	Examinateur
M. Sidi Mohamed Senouci	Professeur, Université de Bourgogne	Examinateur
M. Toufik Ahmed	Professeur, ENSEIRB-MATMECA (Bordeaux INP)	Examinateur
M. Yacine Ghamri-Doudane	Professeur, Université de La Rochelle	Directeur de thèse

To My Father

To My Mother

Acknowledgments

My Ph.D journey was a challenging and enjoyable adventure that could not be achieved without the support of many, to whom I want to give my thanks.

Firstly, I would like to express my sincere gratitude to my advisor, Yacine Ghamri-Doudane, for giving me the opportunity to join his team and work on a very interesting topic. I constantly benefit from his continuous support, confidence and guidance throughout these four years of research allowing me to grow as a research scientist. His valuable advice are priceless for both my research and on my career choices.

I would like to express my special appreciation and thanks to my reading committee, Mrs. Nathalie Mitton and Mr. Hervé Rivano, for accepting to review this dissertation. Their insightful comments and encouragement have helped me to widen my research from various perspectives. Special thanks goes also to Mr. Marco Fiore, Mr. Patrice Raveneau, Mr. Rami Langar, Mr. Sidi Mohamed Senouci and Mr.Toufik Ahmed, for generously offering their time and good will to serve as my dissertation committee.

I thank all my labmates in both, LIGM Lab and L3I Lab, specially, Safa, Camila, Mohamed Ali, Sabrine, Dimitrios, Wafa, Nouredine and Nour, for the stimulating atmosphere and for all the supportive and fruitful discussions we have had during these last four years.

I am indebted to my friends in La Rochelle for providing support and friendship that I needed during the hardest moments as well as all the fun we had together. In particular, I am grateful to Nibal, Nouha, Jana and Aymen. You have always been a second family who believed in me and encouraged my lifetime academic project.

Last but not least, words cannot express how grateful I am to my family; my mother, my sisters: Rihab, Amel and Imene, and my brother-in-law, Meher, for all of the sacrifices that you have made on my behalf. Your prayers for me was what sustained me thus far. My deep appreciation goes to my sister and friend Imene for her lovely support since the first time we met. My heartfelt thanks goes to my best friend Houweida who have stood by me during all my Ph.D ups and downs from overseas and who have spared no effort or time to help me achieve my goal.

Abstract

The ubiquity of sensors-equipped mobile devices has enabled people to contribute data via crowdsensing systems. This emergent paradigm comes with various applications. However, new challenges arise given users involvement in data collection process. In this context, we introduce collaborative sensing schemes which tackle four main questions: *How to assign sensing tasks to maximize data quality with energy-awareness? How to minimize the processing time of sensing tasks? How to motivate users to dedicate part of their resources to the crowdsensing process?* and *How to protect participants privacy and not impact data utility when reporting collected sensory data?*

First, we focus on the fact that smart devices are energy-constrained and develop task assignment methods that aim to maximize sensor data quality while minimizing the overall energy consumption of the data harvesting process. The resulting contribution materialized as a Quality and Energy-aware Mobile Sensing Scheme (QEMSS) defines first data quality metrics then models and solves the corresponding optimization problem using a Tabu-Search based heuristic. Moreover, we assess the fairness of the resulted scheduling by introducing F-QEMSS variant. Through extensive simulations, we show that both solutions have achieved competitive data quality levels when compared to concurrent methods especially in situations where the process is facing low dense sensing areas and resources shortcomings. As a second contribution, we propose to distribute the assignment process among participants to minimize the average sensing time and processing overload compared to a fully centralized approach. Thus, we suggest to designate some participants to carry extra sensing tasks and delegate them to appropriate neighbors. The new assignment is based on predicting users local mobility and sensing preferences. Accordingly, we develop two new greedy-based assignment schemes, one only Mobility-aware (MATA) and the other one accounting for both preferences and mobility (P-MATA), and evaluate their performances. Both MATA and P-MATA consider a *voluntary* sensing process and show that accounting for users preferences minimize the sensing time. Having showing that, our third contribution in this thesis is conceived as an Incentives-based variant, IP-MATA$^+$. IP-MATA$^+$ incorporates rewards in the users choice model and proves their positive impact on enhancing their commitment especially when the dedicated budget is *shared* function of contributed data quality. Finally, our fourth and last contribution addresses the seizing of users privacy concerns within crowdsensing systems. More specifically, we study the minimization of the incurred privacy leakage in data uploading phase while accounting for the possible quality regression. That is, we assess simultaneously the two competing goals of ensuring queriers required data utility and protecting participants' sensitive information. Thus, we introduce a trust entity to the crowdsensing traditional system. This entity runs a general privacy-preserving mechanism to release a distorted version of sensed data that responds to a privacy-utility trade-off. The proposed mechanism, called PRUM, is evaluated on three sensing datasets, different adversary models and two main data uploading scenarios. Results show that a limited distortion on collected data may ensure privacy while maintaining about 98% of the required utility level.

The four contributions of this thesis tackle competing issues in crowdsensing which paves the way at facilitating its real implementation and aims at broader deployment.

Keywords: Mobile crowdsensing, Energy consumption, Quality of information, Users mobility, Sensing preferences, Task allocation, Incentives, Privacy.

Résumé

L'ubiquité des terminaux intelligents équipés de capteurs a donné naissance à un nouveau paradigme de collecte participative des données appelé *Crowdsensing*. Pour mener à bien les tâches de collecte, divers défis relatifs à l'implication des participants et des demandeurs de services doivent être relevés. Dans ce contexte, nous abordons quatre questions majeures inhérentes à ce problème: *Comment affecter les tâches de collecte afin de maximiser la qualité des données d'une façon éco-énergétique? Comment minimiser le temps nécessaire à la collecte et au traitement des tâches? Comment inciter les participants à dédier une partie de leurs ressources pour la collecte?* et *Comment protéger la vie privée des participants tout en préservant la qualité des données reportées?*

Tout d'abord, nous nous intéressons au fait que les ressources énergétiques des terminaux mobiles restent limitées. Nous introduisons alors des modèles de déploiement de tâches qui visent à maximiser la qualité des données reportées tout en minimisant le coût énergétique global de la collecte. Ainsi, notre première contribution se matérialise en un modèle d'allocation appelé, QEMSS. QEMSS définit des métriques de qualité de données et cherche à les maximiser en se basant sur des heuristiques utilisant la recherche taboue. De plus, afin de rendre le processus d'allocation résultante plus équitable, nous faisons appel à un deuxième algorithme, F-QEMSS, extension de QEMSS. Les deux solutions ont permis d'obtenir des niveaux de qualité de données compétitifs principalement dans les situations défavorables des zones de faible densité ou de ressources limitées. En outre, afin de minimiser le temps moyen de collecte et de traitement des données, une deuxième phase d'allocation distribuée est ajoutée. Plus précisement, nous proposons dans cette deuxième contribution de désigner des participants responsables de déléguer des tâches. Ces derniers prédisent le comportement d'autres utilisateurs en termes de mobilité et de préférences de collecte. Par conséquent, nous développons deux types d'allocation; MATA qui ne tient compte que de la mobilité et P-MATA qui tient compte à la fois de la mobilité et des préférences des participants. Les deux allocations démontrent que l'estimation des préférences des utilisateurs minimise le temps de collecte et évite le rejet des tâches. La troisième contribution de cette thèse, IP-MATA$^+$, propose des *incitations* aux participants, ce qui favorise leur engagement aux campagnes de collecte notamment quand le budget dédié est partagé en fonction de la qualité des contributions. Pour finir, nous considérons la problématique de la vie privée des participants au crowdsensing. Particulièrement, nous ciblons la minimisation du risque de divulgation de la vie privée durant la phase du déchargement tout en veillant à l'utilité des données collectées. Ainsi, la quatrième contribution de cette thèse vise à assurer simultanément deux objectifs concurrents, à savoir assurer l'utilité des données nécessaire aux demandeurs et protéger les informations *sensibles* des participants. Pour ce faire, nous introduisons une entité de confiance dans le système de collecte ayant pour rôle d'exécuter un mécanisme qui génère une version altérée de la donnée collectée qui répond au compromis de protection et d'utilité. La solution développée, appelée PRUM, a été évaluée sur des datasets de collecte participative en variant les scénarios d'attaque et de déchargement des données. Les résultats obtenus prouvent qu'une altération limitée des données collectées peut assurer une protection des informations sensibles des participants tout en préservant environ 98% de l'utilité des données obtenue pour les demandeurs.

Pour conclure, nos contributions abordent diverses problématiques complémentaires inhérentes à la collecte participative des données ouvrant la voie à des mises en œuvre réelles et facilitant leur déploiement.

Mots clés: Collecte participative, énergie, qualité d'information, mobilité, préférences, allocation des tâches, incitations, vie privée.

Contents

1 Introduction **1**
1.1 What is Mobile Crowdsensing? 1
1.2 Motivations and Contributions 3
1.3 Organization of the Thesis 5

2 Crowdsensing: Opportunities and Challenges **7**
2.1 Introduction 7
2.2 Mobile Sensing: Scale and Paradigms 8
  2.2.1 Mobile Sensing Scale 8
  2.2.2 Crowdsensing Paradigms 8
    2.2.2.1 Participatory Sensing 8
    2.2.2.2 Opportunistic Sensing 9
2.3 Crowdsensing Opportunities 9
  2.3.1 MCS Applications in Academia 9
    2.3.1.1 Environmental Applications 9
    2.3.1.2 Infrastructure Applications 10
    2.3.1.3 Social Applications 10
  2.3.2 MCS Applications in Industry 11
  2.3.3 MCS Frameworks 11
2.4 Crowdsensing Challenges: Participants Concerns 12
  2.4.1 Energy Consumption 13
    2.4.1.1 Energy Consumption in Sensing 13
    2.4.1.2 Energy Consumption in Processing 14
    2.4.1.3 Energy Consumption in Uploading 14
  2.4.2 Mobile Data Cost 15
  2.4.3 Incentives 15
  2.4.4 Privacy 16
2.5 Crowdsensing Challenges: Requesters Concerns 17
  2.5.1 Quality of Information (QoI) 17
  2.5.2 Budget 18
2.6 Positioning of Dissertation Contributions to the studied Literature 19
2.7 Conclusion 20

3 Quality and Energy-aware Task Assignment Schemes **21**
3.1 Introduction 21
3.2 Related Works 22
  3.2.1 Energy-aware Task Assignment in MCS 23

3.2.2 Quality-aware Task Assignment in MCS 23
3.2.3 Fairness in MCS 23
3.3 Background 24
3.3.1 Proposed Utility Function 24
3.3.2 Tabu Search 25
3.4 Preliminaries: Proposed Measures 26
3.4.1 Quality of Information Attributes and Metrics 26
3.4.1.1 Completeness 26
3.4.1.2 Timeliness 26
3.4.1.3 Energy Metric: Affordability 27
3.4.2 Fairness Metric: Jain Index 28
3.5 Quality and Energy-aware Problem Definition 28
3.5.1 Crowdsensing System Overview 28
3.5.2 Known/Unknown Users Trajectories 29
3.5.3 Problem Formulation 30
3.5.3.1 QoI Maximization Problem 30
3.5.3.2 QoI and Fairness Maximization Problem 31
3.6 Mobile Sensing Schemes: QEMSS Vs F-QEMSS 32
3.6.1 Elements of TS-based Schemes 32
3.6.1.1 Solution $X \in \Omega$ 32
3.6.1.2 Move m 33
3.6.1.3 Tabu List (TL) 33
3.6.2 Phases of TS-based Schemes 33
3.6.2.1 Initialization 33
3.6.2.2 Neighborhood Formation 33
3.6.2.3 Neighborhood Selection 34
3.6.2.4 Update Tabu List 34
3.7 Performance Evaluation 35
3.7.1 Simulation Settings 35
3.7.2 Benchmark 36
3.7.3 Evaluation Metrics 36
3.7.4 Evaluation Results 37
3.7.4.1 Maximum Achieved QoI 37
3.7.4.2 Spatial and Temporal Accuracy 39
3.7.4.3 Fairness Metrics 40
3.8 Discussion 42

4 Preference and Mobility-aware Task Assignment Schemes 45
4.1 Introduction 45
4.2 Motivations and Context 46
4.3 Existing Work on Distributed Crowdsensing 48
4.4 Distributed Task Assignment Problem Definition 49
4.4.1 Semi-Distributed Crowdsensing System Overview 49
4.4.2 Users' Arrival Model 49
4.4.3 Users' Preferences Model 50
4.4.4 Problem Definition 51
4.5 Mobility-aware Task Assignment: MATA 52

- 4.5.1 Offline Mode . . . 52
  - 4.5.1.1 Offline Average Makespan . . . 53
  - 4.5.1.2 Offline Algorithm: MATAF . . . 53
- 4.5.2 Online Mode . . . 54
  - 4.5.2.1 Online Average Makespan . . . 54
  - 4.5.2.2 Online Algorithm: MATAN . . . 55
- 4.5.3 After Thoughts . . . 55
- 4.6 Preference and Mobility-aware Task Assignment: P-MATA . . . 56
  - 4.6.1 Offline Mode . . . 57
    - 4.6.1.1 Offline Average Makespan . . . 57
    - 4.6.1.2 Offline Algorithm: P-MATAF . . . 57
  - 4.6.2 Online Solution: P-MATAN . . . 58
  - 4.6.3 After Thoughts . . . 60
- 4.7 Performance Evaluation . . . 60
  - 4.7.1 Real Traces . . . 60
  - 4.7.2 Requester Selection . . . 61
  - 4.7.3 Simulation Settings . . . 61
  - 4.7.4 Evaluation Results . . . 62
    - 4.7.4.1 Average Number of Assigned Tasks . . . 62
    - 4.7.4.2 Average Achieved Makespan . . . 63
    - 4.7.4.3 The Number of Lost Tasks . . . 65
- 4.8 Discussion . . . 66

5 Extended Preference-aware Task Assignment with Incentive Mechanisms 69
- 5.1 Introduction . . . 69
- 5.2 Motivations and Context . . . 70
- 5.3 Preference-aware and Incentivizing MCS . . . 71
  - 5.3.1 Preference-aware Crowdsensing . . . 71
  - 5.3.2 Incentivizing Mechanisms in MCS . . . 72
- 5.4 Problem Statement . . . 73
  - 5.4.1 Preliminaries . . . 73
    - 5.4.1.1 System Overview . . . 73
    - 5.4.1.2 Users Probabilistic Arrival . . . 73
    - 5.4.1.3 Discrete Choice Model . . . 74
  - 5.4.2 Problem Formulation . . . 75
    - 5.4.2.1 No-incentives-based Assignment . . . 75
    - 5.4.2.2 Incentives-based Assignment . . . 75
- 5.5 Extended Preference-aware Task Assignment: P-MATA$^+$ . . . 76
  - 5.5.1 Offline Mode: P-MATAF$^+$ . . . 76
  - 5.5.2 Online Mode: P-MATAN$^+$ . . . 77
  - 5.5.3 After Thoughts . . . 77
- 5.6 Incentives-based Preference-aware Task Assignment: IP-MATA$^+$ . . . 78
  - 5.6.1 Incentive Policies . . . 78
    - 5.6.1.1 Priority-based Incentives . . . 78
    - 5.6.1.2 Quality-based Incentives . . . 78
  - 5.6.2 Offline Mode: IP-MATAF$^+$ . . . 79
  - 5.6.3 Online Mode: IP-MATAN$^+$ . . . 80

    - 5.6.4 After Thoughts . . . 81
  - 5.7 Performance Evaluation . . . 82
    - 5.7.1 Simulation Settings . . . 82
    - 5.7.2 Performance Analysis . . . 82
      - 5.7.2.1 Average Number of Assigned Tasks . . . 83
      - 5.7.2.2 Average Makespan . . . 84
      - 5.7.2.3 Incentives Policies Performance . . . 85
  - 5.8 Discussion . . . 86

6 Privacy Preserving Utility-aware Mechanism in Data Uploading 89

- 6.1 Introduction . . . 89
- 6.2 Motivations and Context . . . 90
- 6.3 Research Work on Privacy-preserving Mechanisms . . . 92
  - 6.3.1 Privacy-preserving Mechanisms in Participatory Sensing . . . 92
    - 6.3.1.1 Cryptography . . . 92
    - 6.3.1.2 Anonymization Techniques . . . 92
    - 6.3.1.3 Data Aggregation . . . 93
    - 6.3.1.4 Obfuscation/ Perturbation . . . 93
  - 6.3.2 Privacy-utility Trade-off . . . 93
- 6.4 System Model . . . 94
  - 6.4.1 Participatory Sensing Involved Entities . . . 94
    - 6.4.1.1 Participants . . . 94
    - 6.4.1.2 Queriers/Requesters . . . 94
    - 6.4.1.3 Server/Broker . . . 94
  - 6.4.2 Adversary Model . . . 95
  - 6.4.3 Scenarios of Data Reporting Model . . . 95
- 6.5 Problem Formulation . . . 96
  - 6.5.1 Privacy Leakage Metric . . . 96
  - 6.5.2 Data Utility Metric . . . 97
  - 6.5.3 Optimization Problem . . . 98
    - 6.5.3.1 Adversary without Side Information . . . 99
    - 6.5.3.2 Adversary with Side Information . . . 100
- 6.6 Proposed Solution: PRUM . . . 102
  - 6.6.1 Proposed Algorithm . . . 102
  - 6.6.2 Mapping of Distorted Data . . . 104
  - 6.6.3 Data with Large Size Alphabets . . . 104
- 6.7 Performance Evaluation . . . 104
  - 6.7.1 Sensing Datasets . . . 105
    - 6.7.1.1 Occupancy Detection Data . . . 105
    - 6.7.1.2 GPS Trajectory Data . . . 105
    - 6.7.1.3 Crowd Temperature Data . . . 105
  - 6.7.2 Simulations Settings . . . 105
  - 6.7.3 Privacy-Utility Trade-off . . . 106
    - 6.7.3.1 Smart-house Monitoring Scenario . . . 106
    - 6.7.3.2 Traffic Rating Application . . . 107
    - 6.7.3.3 Crowd-Temperature Application . . . 109
  - 6.7.4 Side Information Impact on Privacy Leakage . . . 109

    - 6.7.4.1 Smart-house Monitoring Scenario 109
    - 6.7.4.2 Traffic Rating Application 110
- 6.8 Discussion 112

7 Conclusion and Perspectives **113**
- 7.1 Contributions 113
- 7.2 Perspectives 114

Bibliography **118**

List of Figures

1.1 Mobile crowdsensing phases and involved entities 2

2.1 Example of traffic monitoring crowdsensing scanario 10
2.2 Architecture of the crowdsensing platform: Sensarena 12
2.3 Relation between our contributions and MCS challenges 20

3.1 Mobile crowdsensing centralized architecture 29
3.2 Achieved QoI by different selection schemes 37
3.3 Average achieved QoI by F-QEMSS, QEMSS, DPS and ES schemes 38
3.4 Spatial and temporal data accuracy levels 39
3.5 Fairness measures for different selection schemes 40
3.6 Sensing times for a 500-users density 41

4.1 Distributed assignment in mobile crowdsensing 47
4.2 P-MATAN illustrative example 59
4.3 Assigned tasks by each scheme while varying the number of tasks 63
4.4 CDF of assigned tasks with $m = 20$ for each scheme and trace 64
4.5 Average achieved makespan 65
4.6 CDF of lost tasks with $m = 20$ and $\tau = 1$h 66

5.1 Participant choice model 81
5.2 The average number of assigned tasks by different requesters 83
5.3 The cdf of assigned tasks for each requesters selection policy 83
5.4 Average achieved makespan 84
5.5 Expenditure efficiency of makespan over spent budget for both traces 85
5.6 Budget evolution function of time for both traces 86

6.1 Participatory system with trusted entity 91
6.2 The different scenarios of data reporting and obfuscation 96
6.3 Privacy-Utility plots for Occupancy Data 107
6.4 Privacy-Utility plots for GPS Data 108
6.5 Privacy-Utility for temperature crowdsensing app 109
6.6 Privacy leakage with side information for Occupancy Data 110
6.7 Computation time with exchange Vs erasure distortion techniques 111
6.8 Privacy leakage with side information for Traffic Rating 111

7.1 Évaluation de QEMSS et F-QEMSS en termes de QoI et d'équité 133
7.2 Le temps moyen de collecte réalisé par MATA Vs P-MATA 135
7.3 Le temps moyen de collecte réalisé par P-MATA$^+$ Vs IP-MATA$^+$ 136

7.4 Évaluation de PRUM sur des données de capteurs domestiques 139

Chapter 1

Introduction

Contents

1.1 What is Mobile Crowdsensing? 1
1.2 Motivations and Contributions 3
1.3 Organization of the Thesis 5

1.1 What is Mobile Crowdsensing?

Smartphones are not only mobile phones with advanced computing capabilities and connectivity. They are also equipped with numerous embedded sensors such as accelerometer, GPS, gyroscope, camera and microphone [1] to name a few. This has enhanced their popularity as statistics envision that over a third of the world's population is projected to own a smartphone by 2017 [2], i.e., an estimated total of almost 2.6 billion users in the world. This makes them an important enabler for the future Internet of Things (IoT).

The envisioned tremendous number of smartphones coupled with the proliferation of other smart devices leverages the power of the general public (crowd) to sense, collect and share measurements about different phenomena. Accordingly, Ganti et al. [3] coined the term Mobile CrowdSensing (MCS) to refer to this new paradigm of sensing in a community scale. This emergent sensing model has attracted a lot of attention from both academia and industry given its significant advantages when compared to traditional sensing networks such as Wireless Sensors Networks (WSNs). First, MCS advances the cost limitations of statically deployed sensors and gathers users' data from places not economically feasible before such as in road traffic congestion control applications [4–6]. Moreover, MCS solves the coverage range issue and offers a much broader one given participants mobility. As a consequence, various MCS applications, conducted *opportunistically* [1] or in a *participatory* [7] way, have been introduced ranging from tracking personal activities [8, 9] to monitoring urban infrastructure [4–6, 10] and environment [11–15].

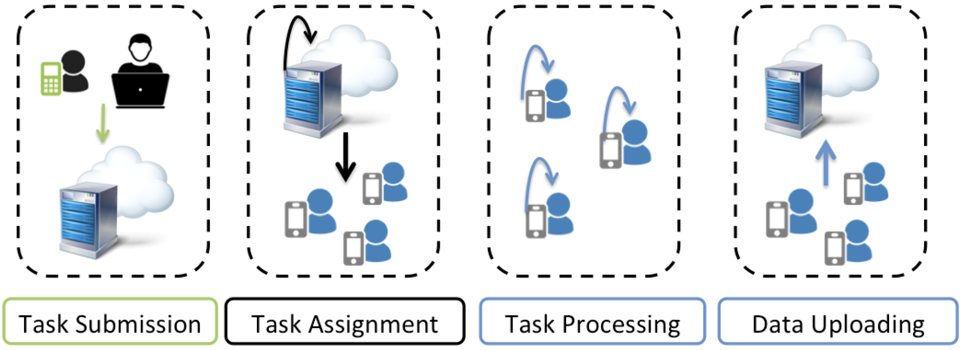


Figure 1.1: Mobile crowdsensing phases and involved entities

In order to do so, mobile crowdsensing systems are usually built around three main entities:

- **Requesters**: users who submit sensing *requests* described also as sensing **tasks** while precising the targeted phenomena and its location among other parameters.
- **MCS platform**: a central entity, usually a server in the cloud, which organizes sensing tasks and delegates them to appropriate users to be performed.
- **Participants**: users who receive requests for sensing tasks and *participate* in the sensing process by devoting their smart devices resources to collect required samples.

Furthermore, we can distinguish four major phases in mobile crowdsensing; *Task Submission*, *Task Assignment*, *Sensing and Processing*, and finally *Data Uploading* as illustrated in Figure 1.1. The former phase is usually initiated by requesters looking for data about a certain event or phenomena. Requesters submit their requirements to the MCS platform and wait for data which can be immediately displayed if already collected. Else, a corresponding sensing task is created in the MCS platform in order to look for participants to handle it. To this purpose, the *Task Assignment* phase is launched to select users who contribute data and respond to sensing requests. Nevertheless, this phase is extremely delicate in terms of how to recruit participants to provide satisfactory sensor data while cutting off the sensing process cost. This cost is usually incurred when participants are executing tasks during the *Sensing and Processing* phase. Indeed, crowdsensing requires users to dedicate their energy-constrained devices, their time and even their human intelligence for some tasks to process the different sensing tasks. Finally, the collected and processed data needs to be reported to the MCS platform in a *Data Uploading* step. Here also, participants may allocate a portion of their mobile data connectivity budget to upload collected samples unless being within a free wireless communication range.

By observing the aforementioned steps, researchers claim that such promising paradigm raises also new challenges [1] given its high dependency on the human factor. As stated, participants dedicate their energetic resources to perform sensing campaigns and may be actively involved in certain tasks such as taking photos or recording videos. Therefore, they are usually recruited on the basis of rewards, defined as “incentives”. Moreover, in addition to their *unwillingness* to allocate mobile resources, data uploading in a large scale augments users privacy concerns. At the same time, requesters are exaggerating their needs in terms of high data quality contributions. Answering both requesters and participants requirements is the main challenge which may hinder crowdsensing real implementation. In this dissertation, we focus on the limitation of the existing research work on mobile crowdsensing and aim to develop adequate task assignment, processing and uploading schemes as explained in the next section.

1.2 Motivations and Contributions

In this dissertation, we aim to address the aforementioned concerns affecting both participants and requesters in MCS campaigns. Particularly, we introduce through the four main contributions of this thesis scheduling algorithms that organize MCS phases while tackling the following questions: *How to assign sensing tasks to maximize data quality with energy-awareness? How to minimize the processing time of sensing tasks? How to motivate users to dedicate part of their resources to the crowdsensing process?* and *How to protect participants' privacy and not impact data utility when reporting collected data?*

Our first contribution focuses on the task assignment phase of mobile crowdsensing. We adopt a centralized architecture of a server coordinating sensing task allocation among registered participants. This entity aims to respond to our first question: *How to assign sensing tasks to maximize data quality with energy-awareness?* Therefore, we first identify the quality attributes of MCS data, denoted as Quality of Information (QoI), and the dedicated resources for its acquisition. Accordingly, for a given set of participants, a sensing area and data quality requirements set by requesters, we look for the subset of users that maximizes QoI in terms of spatial and temporal metrics while minimizing the overall energy consumption and reducing data redundancy during the sensing process. To this purpose, we model our objective as an optimization problem and recall the meta-heuristic Tabu-Search algorithm to design our solution. The introduced Quality and Energy-aware Mobile Sensing Scheme (QEMSS) [16] is also assessed to tackle the final allocation *fairness*. In order to do so, we opt for a multi-objective optimization formulation where we target to maximize the above defined QoI metric along with a fairness measure. Achieving a trade-off among the two conflicting objectives is challenging. Nevertheless, we have proved that the introduced F-QEMSS [17] solution realizes competitive fairness level, measured by the Jain's index, while preserving the same data quality as its preceding solution QEMSS [16].

Our second contribution seizes the question of *How to minimize the processing time of sensing tasks?* To achieve this, we suggest to add a distributed “support” phase of task allocation to the first centralized one. Thus, the MCS platform designates among selected participants those who can carry extra sensing tasks to be delegated on the move to encountered users. Therefore, we study users arrival and preferences models in terms of acceptance/rejection of the proposed assignment. Moreover, we develop a first variant of distributed task allocation that estimates the expected time of sensing (EST) of each potential participant based on historical encounters and decides to whom to assign tasks. The proposed schemes are either only Mobility-aware (MATA) or jointly Preference and Mobility-aware (P-MATA) [18] and are presented in a greedy-based offline and online algorithms. The former algorithm consists of assigning all tasks *at a glance* whereas the second needs to launch the assignment process every time two users encounter. Among the proposed assignment policies, the online preference and mobility-aware has been proved to be the most efficient by achieving the lowest value of average processing time (makespan).

The third contribution of this thesis proposes to build upon the good performances reached by P-MATA. Hence, we propose to extend this variant by introducing a more advanced preferences model which takes into account sensing tasks attributes to compute users' acceptance probability. More importantly, throughout this extension, we have investigated also the incentivizing impact on participants' commitment and as such, the answer to the third question posed in this thesis: *How to motivate users to dedicate part of their resources to the crowdsensing process?* This is performed by varying the incentive policies and conducting extensive simulations to evaluate the new obtained average makespan. Results show that the Incentives-based Preferences and Mobility-aware Task Assignment (IP-MATA$^+$) highly decreases the energetic and time cost of sensing and processing in mobile crowdsensing.

Our fourth and last contribution is rather addressing the *Data Uploading* phase. More precisely, we are interested in enhancing participants commitment to crowdsensing tasks by accounting for their privacy concerns. Thus, we are tackling the question of *How to protect participants privacy and not impact data utility when reporting collected sensory data?* This is due to the fact that major research on privacy-preserving mechanisms in mobile crowdsensing and mainly participatory paradigm *modifies* the collected data. Hence, adding noise to data or aggregating many samples may decrease their final utility. In other terms, protecting participants sensitive information may result in an *unsatisfying* quality of collected data, a major criterion for requesters. Therefore, we look for a trade-off region among privacy-protection and data-quality preservation. The proposed solution, PRUM [19] targets such trade-off while studying adversaries knowledge about participants previously reported data to avoid any correlation that incurs privacy leakage. Simulations show that it is possible to obtain considerable privacy protection level with a slight decrease of data utility.

The contributions presented above are further investigated in details throughout this dissertation as it is described in the next section.

1.3 Organization of the Thesis

This dissertation shows, throughout seven chapters, how to tackle the major identified issues on mobile crowdsensing systems by introducing different scheduling schemes in the various phases of MCS. The organization of this thesis is as follows.

In Chapter 2, we detail the characteristics of mobile crowdsensing systems in terms of scale and paradigms. We also review the existing work on MCS applications and frameworks in both academia and industry. Furthermore, we list the major challenges of MCS and we distinguish participants and requesters concerns. Finally, we present the positioning of our proposed solutions to each MCS issue and involved entities.

Chapter 3 illustrates our first contribution. Specifically, we address requesters quality concerns and participants available resources concerns. The introduced Quality and Energy-aware (QEMSS) optimization model selects users who answer both criteria. Moreover, we intensify the fairness of the final allocation by developing the F-QEMSS solution. Extensive simulations of both QEMSS and F-QEMSS are realized and discussed showing the achieved high data quality levels when compared to state-of-the-art task assignment schemes.

Chapters 4 and 5 present our second and third contributions. Mainly, we define users' arrival and preference models derived from historical encounters in Chapter 4. Accordingly, we develop a distributed assignment which aims to minimize the average makespan of sensing tasks. Chapter 5 develops a further detailed preference model as a regression choice that depends on sensing tasks attributes, especially rewards. This is to study the incentives impact on users' commitment. Simulations realized in both chapters prove that all proposed schemes achieve competitive processing time. Yet, the incentives-based one is the most efficient.

In Chapter 6, we address the fourth and last contribution. Thus, we highlight the necessity of proposing privacy preserving schemes in participatory sensing. Though, we work on the resulted regression of data quality. In other terms, we propose a solution that compromises participants' sensitive information that may be inferred with requesters' data utility requirements. Simulations built upon real traces of crowdsensing applications show that this trade-off is possible to achieve.

To conclude, we resume this dissertation contributions in the final chapter. Furthermore, we explore future research directions such as *how to even more incentivize participants to accept further privacy inference and contribute better data quality*, among other perspectives.

Chapter 2

Crowdsensing: Opportunities and Challenges

Contents

2.1 Introduction 7
2.2 Mobile Sensing: Scale and Paradigms 8
2.3 Crowdsensing Opportunities 9
2.4 Crowdsensing Challenges: Participants Concerns 12
2.5 Crowdsensing Challenges: Requesters Concerns 17
2.6 Positioning of Dissertation Contributions to the studied Literature 19
2.7 Conclusion 20

2.1 Introduction

Crowdsensing is an emergent mobile sensing scheme that comes with various potential applications in different domains. Nevertheless, this new sensing process raises also new issues since it is highly depending on mobile users' *unpredictable* behaviors and limited resources. In this chapter, we review the existing research efforts on Mobile CrowdSensing (MCS) opportunities and challenges. First, we define the different scales and paradigms of MCS. Next, we list the representative MCS applications that we classify by domain. Moreover, we investigate the relevant work on MCS challenges in terms of participants' and requesters' (queriers) concerns. Accordingly, we briefly expose our proposed contributions related to the identified issues in MCS.

2.2 Mobile Sensing: Scale and Paradigms

2.2.1 Mobile Sensing Scale

Lane et al. [1] have distinguished three main mobile sensing scales:

1. **Personal Sensing:** Sensing on smart devices was firstly designed for individual use. A plethora of applications was proposed in this context mainly for tracking users' exercising activities such as UbitFit Garden [20] and SHealth[1] applications, to name a few. In these applications, the generated data and analysis are only shared with the device owner.

2. **Group Sensing:** Sensing can also be developed within individuals sharing the same goal. Particularly, group sensing applications are targeting citizen concerns. For instance, Garbage Watch [7] is a group sensing application that helps students in a campus to capture with their phones' cameras the content of garbage in the aim of improving the recycling process.

3. **Community Sensing:** This type of sensing is the most related to crowdsensing process. The basic idea is to involve large number of participants to collect and share data for the good of a community. This includes monitoring traffic [4–6] or environment [11–15] in a city for instance. Nevertheless, scaling sensing from a personal to a large community comes with several issues related to data analysis and sharing besides users' privacy concerns [1].

2.2.2 Crowdsensing Paradigms

Mobile crowdsensing can be categorized into two major paradigms in accordance to users awareness and involvement during sensing campaigns [21]. These are the participatory [7] and the opportunistic sensing paradigms [1].

2.2.2.1 Participatory Sensing

This paradigm requires an active involvement of smart devices holders. For example, participants are asked to take pictures and record video/audio samples. Besides, they need to make instantaneous and interactive decisions with the different entities of the MCS system [7]. This includes updating their actual location and answering tasks' requests. An advantage of this paradigm is that it leverages the intelligence of participants to get accurate contributions [1]. However, this highly depends on users willingness to participate and dedicate their resources to such campaigns.

---

[1] https://shealth.samsung.com/

2.2.2.2 Opportunistic Sensing

Opportunistic sensing lowers the burden on participants by collecting measurements in a more autonomous way. Hence, users are not aware of active applications gathering data on their sensors [1]. That is, decisions are automatically made by the application regarding data sensing and sharing. However, this system is technically difficult to build since collected data quality depends mainly on handsets' context. For example, a sound measurement would be clearer when the phone is out of the pocket/bag. As a consequence, such paradigm can be limited to only few MCS applications.

2.3 Crowdsensing Opportunities

Several works have been proposed in both academia and industry in the aim of leveraging the potential of mobile crowdsensing. Accordingly, we review hereafter crowdsensing applications and frameworks introduced in the literature.

2.3.1 MCS Applications in Academia

Quite a lot of surveys [1,3,21] have summarized crowdsensing applications into three main categories: Environmental, Infrastructure and Social crowdsensing.

2.3.1.1 Environmental Applications

Researchers [11–15] have studied various environmental phenomena. Rana et al. [11] have presented the design and implementation of an end to end participatory noise mapping system, EarPhone, which comprises signal processing techniques to measure noise pollution at the mobile phone using microphone and GPS sensors. Similarly, the NoiseTube[2] project aims at involving the general public in noise pollution assessment by collecting their geotagged personal noise exposure [12]. Another representative environmental crowdsensing system is PEIR [13]. This application enhances the Personal Environment Impact Report using GPS collected data from users to classify their activities and transportation modes and hence computes some metrics including users Carbon Impact and Fast Food exposure. A comparative participatory air quality sensing mechanism, CommonSense [14] has been introduced to leverage the crowd to collect and share their exposure and air quality measurements. A more recent work on environment monitoring, SakuraSensor [15] has been introduced to extract cherry-lined routes information from videos recorded by car-mounted smartphones and share the information among users in quasi-real time to not miss the cherry blossom period given its *uncertainty*.

---

[2] http://www.noisetube.net

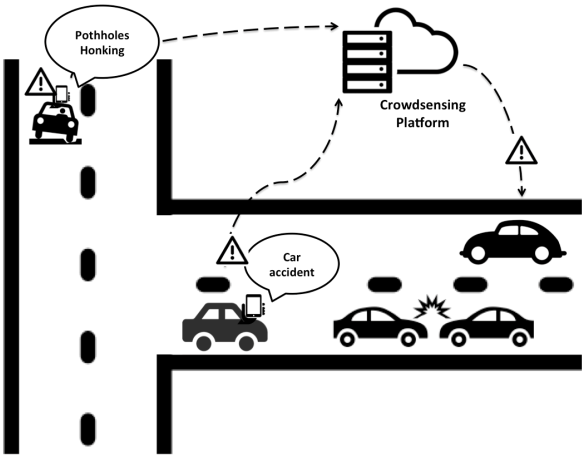


Figure 2.1: Example of traffic monitoring crowdsensing scanario

2.3.1.2 Infrastructure Applications

This type of crowdsensing applications mainly consists of measuring public infrastructure phenomena such as traffic monitoring and mapping [4–6] besides road conditions [4] as described in Figure 2.1 and parking availability [10] detection. Nericell or *TrafficSense* [4,5] and V-track [6] present early research work on traffic congestion detection. These systems perform rich sensing on users' smartphones to measure vehicles speed and traffic delays. Nonetheless, Nericell [4] does not only determine average speed and traffic conditions but also utilizes the accelerometer, microphone and GPS sensors to detect road potholes and honking levels. In the same context, Mobile Millennium[3] is a pilot project on participatory sensing for traffic monitoring which collects traffic data from GPS-equipped mobile phones to estimate traffic conditions in real-time. The collected data is then broadcasted back to users as a traffic map of the target area in order to achieve and enhance more intelligent route decisions. Some other infrastructure applications targeted parking lots availability detection. ParkNet [10] compromises ultrasonic sensing devices installed on cars combined with smartphones applications to detect and share available parking spots in cities.

2.3.1.3 Social Applications

In this category of crowdsensing, individuals are sharing their information to be compared to the rest of a community via social networks. Representative applications are detecting users exercising activities [8, 9, 22] or eating habits [23]. For instance, CenceMe [8] is a personal sensing application that enables participants to share their activities using social networks. Users measured status (walking, sitting, running) by accelerometer and detected context by the microphone besides other habits captured with phones are processed to match users with similar sensing profiles. TripleBeat [9] is a mobile phone based

---

[3] http://traffic.berkeley.edu/

system that assists runners in achieving their workout goals via musical feedback. This application aims to increase personal awareness towards exercising and establishing virtual competition to further motivate participating users. Similarly, BikeNet [22] presents a crowdsensing system that maps cyclists experience. This application urges participants to cycle by aggregating air quality and routes conditions measurements and recommends the most “bikeable” routes. Last but not least, DietSense [23] is a social crowdsensing system enabling users share pictures of what they eat within a community to compare their eating habits. A typical presented example was among diabetics where users control and provide suggestions to each other.

2.3.2 MCS Applications in Industry

In addition to the applications brought by academia, several crowdsensing applications have been designed and implemented in industry. Waze[4] is arguably the most successful crowdsensing based application in the traffic monitoring domain and has already appealed for more than 50 million users. In environment monitoring domain, Stereopublic[5] is a public project collecting geo-tagged noise records from participants to map quiet places in cities. Also, Placemeter[6] is a crowdsensing system using smartphones to record videos on pedestrians and vehicles movement. The processed data helps mapping the crowd of a city. Finally, most tracking activities applications nowadays are based on crowdsensing. Among many other platforms, SportsTracker[7] and Runtastic[8] are smart applications using GPS, accelerometer and heart rate sensors to measure participants achievements and share it within a community.

2.3.3 MCS Frameworks

Several MCS frameworks and platforms [24–29] have been introduced in the literature in order to facilitate the deployment of MCS applications and better assign sensing tasks among participants. Medusa [24] is a programming framework based on the Amazon AMT commercial platform and using a high level programming language, MedScript, to define sensing tasks promoted with monetary incentives. This framework arranges the crowdsensing process into three stages: task submission, workers (participants) selection and incentives management. Vita [26] is a similar framework using open sourced services to develop the customized crowdsensing platform and exchanging service requests via standard web service messages, which is more efficient than the approach of Medusa [24]. McSense [25] is another distributed platform targeting crowdsensing task deployment. In particular, for

---

[4] https://www.waze.com
[5] http://www.stereopublic.net
[6] https://www.placemeter.com/
[7] http://www.sports-tracker.com/
[8] https://www.runtastic.com/fr/

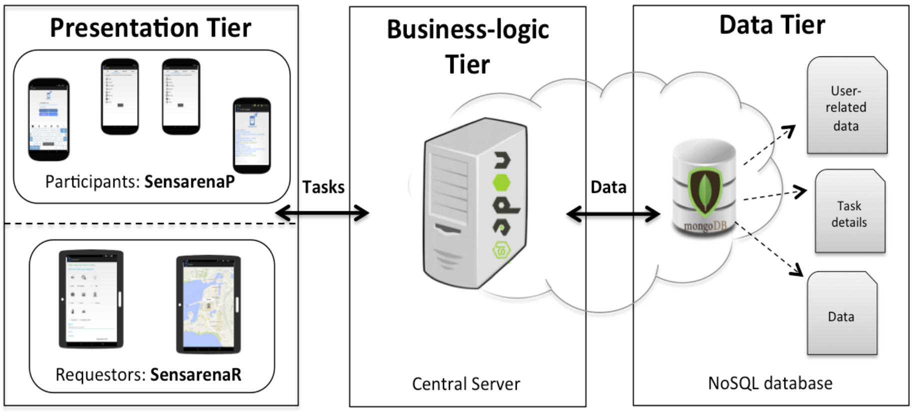


Figure 2.2: Architecture of the crowdsensing platform: Sensarena

each geo-localized task, McSense estimates the time and number of workers required to perform it. Regarding participants' privacy protection, Anonysense [27] has been introduced to anonymize users' IDs during sensor data collection and in network processing. Besides, PRISM [28] addresses tasks deployment combined with security and scalability measurements. Hence, this framework controls sensors access to sensitive data to handle *untrusted* applications. Other ongoing research on crowdsensing platforms has presented Apisense[9] as a platform targeting multiple research communities, and providing a lightweight solution to build and deploy crowdsensing applications for collecting experimental datasets. Also, Sensarena [29] platform is a crowdsensing prototype developed in order to enhance energy-aware and high data quality crowdsensing. This platform is a three-tiered one. The presentation tier introduces two different Android apps; one dedicated to participants to receive tasks notifications and set sensing preferences and the second enabling requesters to submit sensing tasks and aggregate results. The business logic tier and the data tier are illustrated by a Nodejs central server and a NoSQL database (MongoDB), respectively. The latter is organized into three main collections in order to store separately all user-related information, the sensing task details and the collected data. The detailed architecture and core features of Sensarena are illustrated in Figure 2.2.

Despite the great potential of mobile crowdsensing, several issues raise when it comes to its *real* implementation. These issues are mainly related to the human factor, given that users are highly involved in the sensing process, and can be categorized in two classes: participants' concerns and requesters' concerns. Representative work investigating these issues and concurred analysis are detailed in next sections.

2.4 Crowdsensing Challenges: Participants Concerns

The success of mobile crowdsensing is based on participants' willingness and commitment to dedicate their resources (battery, time, processing, etc) and perform sensing tasks. In

---

[9] https://apisense.io/

the following, we list the main challenges that a crowdsensing platform should consider to guarantee users contribution.

2.4.1 Energy Consumption

The primary use of smartphones should be reserved for regular activities of a user (calls, Internet and common apps). Consequently, users accept to participate in MCS only if this process does not use up their handsets batteries [30]. Hereafter, we distinguish smartphones energy consumption within MCS phases: sensing, processing and data uploading.

2.4.1.1 Energy Consumption in Sensing

In this step of MCS, the main incurred energy consumption is due to the activation of required sensors. Indeed, embedded sensors in smartphones are not specifically designed for *heavy* sensing process. Thus, a first measure to reduce their energetic cost is to implement energy-efficient sensors. This should be adopted by industry sensor manufacturers. Meanwhile, one can switch among the less-energy consuming sensors for different measurements. For example, Senseless [31], a system for saving energy consumption in sensing applications for mobile phones, has shown an increase in battery lifetime from 9.2 hours using GPS to 22.2 hours using a combination of accelerometer, GPS and 802.11 for localization samples. A more detailed description on representative values of energy consumption of sensors in a smartphone is illustrated in Table 2.1 based on the study carried by Priyantha et al. [32].

Sensor	Power for active state (mW)
Temperature	0.225
Accelerometer	0.6
Pressure	1.8
Compass	2.7
Gyroscope	19.5
GPS	214

Table 2.1: Energy cost of sensors in a smartphone

In addition, the sensors scheduling during sensing tasks highly impacts the overall energy consumption of the device. In fact, sensors can be activated solely and hence adjusting the sampling frequency is the main factor that minimizes the dedicated power. However, if a sensing application needs a combination of sensors, scheduling becomes more challenging given the necessity to select which sensor to activate, where and when. A lot of works has studied this issue in the aim of minimizing the energy cost of localization. For instance, Constandache et al. [33] have developed the Enloc framework to characterize an optimal localization accuracy for a given energy budget using energy-optimal sensors and prediction-based heuristics. Similarly, Kjaeregaard et al. [34] have introduced EnTracked algorithm to adaptively adjust the GPS sampling rates for users' tracking purpose. Further, authors have extended their solution to an energy-efficient trajectory tracking system named EnTrackedT [35] which considers the adaptive sensor management for other sensors such as compass and accelerometer.

2.4.1.2 Energy Consumption in Processing

This energetic cost is due to the necessary power consumed by mobile phones processors (CPU) during the sensing task. In this context, one of Microsoft Research Projects, Little-Rock [32], developed a tool to reduce the continuous sensing energy overhead by using a dedicated low-power sensing processor (microcontroller) for sampling and low-level processing of sensor data. The dedicated microcontroller minimizes the energy consumed during sensors readings and can transition between sleep and active modes within a very short time. In addition to hardware level implementation, researchers [36, 37] propose to refer to *offloading* techniques where mobile users can delegate the computation to more powerful infrastructure resources generally based on the cloud, thereby extending battery lifetime. Nonetheless, this may come with data uploading energetic cost as detailed hereafter.

2.4.1.3 Energy Consumption in Uploading

Mobile crowdsensing requires a maintained connectivity between different entities to receive tasks assignment and upload collected data. The latter phase can be conducted via 3G/4G, Wi-Fi networks or short range communications such as Bluetooth, Device to Device, etc. However, the aforementioned communication techniques differ in terms of necessary energy such as detailed in Table 2.2. In order to reduce the incurred energy during uploading, Ma et al. [38] present a crowdsensing system which relies on the relatively low-power communications such as Wi-Fi or Bluetooth to upload data rather than directly using 3G/4G. Yet, this can be adopted only for delay-tolerant sensing applications in order to not decrease the temporal-accuracy of samples. Some other works propose paralleling data uploading with phone calls and show that this can save up to 75-90% of energy [39]. According to this assumption, Lane et al. [40] have proposed the Piggyback CrowdSensing (PCS) solution to upload data during phone calls or jointly with the applications of common use.

Network type	Power for connection (mW)
Bluetooth	67
Wi-Fi (802.11)	868
2G (GSM)	500
3G (UMTS)	1400

Table 2.2: Energy cost of mobile phone connection [41]

In summary, the aforementioned measures aim to reduce energy consumption and guarantee individuals commitment to sensing campaigns. Nevertheless, these techniques target the energy issue at a microscopic view. We argue that this should be coupled with energy-efficient task assignment schemes to minimize the overall energetic cost of a sensing task (i.e. handling the energy issue at the macroscopic view as well). This is usually solved by adopting collaborative sensing and selecting a subset of users to perform tasks instead of using up all participants resources. A further study on such assignment methods is presented within Chapter 3 to introduce the related work on our energy-aware task assignment proposals. In such context, Chapters 3 and 4 of this dissertation introduce centralized and distributed assignment approach to handle this issue.

2.4.2 Mobile Data Cost

As we stated in the preceding paragraph, crowdsensing systems require stable connectivity. Therefore, participants utilize Wi-Fi hotspots and if necessary their limited data plans to upload collected data. However, the latter is a key concern for mobile users since it incurs extra fees and affects their 3G budget mainly dedicated for personal use. In order to overcome such issue, the early proposed solution by the research community was to build systems capable of switching between wireless network interfaces on mobile devices in real-time such as MultiNets [42]. This solution does not only reduce users' data plan utilization but also saves energy and achieves higher throughput. Though, offloading data via Wi-Fi may be a limited solution to few applications. Indeed, most of MCS applications require *instantaneous* data uploading such as traffic monitoring where aggregated data must be shared with other users in real time. Therefore, several algorithms have been proposed to efficiently reduce the usage of data networks [43–45]. For instance, authors in [43] encourage participants to allocate a portion of their data budget to participatory sensing tasks by sorting collected data in an ascending utility and upload the most *useful* samples via data networks while offloading the rest via free wireless connections. To this end, a recent work on participatory sensing [44] leverages the phone-to-phone and phone-to-Wi-Fi AP communications besides data prioritization techniques for the uploading phase. This lowers the burden on data networks and decreases the participant dedicated budget for MCS campaigns.

2.4.3 Incentives

Mobile crowdsensing comes with energetic and data costs besides incurring potential privacy leakage to participants. Thus, researchers have introduced incentives to alleviate such issues and encourage users to contribute their sensory data. Incentives can be monetary and non-monetary rewards such as services, 3G budget and games. For a more detailed overview on these incentive mechanisms, we refer to surveys conducted by Zhang et al. [46] and

Jaimes et al. [47]. Nonetheless, we skip in this paragraph monetary rewarding mechanisms as they will be detailed in Chapter 5. Herein, we focus on non-monetary rewards such as entertainment and services as incentives.

Entertainment as incentives is a common adopted policy to attract participants to collect data while using their mobile gaming applications. Particularly, these applications are location-based and can collect data about network infrastructure (Wi-Fi AP or GSM cell towers positions, etc) or map users' trajectories and identified points of interests (Restaurants, libraries, etc). Tremendous mobile games are introduced then to encourage participants to map Wi-Fi coverage [48] or share their location and hence build their trajectories or identify landmarks in cities [49]. The most recent one is the Pokemon Go[10] based MCS framework introduced by authors in [50] in order to incentivize users' contributions via this well-known game.

Furthermore, users may have in some MCS systems both roles of participants and requesters. This includes traffic monitoring applications, namely Waze, as well as parking availability detection [51]. Indeed, users are contributing data in such scenarios and consuming it in the same time. Hence, dedicating a service quota to participants can be an efficient incentivizing mechanism. In this context, Luo and Tham [52] have introduced two incentive schemes: Incentives with Demand Fairness (IDF) and Iterative Tank Filling (ITF) in order to *fairly* share services among users based on their contribution to the framework. Similarly, researchers find the service exchange principle useful in shopping applications. Particularly, in LiveCompare [53] participants use their phone cameras to take pictures of product price tags. By submitting a price data point, participants can receive pricing information for the product at nearby stores.

The common purpose of the aforementioned incentive mechanisms is to ensure an easy deployment of crowdsensing systems in the real-world. However, such non-monetary mechanisms are valid on a limited number of applications that can be designed over existing mobile games [48–50] or proposing services [51–53] as detailed above. Meanwhile, monetary incentives can be adopted in various sensing scenarios, yet with the necessity of requesters dedicating a certain budget to MCS campaigns. The latter is a key concern for requesters as we explain in the next section.

2.4.4 Privacy

Privacy concerns remain the major impediment to users participation in crowdsensing. Undeniably, such paradigm can collect sensitive sensor data pertaining to individuals [3] such as location, time stamps, photos and audio samples. Moreover, sharing collected data in a community scale may induce inferring sensitive information by correlating different

---

[10] http://www.pokemongo.com/fr-fr/

samples reported by a participant. Therefore, countermeasures are needed to encourage participants to perform sensing tasks.

First, MCS applications should enable participants to set their privacy preferences. This was proposed by some MCS frameworks such as PRISM [28] and Sensarena [29] where participants can select which sensors they want to contribute and which to block. Nevertheless, privacy is a subjective perception and a user may not be capable to determine which sensors are divulging his sensitive information. Therefore, privacy preserving mechanisms must be implemented as a transparent process in MCS systems. To this purpose, important research is being conducted to protect participants identities and other sensitive information within the tasking and data uploading phases. For the former phase, the platform may dissociate users Ids with the allocated tasks using anonymization techniques. Hence, the Anonysense [27] platform utilizes Direct Anonymous Attestation (DAA) authentication to send sensing tasks anonymously to participating mobile nodes. Moreover, this platform introduces a k-anonymity reporting scheme in order to guarantee users privacy during the data uploading. That is, each measurement is aggregated with $k-1$ other samples before being reported to the central MCS platform. Further work on privacy preserving schemes in data uploading has been surveyed by Christin et al. in [54]. Proposed techniques include, but are not limited to, data aggregation, data perturbation and encryption-based methods. We enumerate these different methods while investigating their impact on data integrity and utility in Chapter 6. Besides, in that same chapter, we illustrate our PRivacy-preserving Utility-aware Mechanism (PRUM) [19] proposal for protecting different participants sensitive information.

2.5 Crowdsensing Challenges: Requesters Concerns

Data queriers or requesters are users submitting tasks to crowdsensing platforms. They require high data quality samples, thereby they usually dedicate some budget to encourage better contributions. These two concerns may be contradictory given that achieving higher data quality needs requesters to invest more budget on MCS tasks. In the following, we discuss the related work on these issues.

2.5.1 Quality of Information (QoI)

Each sensing task is predefined with time, location besides some additional requirements such as the number of samples to collect and their granularity. Therefore, a broad definition of data quality or Quality of Information (QoI) can be to which level does the collected data answer these characteristics? A more specific definition of QoI was firstly introduced by Bisdikian et al. [55] for wireless sensor networks as the degree of *pertinence* of the information gathered to reflect and understand the measured phenomena.

To better quantify QoI, Sachidananda et al. [56] have listed attributes that can characterize data and have reviewed the proposed metrics in literature. Hence, information can be described by its *accuracy* which is the degree of correctness and precision of the measurement. Indeed, in crowdsensing campaigns, accuracy generally designates the spatial and temporal precision of the uploaded sample, i.e., the time and location of sensing. Accuracy is usually joined with *completeness*, which is an attribute describing the characteristic of information which provides all required facts to a user during the construction of information at the platform [56]. Completeness has been investigated in crowdsensing as the amount of collected data [57] as well as spatial full coverage [58–60] and partial coverage [61] of the target sensing area. These works aim to achieve spatio-temporal accurate and complete measurements. Nevertheless, this assumes that data quality only depends on these criteria which is not always the case. In real life scenarios, sensors may be with low-precision which leads to low quality samples. Besides, participants may be reluctant to optimize the context of measurement or even with malicious behavior which incurs a severe decrease of reported data *utility*. Therefore, MCS platforms should consider other quality attributes that enhance the *reliability* of sensory data. That means, the collected data is free from change [56] and perturbation during all phases from sensing to processing and finally data uploading. This attribute has been investigated for reported data on the web [62] and recently in mobile crowdsensing systems [63].

The above described data quality attributes are the main target to be maximized during sensing tasks. As a consequence, a lot of work on tasks assignment has been introduced to realize a satisfactory level of QoI in terms of amount of data [57] or other quality metrics [16, 17]. These algorithms are further detailed in Chapter 3 as state-of-the-art methods to our proposal on maximizing data quality with energy awareness [16,17]. Nevertheless, achieving high data quality is also depending on other sensing phases such as the data aggregation and computing. In fact, sensing applications may perform additional data processing to perturb the collected samples for privacy preserving reasons. Hence, this may impact the final obtained QoI. Therefore, participants should realize a trade-off between their privacy protection and their contribution quality as we state in our work on privacy-preserving utility aware schemes [19].

2.5.2 Budget

In order to incentivize participants to contribute high quality data, MCS platform and requesters should dedicate a *Budget* to sensing campaigns. The main concern of requesters is then, how to minimize the data cost in terms of allocated budget? and how to efficiently share this budget among participants to obtain satisfying results?

In this context, early work [64] has proposed to minimize the cost of collaborative sensing among participants. The presented algorithm, MCARD, has been designed to select among users with known trajectories in advance those who can provide required data with minimum payments. However, this approach does not guarantee that collected data responds to quality requirements. Therefore, it would be more interesting to tackle both data quality and budget concerns when assigning tasks to participants. In other terms, it is necessary to provide incentives to participants to enhance their commitment while considering requesters available budget. Several task allocation schemes were introduced in the literature to study data quality maximization while respecting a budget constraint. For instance, Liu et al. [57] developed a distributed participant selection scheme (DPS) which computes a satisfaction QoI metric and selects participants that maximize this value. Such approach efficiency is mainly depending on the amount of budget dedicated to pay all selected participants. Other allocation schemes [65,66] propose to pay participants function of their contributions quality in order to optimize both data quality and budget utilization. It is in this trend that we study in Chapter 5 the necessary budget and incentivizing policies to emphasize users' commitment to participatory sensing tasks.

2.6 Positioning of Dissertation Contributions to the studied Literature

The study of the various MCS opportunities and challenges allows us to position our contributions in this thesis with regard to the literature. We focus, in this dissertation, on both participants and requesters encountered issues in the aim of conquering the existing research work and propose adequate solutions. Therefore, we illustrate first which challenge(s) each of our developed methods tackles as detailed in Figure 2.3. Besides, we describe briefly our main contributions positioning in the literature.

Our first work, *QEMSS* [16], studies the data *quality* and *energy* consumption issues in MCS systems. We propose to prevent high energy consumption of participants' devices while achieving competent data quality levels by developing adequate task assignment methods. This work has been extended to consider a *fair* task allocation schedule among participants. *F-QEMSS* [17] targets to jointly maximize data quality and fairness measures while considering energy constraints. In other terms, we propose to answer both requesters' and participants' requirements in the same time, which is slightly introduced in literature but not fully addressed.

In *MATA* and *P-MATA* [18], we extend the task allocation phase in a distributed way. We suppose that some participants can carry extra tasks to delegate them on the move. We suggest then to study users' *mobility* and *preferences* in terms of which tasks to accept. This is made with the aim of minimizing the processing time so that we minimize the *energy* consumption due to computing and maximize data *quality* in terms of temporal accuracy.

IP-MATA$^+$ extends the previous method and studies the impact of *incentives* on participants' commitment. More precisely, we develop a *preference* model which includes, among

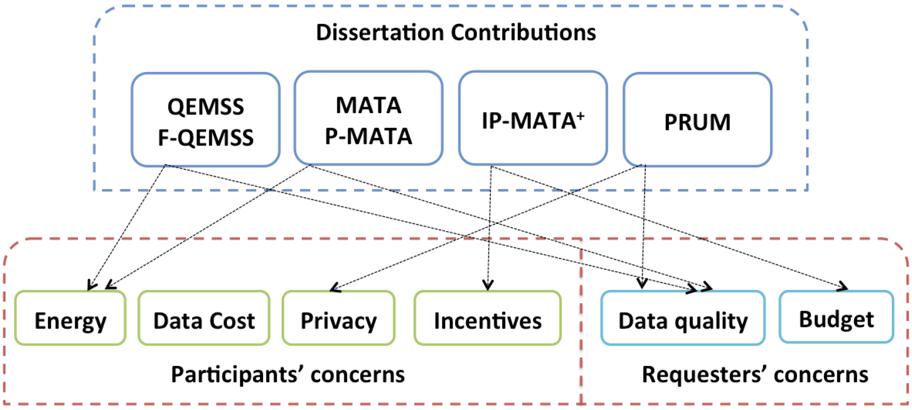


Figure 2.3: Relation between our contributions and MCS challenges

other attributes, rewards. This requires a dedicated *budget* from requesters to be shared as incentives function of users' contributions *quality*. To the best of our knowledge, both *P-MATA* and *IP-MATA*$^+$ are among the early work on distributed task assignment in MCS that consider users' sensing preferences.

Finally, we seize the privacy issue in data uploading phase and introduce the *PRUM* [19] mechanism which obfuscates collected samples before being reported to the MCS platform. However, PRUM accounts also for requesters' data quality constraints and aims to achieve a trade-off which satisfies both participants and requesters. Compared to existing research, our work is among the few ones tackling *privacy* and data *quality* in a row.

2.7 Conclusion

Mobile crowdsensing represents great potential in both academia and industry applications. In addition, it raises new challenges usually described as users' concerns. In this chapter, we reviewed ongoing research work tackling both MCS applications and identified issues to determine the necessary advance on existing literature methods. Accordingly, we propose in this dissertation different methods that leverage collaborative sensing and attempt to satisfy both participants and requesters within the various phases of MCS campaigns as explained in the previous sections and further detailed in the next chapters.

Chapter 3

Quality and Energy-aware Task Assignment Schemes

Contents

3.1 Introduction 21
3.2 Related Works 22
3.3 Background 24
3.4 Preliminaries: Proposed Measures 26
3.5 Quality and Energy-aware Problem Definition 28
3.6 Mobile Sensing Schemes: QEMSS Vs F-QEMSS 32
3.7 Performance Evaluation 35
3.8 Discussion 42

3.1 Introduction

The deployment of mobile crowdsensing campaigns needs to consider both participants' and requesters' concerns. The former users are dedicating their energetic resources and processing different workloads to contribute collected data. Therefore, they are mainly concerned about their devices' energy consumption and the number of assigned sensing tasks. On the contrary, requesters are submitting tasks while requiring a certain level of data-quality, under which the collected samples can be perceived as "useless".

In this context, several works have investigated the aforementioned issues and proposed to prevent them during the task allocation phase [57–59, 67, 68]. For instance, some works studied how to minimize the overall cost of mobile sensing in terms of energetic resources [58, 59, 67, 68] in order to satisfy participants. Other schemes are designed to respond to requesters' data-quality requirements while respecting the platform budget constraints [57]. However, these works mostly focus on one side requirements and slightly studied the others. This may result in unwilling participants to contribute their data. For example, even though being rewarded, participants with low-battery devices may be reluctant to perform tasks. Moreover, minimizing the overall energy consumption of the crowdsensing process does not consider the individual dedicated resources of each participant nor their current devices battery level. In addition, these solutions have studied the overall system performance and did not tackle the following questions: *How fair is our scheduling*? *Were all users ended satisfied when accomplishing sensing tasks?*

In this chapter, we intend to tackle the *Task Assignment* phase in crowdsensing systems while aiming to jointly maximize the quality of collected data when avoiding redundancy, reduce the overall energy consumption, respect an individual energy-level threshold, and ensure a fair workload share among different participants. Particularly, we propose to select a set of participants among the ones present in a specific sensing area within a certain time duration. Participants' devices must verify first an energetic-constraint and provide high-level data-quality and non redundant samples to be selected. We start by setting adequate metrics to quantify the different data quality attributes. Then, we formulate our objective as an optimization problem that we solve using Tabu-Search based algorithm. The proposed scheme is denoted as Quality and Energy-aware Mobile Sensing Scheme, QEMSS [16]. Furthermore, we investigate the fairness of our assignment strategy and we extend our problem to consider fairness as a second objective function. The corresponding solution is named as Fair Quality and Energy-aware Mobile Sensing Scheme, F-QEMSS [17].

The remainder of this chapter is organized as follows. We first enumerate the state-of-the-art methods tackling data-quality, energy and fairness issues in mobile crowdsensing in Section 3.2. We present the analytic and algorithmic tools to design our assignment schemes in Section 3.3. This is followed by a listing of the proposed measures to quantify data quality, energy and fairness. Next, we state the corresponding problem and present proposed assignment solutions in Section 3.5 and 3.6, respectively. The performance evaluation of the proposed mobile sensing schemes, QEMSS and F-QEMSS, is illustrated in Section 3.7. Finally, conclusions and general discussions are given in Section 3.8.

3.2 Related Works

Mobile crowdsensing has been widely studied in terms of applications and challenges as summarized in the preceding chapter. Nevertheless, proposed solutions have tackled different issues separately. Therefore, we review first the energy-aware task assignment schemes in MCS. Then, we investigate the early research work on sensor data quality, denoted as Quality of Information (QoI). Finally, we enumerate the proposed methods that study the fairness issue within a crowdsensing environment.

3.2.1 Energy-aware Task Assignment in MCS

The fact that smart devices are energy-constrained was the major focus of the early research on crowdsensing systems. Sheng et al. [58, 59] proposed to leverage cloud-assisted collaborative mobile sensing to reduce the overall system energy cost. They assumed knowing in advance users' trajectories and estimated the possible resulting coverage of the sensing area. The proposed heuristics proved the potential energy savings of collaborative sensing by designating only a set of users to ensure maximum coverage of the sensing area instead of assigning tasks to all participants. Similarly, Weinschrott et al. [67] adopted the idea of collaborative sensing in the aim of maximizing the sensing area coverage with energy constraints. They proposed to collect *readings* cooperatively and periodically by mobile phones to be mapped into virtual stationary sensors. Authors used two different approaches: the first is based on a central entity in the cloud which predicts mobile users movements and assigns sensing tasks accordingly. Such scheme requires permanent connectivity between *readers* and the central node which is not usually guaranteed. The second approach is based on a distributed coordination of mobile users in an ad-hoc network. The introduced coordination schemes have reduced the average energy consumption per mobile node and reached a competitive spatial coverage considered as a quality metric in this work.

3.2.2 Quality-aware Task Assignment in MCS

One element that hadn't been well addressed in previous works is the quality of collected data, which should be a major requirement for sensing applications. This criteria, denoted as Quality of Information (QoI) [55], has been recently investigated for crowdsensing. In this context, Song et al. [57] studied the QoI issue while providing incentives for selected participants. They defined as QoI satisfaction metric the granularity and quantity of collected data. Then, they estimate the expected amount of collected data by each participant based on his sensing capability, initial location and a probabilistic model which predicts his next move according to his historical trajectory. Based on all these, they established a multi-task oriented mobile users selection strategy, denoted as Dynamic Participant Selection (DPS), to select the subset of participants that maximizes the predefined QoI metric per sensing task subject to a budget constraint. The proposed DPS scheme tackles the QoI notion that we target in this work. Therefore, it is considered as a benchmark when evaluating our proposed assignment schemes. Nevertheless, different from [57], where the energy efficiency is handled very partially, we aim to propose quality and energy-aware task assignment schemes. In addition, we investigate the fairness of our final scheduling.

3.2.3 Fairness in MCS

The fairness issue that arises when assigning tasks to participants was slightly considered in literature. In the following, we enumerate the *fair* allocation methods recently proposed.

First, researchers have tackled this issue for the participatory paradigm of MCS when sharing rewards among users. For example, Luo and Tham [52] introduced a fair incentivizing sensing scheme which targets to maximize both the "social welfare" of the MCS platform and the fairness. Their model presents a demand-based approach to motivate users to participate to sensing tasks. That is, participants receive a regulated quantity of services in return to their contributed data. This has scaled a maximum of participants contribution with an interesting level of fairness among them. However, no specific quality or energy consumption models are attached to this scheme. Few other works addressed the *unfairness* resulted when using participants resources. For instance, Sheng et al. introduced a set of algorithms for energy-efficient sensing scheduling by respecting a fixed number of sensing times for each user in [58,59]. They computed a sensing budget, then assigned tasks to users in a greedy way while maintaining a counter to not exceed this budget value. The proposed algorithms allowed to reach interesting individual energy consumption savings. Zhao et al. [68] developed a fair energy-efficient allocation framework whose objective is to minimize the maximum aggregated sensing time for each participant. Two models were studied; an offline and an online allocation. In the first scenario, a central server is fully aware of the upcoming sensing tasks, which facilitates the fair task assignment to different users. For the online allocation, a dynamic algorithm was established to search for the adequate user to sense while considering the min-max aggregate time objective.

In summary, previous works evaluated their systems fairness by measuring the aggregated sensing time per user. Besides, to the best of our knowledge, no work has considered the quality of collected data which may be highly affected when ensuring only fairness among participants. Therefore, we propose in this chapter the F-QEMSS [17] scheme to solve the problem of maximizing data quality in a fair energy-efficient task allocation and we evaluate our system performance using different fairness metrics.

3.3 Background

In this section, we detail the necessary analytic tools to develop our scheduling schemes. Therefore, we recall the "utility" notion to quantify data-quality in crowdsensing and the meta-heuristic of Tabu Search generally used to solve optimization problems.

3.3.1 Proposed Utility Function

The basic utility theory was developed by Von Neumann and Morgenstern [69]. Since then, it was widely utilized in economy to designate the value of a good/service perceived by a consumer [70]. This theory was also extended to other fields as a "utility function" to quantify some preferences in the process of making a decision. Accordingly, researchers evaluated for example networks or services based on different utility functions such as the exponential, logarithmic or sigmoid forms [71].

In our work, we opt for the utility function proposed in [71] to quantify data quality attributes. The reasons behind this choice is that this function is continuous and normalized. Besides, it assigns different forms depending on the type of the quality criterion x. For example, if x is an upward criterion in $[x_\alpha, x_\beta]$, then the utility of x increases along with its value. For such type of variable, we use the function of Equation (3.1).

$$U(x) = \begin{cases} 0 & x < x_\alpha & \text{(3.1a)} \\ \dfrac{(\frac{x-x_\alpha}{x_m-x_\alpha})^\zeta}{1+(\frac{x-x_\alpha}{x_m-x_\alpha})^\zeta} & x_\alpha \leq x \leq x_m & \text{(3.1b)} \\ 1 - \dfrac{(\frac{x_\beta-x}{x_\beta-x_m})^\gamma}{1+(\frac{x_\beta-x}{x_\beta-x_m})^\gamma} & x_m < x \leq x_\beta & \text{(3.1c)} \\ 1 & x > x_\beta, & \text{(3.1d)} \end{cases}$$

where $\zeta \geq \max\{\frac{2(x_m-x_\alpha)}{x_\beta-x_m}, 2\}$ and $\gamma = \frac{\zeta(x_\beta-x_m)}{x_m-x_\alpha}$ are the tuned steepness parameters, and x_m is the median of $[x_\alpha, x_\beta]$.

3.3.2 Tabu Search

Tabu Search (TS) is a meta-heuristic introduced by Glover [72] to guide a local heuristic search procedure to explore the solution space beyond local optimality [73]. The basic idea is to forbid a *move* that would return to recently visited solutions called tabus. This makes Tabu Search one of the most efficient heuristic techniques in the sense that it finds quality solutions in relatively short running time.

Essentially, given Ω, the set of possible solutions to a problem, for each solution $x \in \Omega$ it exists a subset of Ω called *neighborhood* of x, $N(x)$. The neighborhood contains feasible solutions, each is obtained by making a simple move from the solution x. A move m can be defined as adding/deleting an element to/from the current solution. Besides, the Tabu Search algorithm uses a memory structure called Tabu List (TL) to avoid cycles. A solution among $N(x)$ is selected only if it does not exist in TL. At each iteration, TS algorithm updates TL by adding attributes of the selected solution. Attributes usually do not contain the complete solution to facilitate handling TL. The size of TL is called Tabu Tenure (TT). It is crucial to define an adequate TT, because if it is too small, then there is a high chance to have cycles and hence TS cannot go beyond the local optimal solution. However, if TT is very large, very few options are left for the neighborhood formation.

In general, TS algorithm consists of four steps: initialization, neighborhood formation, neighbor selection and TL update. The former phase is necessary to generate an initial feasible solution x_{init}. Nevertheless, the further is this solution from the optimal one, the greater is the overall time of execution of the method. Therefore, it would be more efficient not to start with a totally random solution. The use of this search in designing our task assignment schemes, QEMSS and F-QEMSS, is presented in details in Section 3.6.

3.4 Preliminaries: Proposed Measures

We aim in this first contribution to respond to participants' energetic and fairness concerns besides accounting for requesters data quality requirements. In this section, we start by associating each targeted mobile crowdsensing issue with appropriate measure.

3.4.1 Quality of Information Attributes and Metrics

Data quality can be described using different attributes. In this work, we refer to the survey conducted by Sachidananda et al. [56] on data characteristics in WSNs. We select three attributes: "Completeness", "Timeliness" and "Affordability". The two first are considered as data quality metrics while the last is utilized as an energy metric.

3.4.1.1 Completeness

This attribute characterizes whether the measured data provides all required facts during the construction of the information [56]. In crowdsensing systems, this can be considered as to which extent the collected sample by a participant device reflects the ground truth to the requester asking for the measured phenomena. Herein, we focus on the spatial completeness and define it as the area potentially covered by a participant. According to [71], the "completeness" metric is an upward criterion, as the more a user covers an area, the more useful is his collected data. Hence, given a variation range of the coverage x, $x_\alpha \leq x \leq x_\beta$ and a middle point of the utility x_m, an adequate utility function for x is the one defined by Equation (3.1). That is, $U_c(x) = U(x)$.

3.4.1.2 Timeliness

"Timeliness" is the attribute describing the freshness of the uploaded information to the crowdsensing platform. Each sample must be collected and uploaded within a limited time interval which is in our case the sensing period. We design the "timeliness" metric as a centered utility function, given that the sensing period is a time interval with the required instant of measurement as the median. This ensures that as the time of measurement gets further from the required moment, its utility decreases. Given a variation range of the instant x, $x_\alpha \leq x \leq x_\beta$ and the required instant of measurement x_r as a middle point of the utility, we formulate the "timeliness" metric as follows:

$$U_t(x) = \begin{cases} 0 & x < x_\alpha, x > x_\beta \quad (3.2a) \\ \frac{(\frac{x-x_\alpha}{x_{m1}-x_\alpha})^\zeta}{1+(\frac{x-x_\alpha}{x_{m1}-x_\alpha})^\zeta} & x_\alpha \leq x \leq x_{m1} \quad (3.2b) \\ 1-\frac{(\frac{x_r-x}{x_r-x_{m1}})^\gamma}{1+(\frac{x_r-x}{x_r-x_{m1}})^\gamma} & x_{m1} < x < x_r \quad (3.2c) \\ 1 & x = x_r \quad (3.2d) \\ 1-\frac{(\frac{x-x_r}{x_{m2}-x_r})^\zeta}{1+(\frac{x-x_r}{x_{m2}-x_r})^\zeta} & x_r < x \leq x_{m2} \quad (3.2e) \\ \frac{(\frac{x_\beta-x}{x_\beta-x_{m2}})^\gamma}{1+(\frac{x_\beta-x}{x_\beta-x_{m2}})^\gamma} & x_{m2} < x \leq x_\beta \quad (3.2f) \end{cases}$$

x_{m1} and x_{m2} are respectively the middle points of the time intervals $[x_\alpha, x_r]$ and $[x_r, x_\beta]$.

The total utility of a data measurement is the product of all its QoI attributes utilities:

$$U_{QoI} = U_c(x) \times U_t(x). \tag{3.3}$$

This formulation makes the collected information "useful" only if it satisfies all data quality requirements set by a requester. That is, in case one quantity is null, the overall data utility provided by the corresponding participant in a specific area is null which forbids, for example, collecting data out of the sensing time interval. Equation (3.3) illustrates the formulation of the first objective function to be maximized by the developed task assignment schemes in this chapter, QEMSS and F-QEMSS.

3.4.1.3 Energy Metric: Affordability

Different from state-of-the-art works [59, 67], we consider here an individual energy consumption measure, denoted as "affordability". This term indicates the ability of a participant to perform a sensing task depending on his handset battery level. The energy cost of data acquisition, represented by the "affordability" metric, is a downward criterion [71] as a participant is less likely to be selected to sense if his handset energy level is low. Thus, we opt for the decreasing function defined by Equation (3.4) to quantify the energy metric; where x designates a device's battery level and varies in $[e_\theta, 100\%]$, and e_θ is an energy threshold level thereby $U_e(e_\theta) = 0$.

$$U_e(x) = 1 - U_c(x). \tag{3.4}$$

The "affordability" metric is a limiting condition in the crowdsensing process. Therefore, it will be considered as a first constraint in our optimization problem.

3.4.2 Fairness Metric: Jain Index

In addition to QoI maximization, we aim in this work to minimize the number of sensing times per user, ns_i, to ensure users commitment to sensing campaigns. In other terms, we want to achieve a system with a maximum of fairness. Thus, we need to answer these two questions: *what type of fairness to target?* and *which measure to use to quantify it?*

Fairness is a subjective performance metric as stated by Shi et al. [74], i.e., a fair system is not necessarily a system with equal allocation. According to this principle, we consider that a fair system does not assign a fixed number of sensing times to all users but rather as comparable as possible since they have different sensing capabilities. Moreover, a fairness model should provide a real number to imply the level achieved by the system. Therefore, we opt for a quantitative fairness model based on the Jain's index, introduced first in [75]:

$$J = \frac{(\sum_{i=1}^{n_p} \delta_i)^2}{p \sum_{i=1}^{n_p} \delta_i^2}. \tag{3.5}$$

where p is the number of selected users and $\delta_i = U(ns_i)$ is the normalized number of sensing times ns_i using the utility function of Equation (3.4); with $ns_i \in [0, ns_{max}]$.

Note that if all users perform an equal number of tasks, the Jain's index achieves 1. Thus, a larger value of this measure represents a fairer task assignment. We target then to maximize J in order to maximize the fairness of our assignment schemes. This is formally modeled with the corresponding optimization problem in the next section.

3.5 Quality and Energy-aware Problem Definition

Our research goal is to design efficient task assignment schemes that jointly maximize QoI criteria set by requesters in MCS campaigns, protect participants' resources and ensure a fair task allocation among them. To model this problem, we describe the crowdsensing system that we consider and the different assumptions.

3.5.1 Crowdsensing System Overview

We consider a $2D$ spatial sensing area representing a city and a sensing period T. The system consists of registered users, $p_i \in P, \forall i \in \{1, \dots, n_p\}$ within a central MCS platform as shown in Figure 4.2. Each user carries his device equipped with built-in sensors and is considered as a candidate participant to sensing campaigns. Users are moving arbitrarily in the sensing region following available paths. The trajectory of each user can be then formally defined as a set of possible visited areas ("locations"). We subdivide the sensing region into m sub-regions, $a_i \in A, \forall i \in \{1, \dots, m\}$, of one or more locations. For simplicity, we consider that tasks are assigned by sub-regions. Hence, the set of sensing tasks S

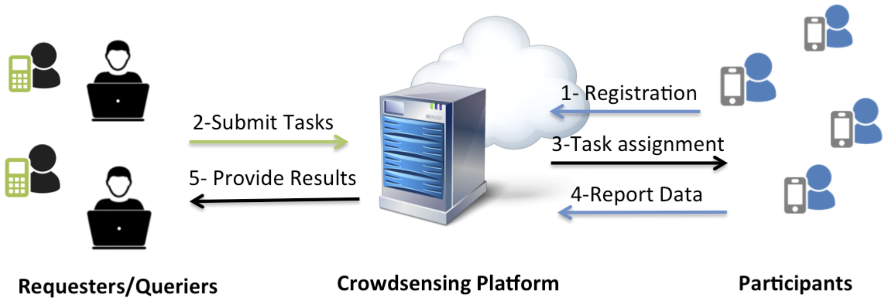


Figure 3.1: Mobile crowdsensing centralized architecture

is with the same size of the set of sub-regions A. Finally, we divide the whole sensing period into time slots $t_i \in T, \forall i \in \{1, \ldots, T\}$. In the beginning of each sensing time slot t_i, the MCS server notifies participants P about the upcoming sensing tasks associated with a set of QoI requirements, $q_i \in \{q_1, \ldots, q_L\}$, such as the quantity of collected data, sensing region and time of measurements. Once notified, each participant should update his location and the remaining energy of his device. In our scenario, we assume that the unit energy consumption of a sensor module on different types of devices is almost identical as presented by authors in [76] and we denote it by e_c. In addition, we set a threshold level on energy e_θ under which the handset owner in question is removed from the candidate list of participants during future sensing periods. This guarantees to participants the necessary energy for normal use of their hand-held devices. Finally, we limit the number of required measurements to one measurement per sub-region a_i to avoid redundant information.

Based on all these, the MCS server selects a subset of users which maximizes the targeted system efficiency. We propose to focus on maximizing only the quality of collected samples while respecting the energetic and redundancy constraints in the first scheme, QEMSS. Further, we aim to study the maximization of both QoI and fairness among participants as a second multi-objective problem, F-QEMSS. According to the required optimization, the server generates a sensing schedule and broadcasts it to designated participants.

3.5.2 Known/Unknown Users Trajectories

We assume that users trajectories can be known in advance, i.e., deterministic, or unknown. The first approach suggests that the MCS server knows in advance each user trip as adopted in [58, 59]. This helps to define a benchmark for the unknown case. The latter scenario supposes that participants announce their initial locations obtained via location services such as GPS to the crowdsensing platform at the beginning of each sensing period. To predict next possible visited areas, we suggest to utilize Markov model for mobility prediction [77]. Variants of this model are widely used to model the transition probability between users current locations and potential destinations [77–79]. In our work, we choose the second-order Markov model, that is to predict the next position based on the two last visited ones. To this purpose, we define the probability of visiting a future area p_{rs} as the number of trajectories Tr_i which have the sequence of positions (v_r, v_s) divided to the number of trajectories which have v_r as introduced in [77]:

$$p_{rs} = \frac{|\{Tr_i|(v_r, v_s) \in Tr_i\}|}{|\{Tr_i|v_r \in Tr_i\}|}, \tag{3.6}$$

where v_r is the last two visited positions vector and v_s is the future sub-area to be visited.

After calculating all probabilities between the different states, a transition matrix Mx is created which can be used to estimate the potential future position of each participant p_i, and hence generate his total predicted trajectory.

3.5.3 Problem Formulation

Given the crowdsensing system described above, we proceed to formulate our task assignment problem. We do so by utilizing the predefined QoI, energy and fairness metrics to design the objective functions and constraints. Nevertheless, we distinguish two optimization problems: first by maximizing only QoI and second by considering both fairness and data-quality maximization.

3.5.3.1 QoI Maximization Problem

The first aim of our assignment methods is to maximize the overall quality of the collected data by selected participants while minimizing the dedicated resources to achieve that. The corresponding optimization problem is formulated as follows:

$$\begin{aligned} &\underset{X}{\text{Maximize}} && \sum_{i=1}^{n_p} \sum_{j=1}^{m} x_{ij} \times U_{QoI}(QoI_{ij}) && \\ &\text{subject to} && U_e(e_i - e_c \sum_{j=1}^{m} x_{ij}) \geq 0 && \forall p_i \in P \\ &\text{and} && \sum_{i=1}^{n_p} x_{ij} \leq 1 && \forall a_j \in A \end{aligned} \tag{3.7}$$

where X is a $users \times areas$ sized-matrix with elements in $\{0, 1\}$. If $x_{ij} = 1$ then the participant p_i is selected to sense in area a_j.

In Problem (7.1), we utilize the QoI general utility function defined by Equation (3.3) to state the objective function. Additionally, we set the energy constraint as the comparison of a device residual energy to the threshold level, e_θ. This is to ensure to each mobile user the necessary amount of energy for common activities. Hence, we denote by e_i the initial battery level of a participant p_i handset. We update this value by subtracting the sensing energetic cost e_c every time the corresponding participant is selected to sense, i.e., $e_i - e_c \sum_{j=1}^{m} x_{ij}$. Finally, we develop a second constraint to avoid redundant information by setting the maximum number of samples collected in a specific area to 1.

3.5.3.2 QoI and Fairness Maximization Problem

We investigate as a second objective the fairness of our assignment policy. That means, we aim to find the subset of participants that maximizes both QoI of collected data and system fairness while minimizing the dedicated resources. To do so, we formulate the following multi-objective optimization problem:

$$\begin{aligned}
&\underset{X}{\text{Maximize:}} \quad U_{QoI}(X) = \sum_{i=1}^{n_p} \sum_{j=1}^{m} x_{ij} \times U_{QoI}(QoI_{ij}) \\
&\underset{X}{\text{Maximize:}} \quad J(X) = (\sum_{i=1}^{n_p} \delta_i)^2 / p \sum_{i=1}^{n_p} \delta_i^2
\end{aligned} \tag{3.8}$$

The above defined two objective functions are conflicting. In general, the maximum achieved QoI is obtained when the best candidates are always selected regardless of the number of sensing tasks they perform. On the contrary, the maximum of system fairness is realized when all participants are assigned to the same number of tasks which may result in lower quality of data. To resolve this issue, we must look for the Pareto optimal solution of this problem. That is, the solution that answers both objectives and it is not possible to move from it to another solution to improve one of the objectives without decreasing the other. Therefore, we opt for an efficient way to resolve such multi-objective optimization (MOO) problem, which is the weighted sum [80]. Accordingly, we transform the predefined objectives into a unique weighted-sum objective function. Note that in order to find the Pareto optimal solution, the associated weights to objective functions must be positive. Furthermore, we refer to the predefined energy-aware and no-redundancy constraints of Problem (7.1) given that we respect the same conditions. Based on all this, the corresponding optimization problem is formulated as follows:

$$\begin{aligned}
&\underset{X}{\text{Maximize:}} && \mu_1 \times U_{QoI}(X) + \mu_2 \times J(X) \\
&\text{subject to:} && U_e(e_i - e_c \sum_{j=1}^{m} x_{ij}) \geq 0 && \forall p_i \in P \\
&\text{and} && \sum_{i=1}^{n_p} x_{ij} \leq 1 && \forall a_j \in A
\end{aligned} \tag{3.9}$$

where μ_1 and μ_2 are positive weights and $\mu_1 + \mu_2 = 1$.

Without loss of generality, we denote the objective function of both Problems (7.1) and (3.9) as the system efficiency: $E(X)$. In order to solve these optimization problems, we look for a "combination" of participants who satisfy both constraints and maximize the objective function. However, combinatorial optimization problems are challenging to solve whenever a large number of entities is involved. Given the fact that we are working on a large scale crowdsensing scenario, looking for the optimal solution can be achieved in a non polynomial-time process [57]. As a consequence, it would be more appropriate to search for a sub-optimal solution using heuristic methods. According to this observation, we adopt the Tabu Search meta-heuristic introduced in Section 3.3 to design our task assignment schemes, QEMSS and F-QEMSS, as illustrated in next section. This choice is enhanced by the fact that crowdsensing large scale assignment should be solved in a short running time. Hence, some scheduling must be forbidden and labeled as tabus such as assigning tasks to users within regions that do not belong to their trajectories.

3.6 Mobile Sensing Schemes: QEMSS Vs F-QEMSS

Throughout this contribution, we present two task allocation methods. First, we aim to resolve Problem (7.1) by introducing a QoI and Energy-aware Mobile Sensing Scheme, QEMSS. Moreover, we propose to solve the MOO Problem (3.9) by a Fair QoI and Energy-aware Mobile Sensing Scheme, F-QEMSS. We make use of Tabu-Search algorithm to design both solutions and we denote by $E(X)$ the system efficiency to be maximized in both cases. Recall that $E(X) = U_{QoI}(X)$ for QEMSS scheme, while it is set to $E(X) = \mu_1 \times U_{QoI}(X) + \mu_2 \times J(X)$ for F-QEMSS. In the following, we detail the different elements and steps of our TS-based assignment solutions, QEMSS and F-QEMSS.

3.6.1 Elements of TS-based Schemes

3.6.1.1 Solution $X \in \Omega$

A possible solution X of the proposed TS-based algorithms is a $users \times areas$ sized boolean matrix, i.e., $x_{ij} \in \{0, 1\}$. This solution must answer the constraints of the optimization Problems (7.1) and (3.9). That means, for each participant p_i, a row of X, the remaining energy of his mobile device, $e_i - e_c \sum_{j=1}^{m} x_{ij}$, needs to be above the defined threshold e_θ. Besides, to have at most a participant per sub-area, the sum of each column, i.e., area a_j, of the feasible solution must be less than or equal to 1 to avoid redundancy.

$$X = \begin{array}{c} \\ p_1 \\ \vdots \\ p_{n_p} \end{array} \begin{array}{c} \begin{array}{ccc} a_1 & \dots & a_m \end{array} \\ \begin{pmatrix} x_{1,1} & \dots & x_{1,m} \\ \vdots & x_{i,j} & \vdots \\ x_{n_p,1} & \dots & x_{n_p,m} \end{pmatrix} \end{array},$$

3.6.1.2 Move m

A move $m \in M(X)$ is a modification applied to an initial solution X_{init} to generate other possible solutions. We consider as a move the swapping of the assignment of two participants present in a same sensing sub-region a_j. Thus, a move m can be presented as a matrix of the same size as X with all elements equals to zeros except those of the old and the new assignment positions, which are set to 1.

3.6.1.3 Tabu List (TL)

The Tabu List is the structure where the TS algorithm stores visited solutions to avoid local optima. In this work, we choose to store attributes of a visited solution X rather than the whole matrix. Thus, we compute the maximum achieved efficiency value for each $X \in \Omega$ and consider it as its main attribute. Further, we update the TL by adding the best achieved value, i.e., $E(X_{best})$ in each iteration as in [73].

3.6.2 Phases of TS-based Schemes

As introduced in Section 3.3, a Tabu-Search meta-heuristic is mainly composed by four phases: Initialization, Neighborhood Formation, Neighborhood Selection and Tabu List update. Hereinafter, we describe the adjusted steps of each phase.

3.6.2.1 Initialization

This phase targets the generation of an initial solution X_{init} to the optimization Problem (7.1) or (3.9). Thus, we need to satisfy first the energy and redundancy constraints of these problems. A possible and *trivial* solution may be the identity matrix which answers the second constraint, i.e., at most one participant per sub-area. However, note that we can not assign a user to an area which does not belong to his trajectory. To avoid this issue, we propose a simple heuristic to find an initial feasible solution. The basic idea is to conduct a “greedy-based” search as presented in Algorithm 1. First, we select a random sub-region a_j in the set of all sensing sub-regions A. Then, we look for the participant p_i with the maximum value of system efficiency $E_{i,j}$ and a residual mobile energy above the defined threshold, i.e., $e_i \geq e_\theta$. We repeat this procedure till we cover all required sensing areas or no more candidates are available.

3.6.2.2 Neighborhood Formation

Starting from a solution $X \in \Omega$, we can generate the neighborhood $N(X)$ by applying one move $m \in M(X)$. As previously described, the move m consists of swapping a designated

Algorithm 1 Initialization: Greedy-based search

```
Require: Set of sensing sub-regions $A$ , Participants $P$, Efficiency matrix $E$.
Ensure: Set of selected users $X$
while $A \neq \emptyset$ do
    Select a random sub-area $a_j \in A$
    while $select = 0$ do
        Select the participant with the maximum efficiency in the area $a_j$:
        $[E_{i_{mx}}, i_{mx}] \leftarrow \max(E_j)$
        Verify the energy constraint:
        if $U_e(e_{i_{mx}}) > 0$ then
            $X(i_{mx}, j) \leftarrow 1$
            $select \leftarrow 1$
        else
            $X(i_{mx}, j) \leftarrow 0$
            $select \leftarrow 0$
        end if
    end while
    $A \setminus \{a_j\}$
end while
Return $X$
```

participant p_{i_1} to sense by the current solution X with another participant p_{i_2} in the same sensing area a_j. This move is based on a generated reference matrix X_{ref}, where $x_{ref}(i,j) = 1$ only if the participant p_i stopped by the area a_j during his trip. All moves are chosen among the 1-elements of the current region (column of the matrix X_{ref}). As a consequence, we obtain a "Reduced Neighborhood". That means, we limit the number of potential moves and corresponding solutions by respecting the availability of participants in each sensing sub-area which facilitates the search among feasible solutions. Similar to the initial solution generation, we verify that each neighbor $X' \in N(X)$ conforms with the constraints of both optimization problems.

3.6.2.3 Neighborhood Selection

After generating the neighborhood of the current solution, we select the neighbor $X' \in N(X)$ with the far best system efficiency E_{best} to be considered as the initial solution for the next iteration.

3.6.2.4 Update Tabu List

Finally, we update the Tabu List by adding only the attributes, i.e., the system efficiency of the selected solution and not the whole matrix X. This does not only forbid recycling to already visited solutions but also not to visit solutions with the same system efficiency. Consequently, we reduce both the time of computation and the required memory for TL.

Algorithm 2 TS for maximizing the overall System Efficiency

```
Require: Set of sensing sub-regions $A$, Participants $P$, Efficiency matrix $E$.
Initialization: Generate an initial solution $X_{init}$ and compute its system efficiency $E(X_{init})$.
$X \leftarrow X_{init}$
$X_{opt} \leftarrow X_{init}$
$E_{opt} \leftarrow E_{init}$
while $iter < \max(iter)$ do
    Neighborhood Formation: Generate all neighbors of the current solution $X$ by applying moves $m \in M(X)$ except those in TL.
    Neighborhood Selection: Select the neighbor solution $X' \in N(X)$ with $E_{best}$ to be considered as the initial solution for the next iteration.
    if $E_{best} > E_{opt}$ then
        $X_{opt} \leftarrow X_{best}$
        $E_{opt} \leftarrow E_{best}$
    end if
    Update TL: Add the attributes of $X_{best}$ to TL.
    $X \leftarrow X_{best}$
end while
Return $X_{opt}$
```

Algorithm 2 designs in details the search strategy for both QEMSS and F-QEMSS assignment schemes. The evaluation of these methods is conducted via simulations while compared to existing concurrent task allocation strategies in mobile crowdsensing from the literature as described in the following section.

3.7 Performance Evaluation

This section describes the simulation environment set to evaluate the task assignment methods proposed in this chapter. Then, and after presenting the evaluation metrics, we discuss the achieved results when comparing our schemes, QEMSS and F-QEMSS, to state-of-the art methods.

3.7.1 Simulation Settings

We implemented the proposed allocation methods on the Matlab environment. We set, as a simulation area A, a city of $4000m \times 6000m$ scale which includes 20 horizontal “roads” and 20 vertical ones considered as sub-areas a_j. Given that A is comparable to the well known Manhattan-city model, we use mobility traces of the Manhattan mobility model generated by Bonnmotion [81]. Each participant enters a road with a random speed picked from $[1, 3]$ m.s^{-1}, moves straight with a probability of 0.5, turns left or right with respective probabilities of 0.25 each. The number of participants is varied from 100 to 1000. Moreover, we set the energy threshold level e_θ to 30%. This limit is purposely set relatively high in order to ensure to participants normal phone use. Accordingly, we generate users handsets initial energy as a uniformly-distributed random variable in $[e_\theta, 100\%]$. Finally, we set the weighted sum coefficients of F-QEMSS objective function (3.9) to $\mu_1 = \mu_2 = 0.5$. We carried out simulations for $T = 16$ h subdivided into 16 time slots of 1 h each.

3.7.2 Benchmark

We compare our first contribution schemes to three main benchmark:

- QEMSS is compared to a Random Selection (RS) algorithm which consists of randomly selecting participants to perform sensing tasks with no QoI or energy consideration. In addition, we implement the DPS assignment scheme [57], a greedy-based strategy, which selects users who maximize a QoI metric when respecting a budget constraint. The evaluation of QEMSS is studied with known and unknown trajectories to investigate the impact of mobility prediction.
- F-QEMSS is compared to its predecessor QEMSS and DPS [57] which both target only maximizing QoI with no fairness consideration. Besides, we consider a baseline method defined as Equal Sensing (ES) which assigns an equal number of tasks to all participants independently of their contributions' quality. It is worth noting that, without loss of generality, this evaluation was conducted only on known trajectories for simplicity reasons.

3.7.3 Evaluation Metrics

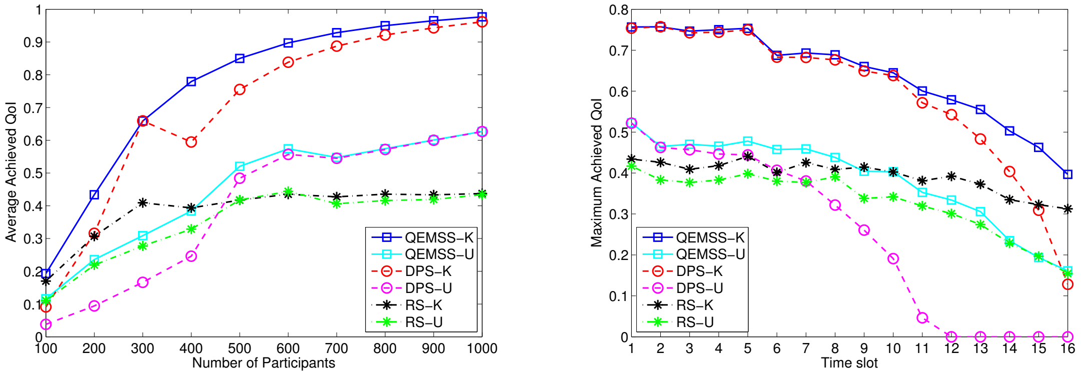
To evaluate the performance of QEMSS and F-QEMSS, we introduce four main metrics:

- *Achieved QoI*: this metric computes the average achieved QoI by all participants as defined by Equation (3.3).
- *Spatial Accuracy*: we used the Manhattan-distance function to compute the distance between the spatial coordinates of the required sensing area and the collected data:

$$s_a = 1 - \sum_{i=1}^{m} |X_i - Y_i|, \tag{3.10}$$

- *Temporal Accuracy*: this metric quantifies the “timeliness” of the collected measurement time compared to the required instant of sensing as introduced in Equation (3.2).
- *Fairness metrics*: we measure the fairness achieved by the Jain index defined by Equation (3.5) and we compute the variance of the number of sensing, ns:

$$V(ns) = \frac{1}{n_p - 1} \sum_{i=1}^{n_p} (ns_i - \bar{ns})^2 \tag{3.11}$$



(a) Average QoI for different number of users (b) Maximum achieved QoI by 400 users

Figure 3.2: Achieved QoI by different selection schemes

where ns_i is the number of sensing times for a participant p_i and $\bar{ns}$ is the mean value of $ns_i, \forall i \in 1 \ldots n_p$.

3.7.4 Evaluation Results

As a first evaluation, we investigate first the performance of QEMSS when compared to DPS and RS in terms of achieved QoI, then in terms of spatial and temporal accuracy. These two evaluations are drawn while considering both known and unknown trajectories. After that, we propose to study the efficiency of our proposed *fair* allocation, F-QEMSS, when compared to its predecessor QEMSS, as well as to DPS and ES.

3.7.4.1 Maximum Achieved QoI

In Figure 3.2, we plot the data quality levels realized by the three task assignment methods: QEMSS, DPS and RS. We associate the $-K$ and $-U$ symbols to all notations to designate the use of Known or Unknown participants' trajectories.

The *average* value of QoI measured during each sensing time slot t_i while varying the number of participants $n_p \in [100, 1000]$ is shown in Figure 3.2(a). Naturally, the achieved data quality level increases as function of the number of users available in the sensing area A. First, we observe that QoI level values realized by participants selected via the random selection (RS) method are the lowest. This is due to the fact that this scheme assigns randomly tasks to participants without comparing their estimated contributed data-quality which yields to *poor* quality levels. Differently, our quality and energy-aware solution, QEMSS, and its benchmark DPS perform better and in a comparable way, especially for high density areas, i.e., $n_p \geq 800$. Nevertheless, for low density sub-regions, $100 \leq n_p \leq 400$, QEMSS outperforms both DPS and RS. This is observed for both deterministic and prediction algorithms, i.e., for known and unknown trajectories. It is worth noting that the

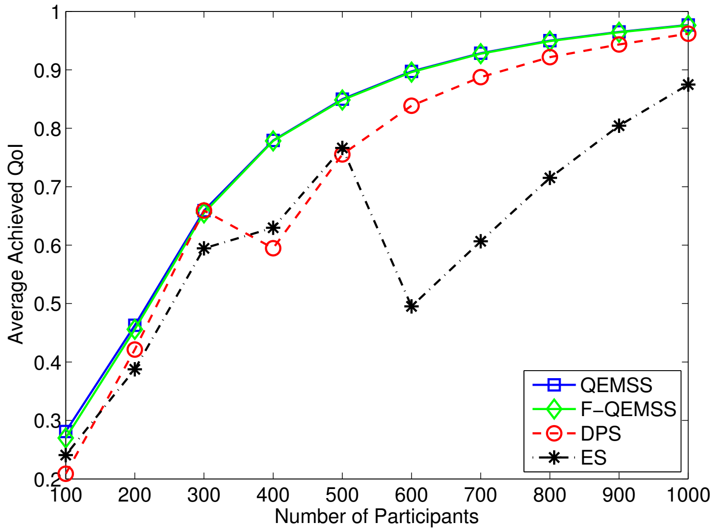
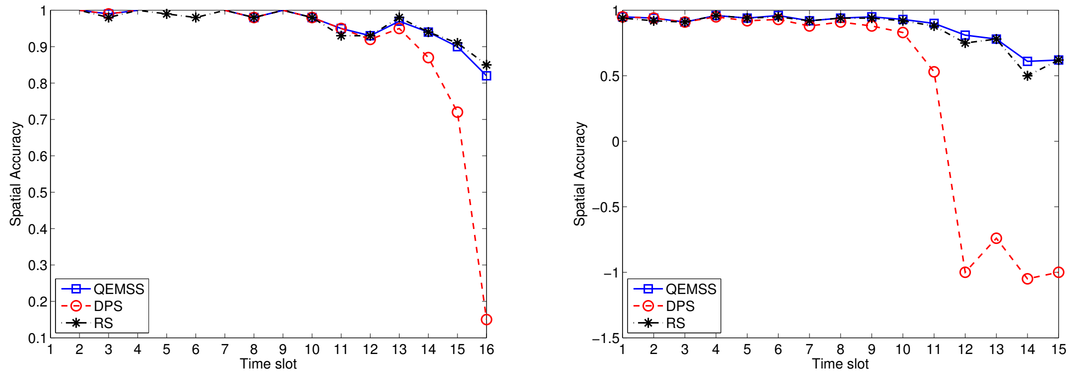


Figure 3.3: Average achieved QoI by F-QEMSS, QEMSS, DPS and ES schemes

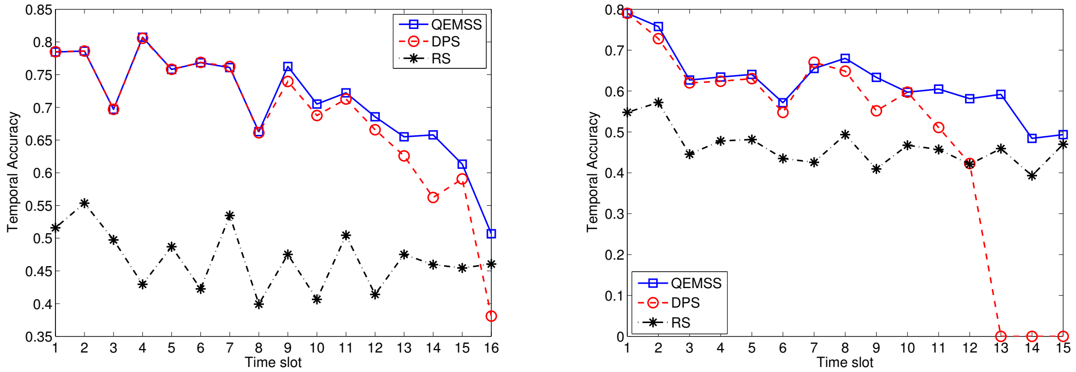
maximum achieved QoI value, in the case of unknown trajectories, is lower for all algorithms due to the prediction resulted errors. Yet, QEMSS still performs better than the other two evaluated schemes.

In Figure 3.2(b), we plot the evolution of the *maximum* achieved QoI level, realized by 400 participants available in the sensing area, per time slot. In the beginning of the sensing period, the performance of QEMSS is comparable to DPS. Though, the gap between QEMSS and DPS realized data quality level is widened in the end of the sensing period to reach up to 27%. This is due to the selection strategy. Indeed, the TS-based algorithm used by QEMSS scheme conducts extensive search at each sensing interval to diversify the selected participants which makes their devices' batteries last more. However, using the DPS greedy-based search, the same users are frequently selected to sense which harvest quickly their battery resources. This yields to less interesting selection choices at the end of the studied period, leading to a decrease in the performance achieved by DPS. The random selection (RS) has clearly and obviously the lower performance in terms of achieved QoI.

Similarly, we plot in Figure 3.3 the average value of data quality achieved by the following selection schemes: F-QEMSS, QEMSS, DPS and ES. This is performed to compare the realized QoI by all the studied alternatives including the two dealing explicitly with the fairness. We observe that the equal sensing (ES) algorithm reaches the lowest QoI values. Indeed, this assignment strategy assigns sensing tasks respecting a sensing budget, $ns_{max} = \frac{|A|}{n_p}$, which decreases as function of the number of participants n_p in the area A. This may yield to selecting users only among those who did not exceed their sensing budget and consequently affect the data quality of collected samples. As stated earlier, the DPS solution realizes less important data quality levels than our TS-based solutions, QEMSS and F-QEMSS. Moreover, the fair allocation scheme, F-QEMSS, reaches as important data



(a) Spatial accuracy - $n_p = 400$ (b) Spatial accuracy - $n_p = 200$



(c) Temporal accuracy - $n_p = 400$ (d) Temporal accuracy - $n_p = 200$

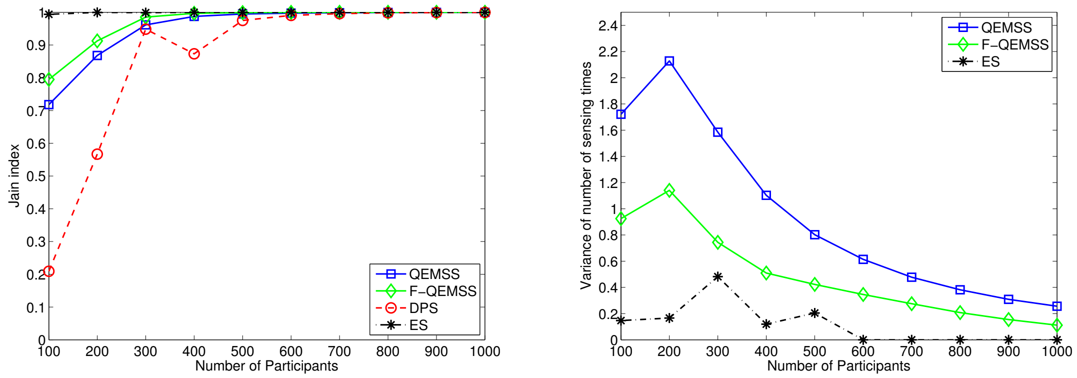
Figure 3.4: Spatial and temporal data accuracy levels

quality levels as achieved by the only maximization QoI scheme, QEMSS. This can be explained by the associated coefficients to the F-QEMSS objective function; $\mu_1 = \mu_2 = 0.5$, where we suggest to maximize the QoI measure and the fairness index equally.

3.7.4.2 Spatial and Temporal Accuracy

In this part, we measure the spatial and temporal accuracy of the different discussed schemes. That is, we run the different selection strategies to obtain the set of selected users and collect their reported samples respecting the mapped sensing sub-areas for each. Accordingly, we measure the data accuracy of the collected samples.

Spatial Accuracy: We plot in Figures 3.4(a) and 3.4(b) the spatial accuracy computed by the Manhattan distance defined in Equation (3.10). Note that the larger is this distance, the less accurate is the collected sample. We observe that the spatial accuracy of DPS method is decreasing as function of time and is barely not accurate at the end of the sensing period T. This conforms with the evolution of data quality shown in Figure 3.2.



(a) The Jain's fairness index

(b) The variance of sensing times number

Figure 3.5: Fairness measures for different selection schemes

On the contrary, QEMSS and Random Selection (RS) realize comparable spatial accuracy since they both cover most of the required regions.

Temporal Accuracy: For the "timeliness" criterion quantified by the utility function $U_t(x)$ defined by Equation (3.2), we observe in Figures 3.4(c) and 3.4(d) that the RS scheme is the less accurate selection strategy. In fact, selected participants by RS cover most of the targeted sensing area as shown by the spatial accuracy. However, they may collect samples in *random* instants which can be far from the required time of measurement. Even though DPS is as accurate as QEMSS for the first half of the sensing period, the error rate is leveled up to 62% by the end of T. This is also due to the limited energy resources of available users at this time period, which results in quasi-random assignment by DPS and thus samples with low accuracy.

3.7.4.3 Fairness Metrics

Through this first contribution, we managed to prove the efficiency of the proposed solutions, QEMSS and F-QEMSS, in terms of maximizing data quality levels. In the following, we investigate the fairness level of the generated scheduling by QEMSS and F-QEMSS when compared to state-of-the art assignment methods. Recall that F-QEMSS has succeeded to reach the same data quality level as its predecessor QEMSS.

Jain index: Figure 3.5 summarizes the evaluation of the fairness level realized by our proposed task assignment solutions compared to DPS and ES schemes. The line graph 3.5(a) shows the evolution of the Jain's index detailed in Equation (3.5). We remark that this fairness measure increases according to the number of available participants in the sensing area A. That means, a *fairer* allocation is rather possible when having an important number of participants to share the sensing load. The *fairest* method is proved to be the Equal Sensing

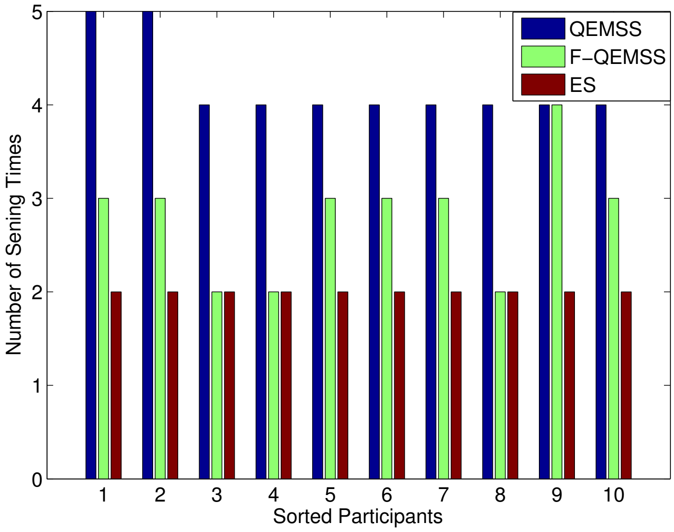


Figure 3.6: Sensing times for a 500-users density

(ES) as it reaches the maximum value 1 by affecting the same number of measurements to all participants, yet, to the cost of realizing low data quality as previously shown. On the contrary, the less fair scheme is the DPS algorithm. Particularly, in low dense areas, i.e., $n_p \leq 400$, more than 50% of participants may be unsatisfied when assigned sensing tasks. For the two proposed schemes of this chapter, QEMSS and F-QEMSS, Jain's index values vary between 0.78 to 1. Consequently, most participants are satisfied while dedicating their resources to achieve competitive data quality levels. It is worth mentioning that F-QEMSS realizes important fairness level while ensuring the same QoI achieved value as QEMSS. Indeed, F-QEMSS enhances the fairness by adding up to 10 more satisfied users among 100 compared to its predecessor.

Variance of Sensing Times: Additionally, we utilize the function defined by Equation (3.11) to investigate the efficiency of the three methods: QEMSS, F-QEMSS and ES in terms of the variance of the number of sensing. We plot the average variance for all schemes in Figure 3.5(b). Similar to the Jain's index results, the Equal Sensing scheme is observed to be the fairest method when reaching the lowest values of variance. Also, our assignment solution F-QEMSS performs well by decreasing the value of variance of the basic maximum QoI scheme, QEMSS, while preserving the same data quality. Particularly, for 200 participants, F-QEMSS and QEMSS achieve both 50% of the required level of data quality, however, F-QEMSS decreases the variance of ns by 1. Note that we did not plot the DPS algorithm results since it realizes very high values of variance.

To conclude, the proposed task allocation model, F-QEMSS, achieves high QoI level and a competitive system fairness level. This is also illustrated in the bar graph 3.6 showing the distribution of the number of sensing times over the 10 users who sense the most among 500. Naturally, the number of sensing times is distributed equally by the Equal Sensing scheme, but at the cost of very bad QoI values compared to QEMSS and F-QEMSS. Meanwhile, F-QEMSS outperforms QEMSS by assigning a lower number of sensing tasks to the same participants while preserving the same data quality levels.

3.8 Discussion

This chapter has tackled essentially the energetic and data quality issues in MCS campaigns. These issues are mainly encountered when collecting sensing data from all participants in a specific sensing region. Especially, when the MCS platform collects data with no consideration of users' handsets characteristics such as battery level, location precision or time of measurement to name a few. This yields to a heavy sensing process in terms of dedicated resources besides not providing accurate sensor data.

More precisely, we have studied how to prevent these challenges during the MCS task assignment phase. We suggested to consider registered users' movements in a sensing region in order to estimate their contributions quality. Also, we selected a subset of participants among all available ones who provide the same data quality level, yet, with lower energy cost level. Another energetic constraint was set in the form of a battery threshold level under which a participant handset is removed from the candidate list. This was performed in order to ensure to participants a normal use of their devices. Based on all these, we introduced two quality and energy-aware task assignment schemes for mobile crowdsensing systems: QEMSS and F-QEMSS. The basic idea is to quantify different QoI attributes and fairness by introducing adequate measures to each criterion. Furthermore, we formulated the corresponding optimization problems targeting the maximization of the system efficiency. For the different objectives, we made use of the Tabu-Search strategy to design our assignment algorithms. The former solution targets only maximizing data quality with energy and no-redundancy constraints. While the second proposal, F-QEMSS, aims at satisfying both participants and requesters by realizing a high quality sensing process with fairness consideration.

The evaluation of our schemes was conducted when compared to the DPS scheme [57] as well as two other baseline schemes. Simulation results show that both QEMSS and F-QEMSS have realized important QoI levels in high density areas compared to other benchmark schemes. Nevertheless, the proposed schemes have obtained a significant gain in the achieved QoI level and spatial and temporal accuracy of data in challenging situations such as low density sensing areas and/or low battery equipped participants at the end of a sensing period. In addition, when assessing the fairness of the different selection strategies, we demonstrated that our allocation model F-QEMSS realizes a high level of fairness while maintaining as important QoI level as its predecessor QEMSS. Hence, our first contribution can be considered as an effective task allocation scheme that scales an excellent trade-off of satisfying both requesters and participants in the crowdsensing process.

In summary, the proposed work in this chapter aims at optimizing the task assignment in mobile crowdsensing systems. However, the centralized approach adopted by the proposed solutions may raise new challenges. In particular, the assumption of knowing trajectories in advance is rather to generate efficient assignment schemes as benchmarks. While the results of prediction-based solutions were shown to be slightly less effective. In the next chapter, we target improving this crowdsensing scenario by focusing on a more interactive paradigm which is the participatory sensing. In such context, we seize the issue of learning users' arrival model to enhance the accuracy of collected measurements. Moreover, we propose to extend the task assignment phase in a distributed architecture in order to overcome the traditional issues related to centralized scenarios.

Chapter 4

Preference and Mobility-aware Task Assignment Schemes

Contents

4.1 Introduction 45
4.2 Motivations and Context 46
4.3 Existing Work on Distributed Crowdsensing 48
4.4 Distributed Task Assignment Problem Definition 49
4.5 Mobility-aware Task Assignment: MATA 52
4.6 Preference and Mobility-aware Task Assignment: P-MATA . . 56
4.7 Performance Evaluation 60
4.8 Discussion 66

4.1 Introduction

We introduced in the previous chapter two quality and energy-aware assignment schemes for mobile crowdsensing. The introduced solutions foresee participants movements to estimate their contributions' quality in terms of spatial and temporal accuracy. Besides, we have set an energy level threshold and a fairness measure to respect the normal use of users' devices. Proposed methods proved to be highly efficient especially when associated to known trajectories in advance which may not be feasible in large scale crowdsensing systems.

In this chapter, we target the *Sensing and Processing* phase of mobile crowdsensing. More precisely, we present our contribution on minimizing the average time of sensing, defined as makespan. In order to do so, we propose to assess the centralized task assignment and delegate it to individual participants. We suggest to add a distributed "support" phase where each sensing platform can assign extra tasks to some selected participants to be reallocated on the move when "encountering" other users. These participants, denoted as *requesters* throughout this chapter, perform distributive task assignment as two different policies. The first policy estimates users arrival, i.e., the inter-meeting time, and decides accordingly to whom to delegate tasks. The second variant accounts for the probability of acceptance/ rejection of the proposed assignment that we define as users preferences. Thus, each requester estimates a probabilistic inter-meeting time and assigns tasks to participants who minimize the average makespan. Both task assignment variants are conducted in offline and online modes. That means, requesters generate their sensing schedule in the beginning of each sensing period or whenever they encounter a participant, respectively.

In the following, we present the motivations behind adopting the distributed assignment for MCS in Section 4.2. In Section 4.3, we illustrate a literature review of existing works on distributed task assignment in MCS. Furthermore, we present our system model and we state the investigated problem in Section 4.4. The proposed assignment schemes are designed in details in Sections 4.5 and 4.6. Finally, we evaluate the performance of the two proposed variants on real-traces in Section 4.7. Discussions conducted based on obtained results are illustrated in Section 4.8.

4.2 Motivations and Context

Motivated by crowdsensing potential applications, researchers have widely studied the problem of sensing campaigns assignment among different users [52,57–59,64,67,68]. Along with this, we introduced in the previous chapter our works on quality and energy-aware allocation [16,17]. However, all the aforementioned schemes are central-based approaches where a unique entity in the cloud is responsible of assigning tasks to registered participants. In this context, various issues may arise during and ahead of the allocation phase.

First, the central platform may not be capable to predict all participants trajectories given their important number within a large scale crowdsensing scenario. Indeed, human mobility is hard to model especially for a long period. Besides, available analytic tools such as Markov model [77,79] or Bayesian inference [82] need a prior knowledge of certain previous states. To overcome this problem, authors [77,78] propose to identify common visited locations of each participant (work, home, restaurant ..) and assign accordingly time-tolerant tasks to be processed when reaching these locations. To do so, participants need to turn on their location sensors (GPS) and upload their trajectories. Yet, this comes with an important energy cost [3,59]. In addition, users are usually reluctant to share their locations for privacy reasons. Therefore, it would be better to delegate tasks to participants based on a local estimation of their movements rather than a global one.

Furthermore, participants may be registered within more than one crowdsensing platform in order to maximize their profit in terms of received "incentives". This might lead to time-overlapping tasks assignments and to a number of tasks exceeding their sensing or

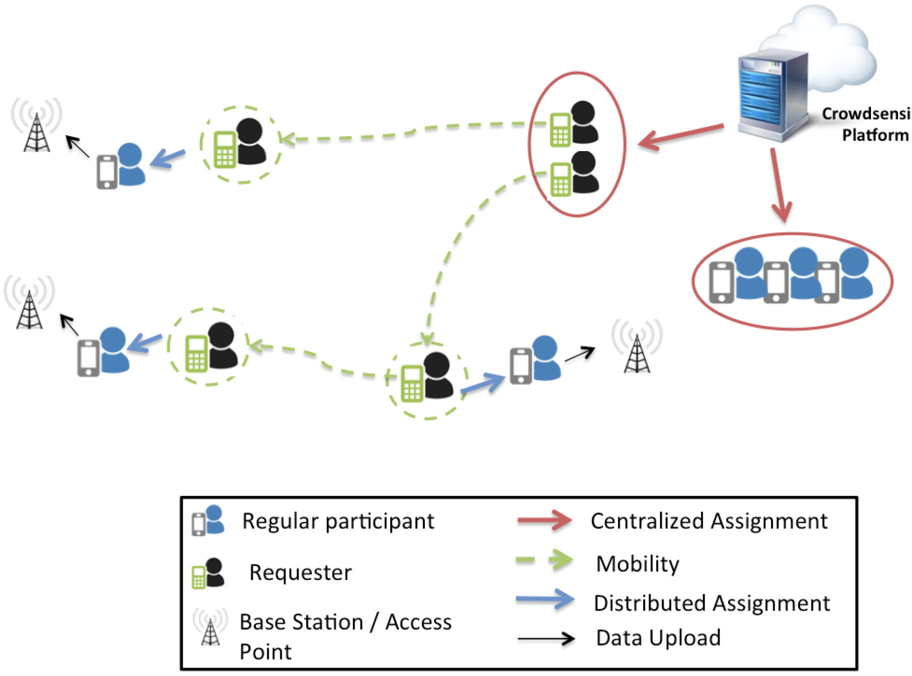


Figure 4.1: Distributed assignment in mobile crowdsensing

processing capabilities. As a consequence, participants with important sensing load may be unable to perform assigned tasks on time which results on reported samples with poor temporal and even spatial accuracy, i.e., low data-quality. This yields to low rewarding function of users' contributions thereby unsatisfied participants and requesters. Hence, it would be more appropriate to account for participants current assignment to better estimate their availability for sensing, thus a more accurate sensing process. Last but not least, some users may not be reachable during the allocation phase due to connectivity issues such as being outside Wi-Fi access point range or having limited 3G data budget. As a result, the MCS platform may miss recruiting participants with potential high quality contribution. Therefore, it would be beneficial to add a distributed phase to recruit as more possible participants as possible and optimize the quality of collected data.

With all this in mind, we design in this chapter a distributive crowdsensing "support" phase to better share the load among participants and consequently minimize the overall sensing and processing time. The identified *requesters* carry tasks to be assigned to adequate participants according to the adopted assignment policy. The latter collect and upload data via different networks depending on their 3G/4G budget and the data time-sensitivity as illustrated in Figure 4.1. Specifically, our main contributions are:

1. We model participants arrival and preferences models based on their historical encounters and we derive the expected time of sensing and processing.
2. We develop a Mobility-Aware Task Assignment scheme, MATA, based on only users arrival model and we introduce it in both offline and online assignment models.

3. We develop a Preference and Mobility-Aware Task Assignment scheme, P-MATA, which jointly considers users mobility and preferences. P-MATA is also introduced in both offline and online modes.

4.3 Existing Work on Distributed Crowdsensing

By far, several assignment schemes are proposed to optimize the crowdsensing process. Yet, few works have tackled this issue in a distributed based or semi-distributed way.

The Bubble-sensing model [83] is a *hybrid* alternative to the purely centralized schemes, i.e., based on distributed allocation of tasks but requires the presence of a central entity to maintain the sensing tasks. Participants, named as *bubble creators*, can create sensing areas at defined places of interests. They initiate a sensing request as a *bubble* defined by a geographical location, time and necessary sensors and broadcast it to potential *bubble carriers*, i.e., other participants, in a distributed fashion. Bubble carriers may then move to the task location, perform the sensing and report the collected data to the central *bubble server* to be retrieved further by the bubble creator. A major issue of this approach is that the persistence of bubbles during the sensing period is ensured by *bubble anchors* who maintain bubbles on behalf of their creators. This requires a limited mobility of different users and a persistent connectivity among them.

Different from this hybrid scenario, Cheung et al. [84] have designed a more self-organized crowdsensing system. The introduced Asynchronous and Distributed Task Selection (ADTS) algorithm aims to help users plan their tasks selection on their own. Accordingly, participants designate in a non-cooperative game paths that maximize their profit. Nevertheless, authors did not investigate the processing time of tasks. Note that users, even though being rewarded for their contributions, may be reluctant to perform long sensing campaigns that use up their devices batteries. In this context, Xiao et al. [85] studied the task assignment problem in Mobile Social Networks (MSN) with the aim to minimize the average sensing and processing time. Authors investigated users' encountering based on their historical traces. In light of this, data queriers can recruit participants to perform sensing tasks. The proposed methods are formulated as an offline (FTA) and online (NTA) assignment strategies. However, authors have only considered time-dependent crowdsensing, while in fact the location of collected samples matters as much. Hence, an assignment scheme should be based on users' mobility in terms of different locations rather than estimating only their meetings. Moreover, previous works did not consider the issue of participants' sensing preferences, i.e., the ability to accept or reject the assignment strategy. Such ability has been very recently discussed by authors in [66] in order to select the *workers* who maximize an expected sum of service quality. The proposed framework, Crowdlet, is based on dynamic programming which enhances distributed self-organized mobile crowdsourcing. Yet, the time cost of conducting such quality-aware sensing was not investigated. From this perspective, we assess in this chapter the distributed task assignment problem while accounting for users mobility and sensing preferences.

4.4 Distributed Task Assignment Problem Definition

In this section, we describe the MCS distributed scenario. We also design user arrivals and preferences models and we formulate the problem that states the goal of our proposal.

4.4.1 Semi-Distributed Crowdsensing System Overview

In this chapter, we argue that a crowdsensing system should be split into two phases: a centralized and a distributed one. In the first phase, the central unit proceeds with a task allocation comparable to the one introduced in the previous chapter [16, 17]. Nevertheless, the platform can designate some participants to continue assigning tasks in a distributed fashion as illustrated in Figure 4.1. In this work, we are interested in the latter phase as a dynamic *Participatory Sensing* (PS) paradigm given the involvement of participants during the assignment and sensing process.

We consider N mobile users in a crowdsensing area divided into sub-regions, defined as compounds $C = \{k, k \in 1..n_C\}$. These compounds are with different characteristics thereby with various users mobility behavior. For instance, an eating area is rather dense during feeding times with a very low mobility whereas a shopping area is both a crowded and dynamic area. Accordingly, users move with various speed values between different compounds and can be at a given time with a probability q_k in a compound k. For simplicity, let $R = \{r_1, r_2, \ldots, r_{n_r}\}$ be the set of *requesters*, i.e., the participants responsible for distributed assignment, and $P = \{p_1, p_2, \ldots, p_{n_p}\}$ the set of regular participants with $n_r + n_p \leq N$. We assume that users encounter when they move around in a same compound k and are close enough to allow direct communications such as Device-to-Device (D2D) communications. More specifically, a requester selects a participant among the ones present in his device communication range and delegates to him a set of sensing tasks. Further, collected data samples can be uploaded via Wi-Fi within access point range or data networks for only time-sensitive data. Without loss of generality, we assume that the inter-meeting time of a requester with a user is enough for exchanging tasks. In order to better estimate this time, we investigate hereafter the arrival model of users.

4.4.2 Users' Arrival Model

The task-oriented participatory sensing is critically depending on users' mobility. Hence, we base our work on a widely-used mobility model in mobile social networks [86, 87]. We assume that the inter-meeting time between a requester r_i and a participant p_j follows an exponential distribution with a contact rate parameter λ_{p_j}. As a result, the inter-encounter time of a requester with two consecutive participants follows also an exponential distribution with parameter $\lambda = \sum_{p_j \in P} \lambda_{p_j}$. Thus, the arrival of participants to a requester follows a Poisson process. Moreover, we suppose that λ_{p_j} can be derived from historical encounters between a requester and each participant as stated in [66, 85]. For simplicity, we consider users encounter only while being in the same compound k. We examine the probability of each user to be in a compound and define the inter-meeting time as follows:

Definition 4.1. The inter-meeting time between a requester r_i and a participant p_j in a compound k is an exponential distribution with rate parameter $F_{(k,p_j)} = q_k(r_i)q_k(p_j)\lambda_{p_j}$

$$A_{i,j,k} = \int_0^{\infty} F_{(k,p_j)} t e^{-F_{(k,p_j)}t} \,\mathrm{d}t = \frac{1}{F_{(k,p_j)}} = \frac{1}{q_k(r_i)q_k(p_j)\lambda_{p_j}}$$

where $q_k(r_i)$, $q_k(p_j)$ are the probabilities of a requester r_i and a participant p_j being in a compound k, respectively.

Based on the above defined inter-encounter time, each requester can estimate the location and the time to meet a participant. Accordingly, the requester can perform the task assignment and generates the set of users to whom to delegate the tasks. Such mobility-aware assignment strategy is presented in details in Section 4.5.

4.4.3 Users' Preferences Model

The aforementioned Poisson process models the arrival of participants to a requester but does not take into account their preferences as they can accept or reject the proposed assignment. Therefore, we define here a user acceptance probability to perform assigned sensing tasks and we denote it as p_a. Thus, the probability of rejection is defined as $p_r = 1 - p_a$. In practice, the former factor can be calculated from historical statistical data or requesters' experiences as introduced in [85]. Based on this probability, the acceptance of a task by a participant can be modeled as a Bernoulli process. That is, for each participant p_i, the set of answers are associated with a random variable X in $\{0, 1\}$, where $X = 1$ with probability p_a and $X = 0$ with probability $p_r = 1 - p_a$. Accordingly, we consider a requester encountering n_p participants and derive the number of those who accept to participate as a Binomial distribution, $B(n_p, p_a)$. Therefore, we extend the arrival model of these participants as a composition of the Poisson process and the Binomial distribution which leads to a Poisson distribution with parameter $(\beta_k = p_a \times F_{(k,p_i)})$ [88].

It is worth noting that we name a participant who is predicted to accept his assignment as "positive" participant. In Lemma 4.2, we proceed to compute the necessary time to encounter such participant.

Lemma 4.2. *The mean time for a requester r_i till meeting a "positive" participant p_j within n time slots is:*

$$\Pi_{i,j,k} = \Big[\sum_{l=1}^{n} p_r^l p_a \frac{1}{q_k(r_i)q_k(p_j)\lambda_{p_j}}\Big]'$$

Proof. We assume that during the assignment phase, a requester r_i is likely to meet any participant p_j more than once. However, the latter accepts the assigned tasks with p_a. Suppose that a participant accepted tasks within n meetings, thus, the mean time of meetings this "positive" participant can be expressed as: $p_a A_{i,j,k} + p_r p_a A_{i,j,k} \times 2 + \ldots + p_r^{n-1} p_a A_{i,j,k} \times n = p_a A_{i,j,k}(1 + 2p_r + 3p_r^2 + \ldots + np_r^{n-1})$. In this expression, each term of the sum can be a derivative of p_r^l with $l \in \{1, \ldots, n\}$. Hence, we denote by $[.]'$ the derivative operator and model $\Pi_{i,j,k}$ as $p_a A_{i,j,k}\Big[p_r + p_r^2 + \ldots + p_r^n\Big]' = p_a A_{i,j,k}\Big[\sum_{l=1}^{n} p_r^l\Big]'$. Finally, we substitute $A_{i,j,k}$ with its corresponding expression from Definition 4.1 to obtain the expression of $\Pi_{i,j,k}$. ☐

The probabilistic inter-meeting time presented above is utilized to build a preference and mobility aware assignment strategy. We illustrate this assignment policy in Section 4.6 and compare it to the pure mobility-aware variant to evaluate their efficiency.

4.4.4 Problem Definition

In this chapter, our aim is to minimize the overall sensing and processing time when relying on distributed assignment phase. We refer to the scenario of a requester r_i carrying m sensing tasks to be assigned to encountered participants. Let $S = \{s_1, s_2, \ldots, s_m\}$ be the set of sensing tasks that a requester intends to assign. We assume that these tasks differ in terms of average workload, i.e., the time of sensing and processing, that we denote as $\{\tau_1, \tau_2, \ldots, \tau_m\}$. Also, we define by makespan of a task $s_i \in S$, the time of being assigned and processed and we denote it as $M(s_i)$. Note that we exclude here the reporting phase since we assume it can be instantaneous. Furthermore, we define by assignment strategy $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_{n_p}\}$ the resulted schedule of sensing. We suggest that each task is assigned only to one participant, i.e., $\gamma_i \cap \gamma_j = \emptyset \ \forall p_i, p_j \in P$. Moreover, if a participant $p_i \in P$ has not received any task to be processed for a given period of assignment from encountered requesters, then his assignment set $\gamma_i = \emptyset$. With all this in mind, we define the average makespan of all assigned tasks to different participants as stated in [85]:

Definition 4.3. The average makespan is the average time of assigning, sensing and processing tasks by all encountered participants can be formulated as:

$$AM(\Gamma) = \frac{1}{m}\sum_{s_i \in S} M(s_i)|_\Gamma$$

In the following, we advocate how to minimize the average makespan expressed in Definition 4.3 in two variants of scenarios.

First, we suppose that all encountered participants in a certain compound accept *unconditionally* the suggested tasks. Hence, we refer to users arrival model of Definition 4.1 and formulate the objective function. We introduce this variant as the Mobility-Aware Task Assignment, MATA, and we develop two different modes; offline and online assignment strategy. The former mode indicates that a requester, and based on the expected arrival time of each participant, decides his assignment strategy which we denote throughout this chapter as Γ_{AF}. Once the selection is done, the requester assigns tasks to the designated participants when he encounters them. As for the online mode, a requester starts his assignment strategy computation only when he meets a participant. If the met user is among the selected ones, he receives his assignment. Otherwise, the requester keeps the set of tasks to be assigned for a next meeting. This latter strategy enhances the dynamic assignment scheme whose result is denoted as Γ_{AN}.

For the second variant, we jointly take into account the mobility of users and their sensing preferences. Then, we refer to the acceptance model of Lemma 4.2 and define the Preference and Mobility-Aware Task Assignment, P-MATA. Similarly, we develop an online and offline assignment strategies that we denote by Γ_{PN} and Γ_{PF}, respectively. Note that the offline and online assignment principles are the same for both variants. However, the average makespan expression varies respecting the probabilistic nature of users inter-meeting time in the preference model. Sections 4.5 and 4.6 illustrate the proposed distributed assignment schemes, MATA and P-MATA, while adapting the average makespan expression to each variant and mode.

4.5 Mobility-aware Task Assignment: MATA

Essentially, the assignment of tasks conducted by each requester r_i targets the minimization of the average makespan expressed in Definition 4.3. However, this formula varies as a function of the estimated users arrival model. In this section, we focus on users mobility and we ignore, for now, their ability to reject sensing campaigns. As a result, we formulate the average makespan for the mobility-aware distributed assignment and we introduce two greedy-based offline and online algorithms.

4.5.1 Offline Mode

In the offline case, we assume that each requester r_i can compute, based on his statistical historical records [85], the expected inter-meeting time with any participant p_j using the expression of Definition 4.1. As a result, we derive first the expression of the average makespan of an offline assignment mode then we describe the proposed algorithm aiming at minimizing this expression.

4.5.1.1 Offline Average Makespan

Let $\Gamma_{AF} = \{\gamma_1, \gamma_2, \ldots, \gamma_n\}$ be the assignment strategy decided by a requester r_i regarding n potential encountered participants. Correspondingly, we express the average makespan of all tasks in the following theorem:

Theorem 4.4. *The average makespan of m tasks in an offline mobility-aware task assignment strategy, $AM(\Gamma_{AF})$, is expressed as follows:*

$$AM(\Gamma_{AF}) = \frac{1}{m} \sum_{j=1}^{n} \sum_{l \in \gamma_j} \left(\frac{1}{q_k(r_i) q_k(p_j) \lambda_{p_j}} + \tau_l \right)$$

Proof. Given that γ_j is the set of tasks to be assigned to a participant p_j and τ_l is the workload of the task $s_l \in \gamma_j$. Then, the makespan of all tasks is the sum of workloads. Plus, we consider the necessary time to meet the participant in question which leads directly from the arrival time computed by Definition 4.1 as $\frac{1}{q_k(r_i) q_k(p_j) \lambda_{p_j}}$. Finally, we generalize this expression for all n met participants to consider all m tasks held by a requester and obtain the formula of Theorem 4.4. □

4.5.1.2 Offline Algorithm: MATAF

In order to minimize the average makespan expressed in Theorem 4.4, we propose a greedy-based solution that we denote as MATAF for Mobility-Aware Task Assignment oFfline. The basic idea of this assignment scheme is that each requester r_i needs to compute an Expected Sensing Time (EST) for each participant $p_j \in P$. This factor includes the inter-encounter time to meet p_j in a certain compound k, i.e., $A_{i,j,k}$ plus the sum of the possible assigned tasks loads. To this purpose, we consider first that all tasks held by requesters are sorted in an ascending way. That means, $\forall s_i, s_j \in S$, if $i \leq j$ then $\tau_i \leq \tau_j$. Also, we initialize all participants Expected Sensing Time to their inter-meeting time with the current requester r_i, $EST_j = A_{i,j,k}, \forall j \in 1 \ldots n_p$. Next, for each task $s_l \in S$, we look for the participant with the smallest EST_j. We assign the current task to the selected participant and update the latter $EST_j = A_{i,j,k} + \tau_l$. We repeat this strategy until assigning all tasks of requester r_i or the expiry of the sensing period.

We remind the fact that inter-meeting time varies as a function of the compound k given that users have different mobility behavior in each compound. For example, a user may stay a long period in a compound representing his work/housing area but spends less time in another compound modeling an eating area. Consequently, a requester may meet a participant only in certain compounds and more than once. Therefore, we compute the k possible inter-meeting times between each requester and the rest of users and we map the corresponding result into a $k \times n_p$ matrix where each row represents all participants EST in a compound; $EST|_k = [EST_1, EST_2, \ldots, EST_{n_p}]$. As described above, EST_j is initialized

Algorithm 3 MATAF Assignment Algorithm

```
Require: Requester $r_i$, Set of sensing tasks $S = \{s_1, s_2, \dots, s_m : \tau_1 \leq \tau_2 \leq \dots \leq \tau_m\}$,
    Participants $P = \{p_1, p_2, \dots, p_{n_p}\}$, Matrix of Expected Sensing Time $EST$.
Ensure: Assignment strategy $\Gamma_{AF} = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$
for $s_l \in S$ do
    $min_e \leftarrow \infty$
    for $k \in C$ do
        for $p_j \in P$ do
            $A_{i,j,k} = \int_0^\infty F_{(k,p_j)} t e^{-F_{(k,p_j)} t} \,\mathrm{d}t = \frac{1}{F_{(k,p_j)}} = \frac{1}{q_k(r_i) q_k(p_j) \lambda_{p_j}}$
            $EST_{k,j} = A_{i,j,k} + \tau_l$
        end for
        $[min_k, j_k] \leftarrow argmin(EST|_k)$
        if $min_k \leq min_e$ then
            $min_e \leftarrow min_k$
            $j_{min} \leftarrow j_k$
        end if
    end for
    Assign the task: $\gamma_{j_{min}} = \gamma_{j_{min}} + \{s_l\}$
    Update EST: $EST_{k,j_{min}} = EST_{k,j_{min}} + \tau_l$; $\forall k \in C$
    Update the set of tasks: $S = S \setminus \{s_l\}$
end for
Return $\Gamma_{AF}$
```

to $A_{i,j,k}$ and updated every time a new task is assigned to participant p_j. The detailed mobility-aware offline assignment scheme, MATAF, needs to be run by each requester $r_i \in R$ as detailed in Algorithm 3.

4.5.2 Online Mode

The second mode of our Mobility-Aware Task Assignment scheme is an oNline strategy, MATAN. The principle here is that each requester r_i moves around in a specific compound and starts requesting crowdsensing support every time he gets close to a participant p_j. In such a case, we opt for an instantaneous inter-encountering time which results in a different expression of the average makespan to be minimized. Herein, we develop this new expression and we present the respective online algorithm.

4.5.2.1 Online Average Makespan

The assignment phase in the online mode is launched at every encounter. First, the requester r_i asks the encountered participant p_j to update his expected sensing time, EST_j. However, the latter may have previously received other assignments from other requesters and currently processing them. Thereupon, we consider merging the expected sensing time to an Instant Sensing Time (IST) which computes only the rest of workload held by the current participant p_j as follows:

$$IST_j = \sum_{l \in \gamma_j} \tau_l - T_{j,\gamma_j} \quad \forall j \in 1 \dots n_p. \tag{4.1}$$

where $T_{j,\gamma_j} = t_c - t_{s,\gamma_j}$ is the time elapsed since participant p_j has started performing his previous assignment γ_j, t_c is the current time and t_{s,γ_j} is the starting time of γ_j.

We refer to the Instant Sensing Time expression presented above in order to deduce the online average makespan as detailed in Lemma 4.5.

Lemma 4.5. *The average makespan of m tasks in an online mobility-aware task assignment strategy, $AM(\Gamma_{AN})$, is expressed as follows:*

$$AM(\Gamma_{AN}) = \frac{1}{m} \sum_{j=1}^{n} \sum_{l \in \gamma_j} IST_j + \tau_l$$

4.5.2.2 Online Algorithm: MATAN

In this mobility-aware online mode, each requester r_i performs an assignment strategy comparable to the offline method while setting EST_j as IST_j for the encountered participant p_j. The resulted assignment strategy Γ_{AN} contains the list of selected participants and their associated tasks. If the encountered participant p_j is present in this list, he receives his corresponding assignment. For the rest of users, this assignment is temporary and can change based on the next encountered participants. That means, if the next met participant presents lower IST, then all the assignment strategy Γ_{AN} can be modified accordingly. Similar to the offline mode, MATAF, we assume that the sensing tasks are sorted in an ascending way as a function of their loads. Moreover, the requester adopts the same principle of assigning the current task to the participant with the smallest EST. In general, each requester r_i proceeds to run the online mobility-aware assignment method detailed in Algorithm 4.

4.5.3 After Thoughts

For the developed mobility-aware assignment (MATA), we assume that users can estimate their inter-meeting time based on historical encounter records as suggested by comparable allocation schemes in the literature [66, 85]. Nonetheless, different from these works, we assess the prediction of users arrival by including location-aware model rather than only time-dependent one. This was achieved by introducing first the probability of users being in different compounds, q_k. Also, we model users encountering when only being in a same compound which is highly realistic given that two users can not meet if being in different places such as a restaurant and an office.

In the following, we study a key issue in assignment schemes, *Do all users accept the proposed tasks?* Hence, we suggest empowering participants by a “preference” option.

Algorithm 4 MATAN Assignment Algorithm

```
Require: Requester $r_i$, Set of sensing tasks $S = \{s_1, s_2, \ldots, s_m : \tau_1 \leq \tau_2 \leq \ldots \leq \tau_m\}$,
    Participants $P = \{p_1, p_2, \ldots, p_{n_p}\}$, Matrix of Expected Sensing Time $EST$.
Ensure: Assignment strategy $\Gamma_{AN} = \{\gamma_1, \gamma_2, \ldots, \gamma_n\}$
When the requester meets a participant $p_j$ in a compound $k \in C$
Set $EST_{k,j} = IST_{k,j}$
$P = P \setminus \{p_j\}$
for $s_l \in S$ do
    $min_e \leftarrow \infty$
    for $k \in C$ do
        $[min_k, j_k] \leftarrow argmin(IST_{k,j} + EST|_k)$
        if $min_k \leq min_e$ then
            $min_e \leftarrow min_k$
            $j_{min} \leftarrow j_k$
        end if
    end for
    if $(j_{min} = j)$ then
        Assign the task to this participant: $\gamma_j = \gamma_j + \{s_l\}$
        Update EST: $EST_{k,j} = EST_{k,j} + \tau_l$; $\forall k \in C$
        Update the set of tasks: $S = S \setminus \{s_l\}$
    else
        Temporary assignment : $\gamma_{j_{min}} = \gamma_{j_{min}} + \{s_l\}$
        Temporary Update of EST: $EST_{k,j_{min}} = EST_{k,j_{min}} + \tau_l$; $\forall k \in C$
    end if
end for
Return $\Gamma_{AN}$
```

That means, a requester needs to consider the fact that the potential participant may accept or reject assigned tasks. We model such preferences, the resulting average makespan expression and the corresponding assignment scheme in details in the next section.

4.6 Preference and Mobility-aware Task Assignment: P-MATA

Previously, we introduced the variant of our assignment scheme which foresees users arrival model and decides accordingly to whom to delegate tasks. Yet, this decision policy did not tackle the question of preferences for each participant which may impact their commitment to the crowdsensing process. In fact, any participant registered within a crowdsensing platform may be reluctant to perform sensing tasks due to current sensing workload, non availability or other constraints in terms of energetic resources or data budget to name a few. Considering this, we develop a preference and mobility aware task assignment, P-MATA, which estimates participants acceptance thereby guarantees their commitment as well as their satisfaction while performing sensing campaigns. More specifically, we refer to the model of acceptance probability described in Lemma 4.2 and we introduce an offline and online Preference and Mobility-Aware Task Assignment algorithms.

4.6.1 Offline Mode

In this mode, we recall the assumption of requesters who estimate “a priori” the different inter-meeting time of participants and select those who can minimize the average time of sensing and processing. Nevertheless, we account for users preferences and formulate first the new derived expression of the average makespan. Besides, we update the previous proposed offline mobility-aware assignment algorithm to cope with the new requirements.

4.6.1.1 Offline Average Makespan

In the beginning, we remind that we define a participant who may potentially accept his assignment based on the associated probability of acceptance p_a as a “positive” participant. In order to derive the average makespan of potential assigned tasks to “positive” participants, we need to jointly take into consideration the mobility and the acceptance probability of users. Hence, we consider $\Gamma_{PF} = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$ as the assignment strategy decided by a requester r_i regarding n potential “positive” participants. Also, we refer to Lemma 4.2 to estimate for requester r_i the next time slot in which he can meet a “positive” participant p_j and we propose the following theorem:

Theorem 4.6. *The average makespan of m tasks in an offline preference and mobility-aware task assignment strategy $AM(\Gamma_{PF})$ is expressed as follows:*

$$AM(\Gamma_{PF}) = \frac{1}{m}\sum_{j=1}^{n}\sum_{l\in\gamma_j}\Big(\Big[\sum_{x=1}^{t} p_r^x p_a \frac{1}{q_k(r_i)q_k(p_j)\lambda_{p_j}}\Big]' + \tau_l\Big).$$

Proof. Let γ_j be the set of tasks to be assigned to a participant p_j with τ_l is the workload of the task $s_l \in \gamma_j$. The makespan of all tasks is the sum of workloads to each participant plus the passed time before the first acceptance. The latter factor can be deduced from Lemma 4.2 as $[\sum_{x=1}^{t} p_r^x p_a \frac{1}{q_k(r_i)q_k(p_j)\lambda_{p_j}}]'$ where x is the number of time slots to get an acceptance from a participant and $[.]'$ is the derivative operator. Similarly to the previous approach, we generalize this expression for all n met participants to consider all m tasks held by a requester, so we obtain the formula of Theorem 4.6. □

4.6.1.2 Offline Algorithm: P-MATAF

In the aim of minimizing the average makespan expressed above, we opt for the same policy as in MATAF in which we introduce users probability of acceptance/rejection. In other terms, we look for the corresponding assignment strategy $AM(\Gamma_{PF})$ that minimizes this entity. To this purpose, each requester r_i needs to proceed according the following steps:

1. Generate the acceptance model of each participant p_j based on estimated sensing preferences from historical records, p_a and p_r.

2. Compute the inter-encounter time $\Pi_{i,j,k}$, as in Lemma 4.2, to generate the list of potential "positive" participants, i.e, the ones with the highest probabilities p_a.
3. Set the Expected Sensing Time of all participants to $\Pi_{i,jk}, \forall p_j \in P$ and look for the smallest EST_j.
4. Select the corresponding participant and update his $EST_{j_{min}}$ as in Algorithm 3 until assigning all tasks.

We account also, as in the mobility-aware offline mode, for the differences among sensing sub-regions. As a matter of fact, users' preferences also may depend on their current location. For instance, if a participant p_j is within an area of work, he is more likely to reject *heavy* tasks to not disturb his working process. Therefore, we compute the various probabilistic arrival model of "positive" participants in each sensing sub-region (compound) and we map a $k \times n_p$ matrix of expected sensing time as detailed in the previous sections.

Finally, we need to accentuate the fact that, in the P-MATAF solution, the preference of a participant p_j is only considered in the pre-assignment phase. More precisely, a requester r_i estimates the behavior of each participant based on his historical data to generate the list of selected users with assigned tasks Γ_{PF}. However, when encountering designated participants, the requester transmits to each participant p_j his assignment γ_j and does not wait for the confirmation or rejection. The error in such estimation may result in a non perceived rejection and consequently non performed tasks. Therefore, we believe that the online preference solution is more realistic to overcome this issue. This is described below.

4.6.2 Online Solution: P-MATAN

Estimating users' preferences based on only their historical records is rather not accurate enough. These parameters are highly depending on a participant current location, resources and time of receiving tasks among other conditions. Thus, we opt for an oNline Preference and Mobility-aware Task Assignment, P-MATAN, which considers an updated status of each candidate before assigning sensing tasks.

First, we refer to the formulation of Lemma (4.5) in order to model the online average makespan given that the Instant Sensing Time (IST) is the same for both mobility-aware and joint preference and mobility aware variants of assignment. Furthermore, we conduct an online assignment method comparable to MATAN but with a considerable interaction between requesters and participants. More precisely, we require that whenever a requester encounters a participant, he needs to compute the list of selected participants based on their expected sensing time as described in Algorithm 4. Nonetheless, if the encountered participant is among the resulting assignment strategy Γ_{PN}, the requester must wait for his confirmation before updating his list of tasks. As a consequence, the design of the

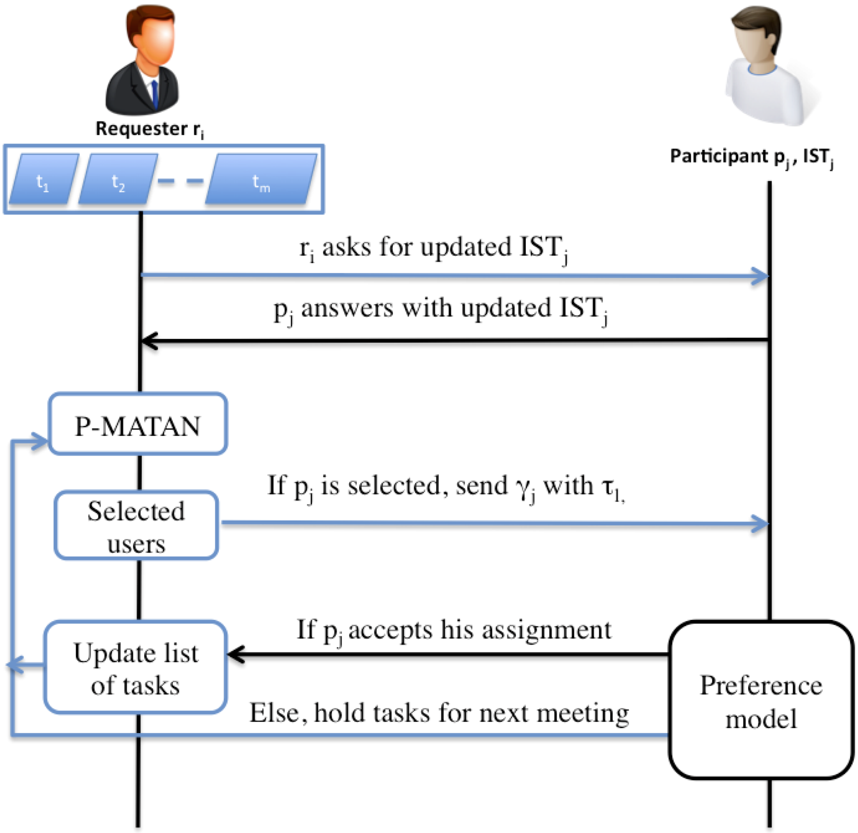


Figure 4.2: P-MATAN illustrative example

proposed online task allocation can be described as two phases: an online selection phase and an after-decision one.

Online selection phase: A requester r_i moves around and travels between different compounds $k \in C$. Whenever r_i encounters a participant p_j, he asks him to update his Instant Sensing Time IST_j based on Equation (4.1) in order to estimate the necessary time to perform the proposed assignment. After receiving the participant's IST, the requester conducts an online greedy-based selection as in Algorithm 4. Hence, if the current participant p_j is identified among the list of selected users, i.e., $\gamma_j \neq \emptyset$, the requester needs to send the designated tasks along with their workloads and wait for the participant confirmation.

After-decision phase: The encountered participant decides to accept or reject such assignment based on his associated probability of acceptance p_a and sends back his response to the requester r_i. If the participant p_j has accepted this assignment, the list of tasks is updated as well as his IST_j. Otherwise, the requester r_i holds the corresponding tasks and continues moving until encountering another participant. This second phase highly impacts the time of assignment, since the less rejections a requester receives, the faster the online assignment is conducted. To avoid such an issue, each requester stores a list of met participants to not consider a previous encountered user in the current online assignment.

The illustrative diagram given in Figure 4.2 details the steps of the preference and mobility-aware task assignment online mode introduced in this work.

4.6.3 After Thoughts

In the preferences and mobility-aware assignment policy, we suggest to *respect* participants willingness to perform or not the proposed sensing tasks. Particularly, we estimate users' acceptance probability and the resulted inter-meeting time in two different methods:

- In the offline mode, P-MATAF, it is up to the requester to estimate encountered participants acceptance probabilities based on his historical records, then compute the inter-meeting time for "positive" ones. That means, a participant preference is deduced based on his previous *choice*.
- In the online mode, P-MATAN, the requester asks for encountered participants' ISTs and designate accordingly those who minimize the average makespan. After receiving his assignment, a participant accepts or rejects tasks based on his current acceptance probability which is the more updated one.

In order to compare the efficiency of all introduced assignment algorithms, we propose to conduct simulations on real mobility traces as detailed in the following section.

4.7 Performance Evaluation

We evaluate the performance of the proposed task assignment variants, MATA and P-MATA, thus accounting for their offline and online solutions, through extensive real-traces based simulations. Furthermore, we dedicate a first phase of this evaluation to investigate the possible selection policies that a MCS platform can conduct to designate requesters among registered users. In the following, before discussing results, let us first describe the utilized mobility traces, the simulation settings as well as the evaluation metrics.

4.7.1 Real Traces

We opt for real mobility traces to design our crowdsensing scenario. Hence, we refer to two well-known user traces within a campus [89]. These traces include daily GPS track logs from two university campuses; North Carolina State University (NCSU) and Korea Advanced Institute of Science and Technology (KAIST). For the first trace, GPS readings are collected at every 10 seconds and recorded into a daily track log by 34 randomly selected students who took a course in the computer science department. Thus, we divide the corresponding area into four major compounds where we consider the two densest as computer science labs and the two other as food and administrative areas, respectively. As for KAIST trace, GPS readings were sampled each 10 seconds as well by students living in the dormitory of the campus between 2006 and 2007. Hence, we notice a higher density and lower speed compared to the first trace. We subdivide KAIST campus area also into 4 compounds; two dormitory sections with the highest densities, one food area and one studying area. According to each trace characteristics, we estimate users probabilities to be in different compounds q_k. Furthermore, we compute the inter-meeting time parameter between two users as $\lambda_i = N_i/T$, where N_i is the total number of encountering times with a distance set at $10m$ and T is the total duration.

4.7.2 Requester Selection

We set the number of requesters to be $\simeq 20\%$ of the total number of users in each trace. As a result, the number of requesters in the NCSU trace is equal to 7 while the one in the KAIST trace is equal to 20, as detailed in Table 4.1. Also, we investigate three different selection methods that a crowdsensing platform can use to identify requesters among the different registered participants. We propose to designate requesters *randomly*, by considering those with the highest estimated number of meetings λ_i or among the *fastest* mobile users. The selection of requesters is very important since it may highly impact the performance of our assignment schemes. For example, when selecting requesters randomly, we had *bad* ones, i.e., requesters that rarely encounter other users due to their mobility behavior and consequently, they can not delegate some of their sensing campaigns. This was the case of 2 requesters among the randomly selected ones in the NCSU campus trace, as illustrated in Table 4.1.

4.7.3 Simulation Settings

To simulate our participatory sensing scenario, we generate a set of sensing tasks S to be assigned by each of identified requesters to encountered participants. We vary the number of tasks and associated workloads to study their impact on our proposed assignment schemes. The number of tasks is then selected from $\{20, 40, 60, 80, 100\}$ and the average workload of all tasks τ is selected from $\{1, 3, 5\}$ (hours). Moreover, we vary the workload among different tasks by randomly selecting each task load, τ_j in $[0, 2\tau]$. Furthermore, we generate for each participant p_j the associated probability of acceptance p_a. Based on this probability, we determine also the set of possible *answers* an_{p_j} as a Bernoulli variable. The latter is rather used in P-MATAN. Accordingly, a participant accepts his assignment, i.e., $an_{p_j} = 1$, with probability p_a and rejects it with probability $p_r = 1 - p_a$ $(an_{p_j} = 0)$.

Table 4.1: Traces characteristics

Trace	Length	Requesters	Participants	Bad Requesters
NCSU	22 (h)	7	27	2
KAIST	24 (h)	20	72	0

Simulations are conducted under the network simulator ns-3.24. We run two different groups of tests, first while varying the number of tasks and setting the average workload to 1h and second while varying the average workload of tasks and fixing the number of tasks to 20. Within each group, 30 runs are performed for each trace (NCSU/ KAIST) and each assignment scheme: MATAF, MATAN, P-MATAF and P-MATAN. Results are illustrated and detailed herein.

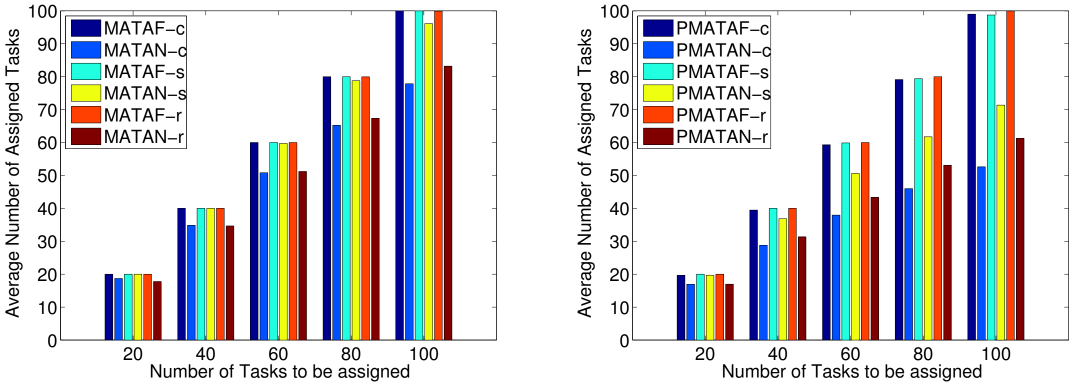
4.7.4 Evaluation Results

First, we aim to highlight the most efficient requester selection policy among the proposed ones. Thus, we compare the number of assigned tasks achieved by different requesters selected either *randomly*, *by-contact* or *by-speed*. Further, we proceed with comparing the achieved makespan by each assignment scheme. Finally, we plot the number of rejected tasks to analyze better the obtained results.

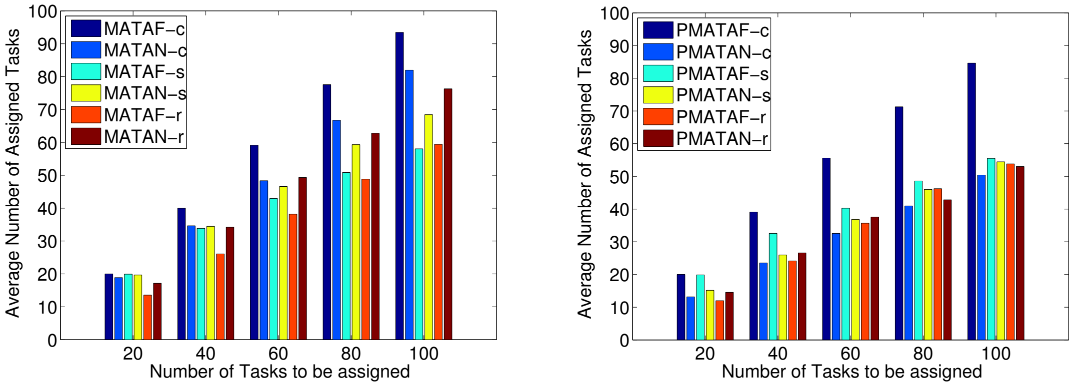
4.7.4.1 Average Number of Assigned Tasks

As a first evaluation metric, we start by measuring the achieved number of assigned tasks by each assignment scheme while varying the requesters selection policy. Therefore, we run simulations with different number of tasks and an average workload $\tau = 1$h on the two considered mobility traces KAIST and NCSU. Results illustrated in Figure 4.3 show that the offline mode achieves the highest values of assigned number of tasks for all requesters selection policies and for both MATA and P-MATA variants. Regardless, the preference and mobility aware variant, P-MATAF, realizes slightly lower values since it considers the ability for a participant to reject the assignment, which reduces the total final number of assigned tasks. Moreover, for the NCSU trace, the number of requesters and users in general, 7 and 34, is smaller than those in the KAIST trace, which limits the availability of participants for sensing, which results on lower number of assigned tasks.

Similarly, we investigate the distribution of the assigned number of tasks for various requesters selection policies while setting the required tasks number to 20. The results in Figure 4.4 conform with our observations. That is, the offline modes of mobility-aware assignment MATA for both traces outperform the online ones by assigning all and 80% of sensing tasks for the KAIST trace and the NCSU trace, respectively. Moreover, the distribution of assigned tasks by P-MATA describes the competitive results achieved by



(a) MATA for KAIST (b) P-MATA for KAIST



(c) MATA for NCSU (d) P-MATA for NCSU

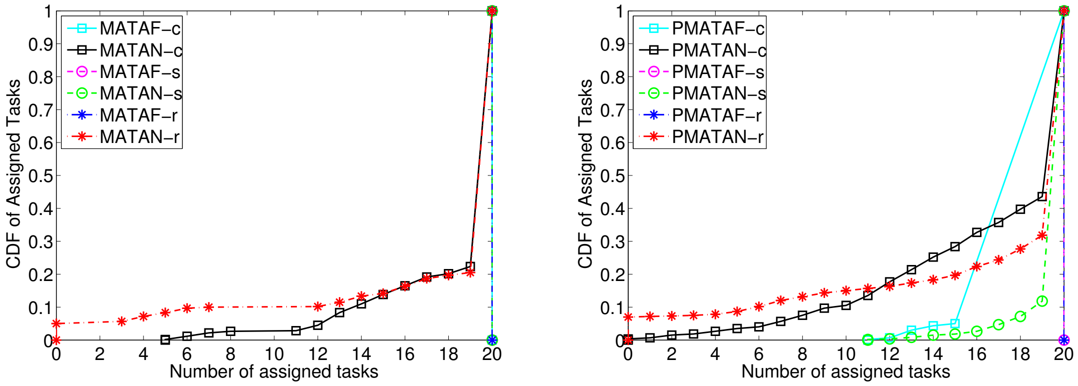
Figure 4.3: Assigned tasks by each scheme while varying the number of tasks

assigning, for all requesters selection types, more than 50% of the required tasks in 70% of the experiments for the NCSU trace and in 80% of the cases for the KAIST trace.

To this end, we compare the obtained results among different selection policies of requesters. We observe that for the two real traces, and the different variants and modes of assignment, the selected requesters *by-speed* are the ones with the most competitive results. For example, identified requesters using the *by-speed* policy assigned, by MATAF and P-MATAF, all sensing tasks for KAIST trace and more than 90% for NCSU. Moreover, this selection policy enhances the online assignment mode results by realizing the most important values of assigned tasks for all traces and schemes. Accordingly, we adopt for the rest of our evaluations the *by-speed* identified requesters and compare the efficiency of different algorithms based on this.

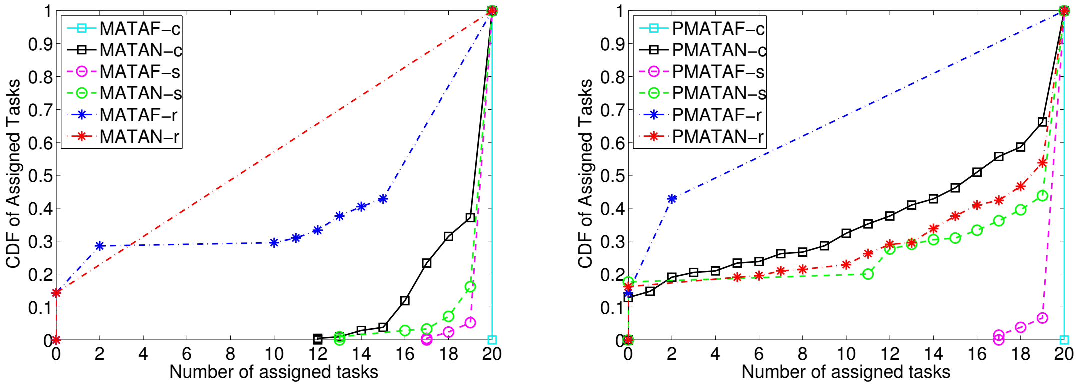
4.7.4.2 Average Achieved Makespan

In this part of the evaluation, we measure the achieved value of the average makespan, the objective to be minimized in this work, while varying first the number of tasks and



(a) MATA for KAIST

(b) P-MATA for KAIST



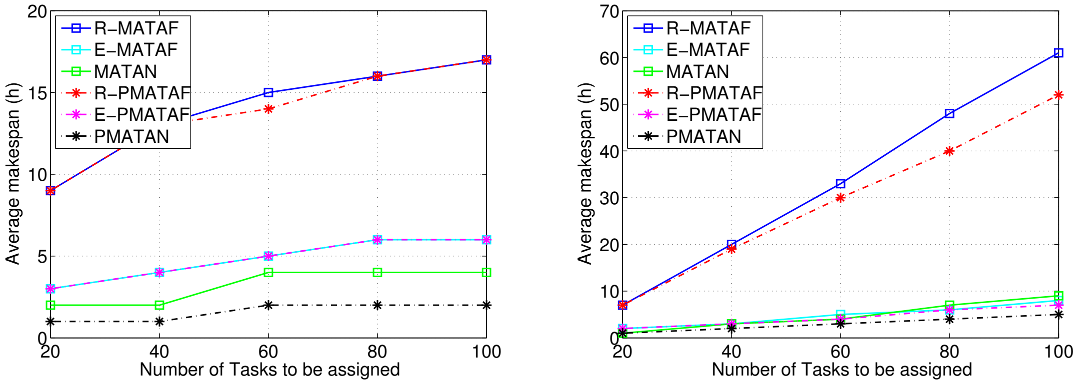
(c) MATA for NCSU

(d) P-MATA for NCSU

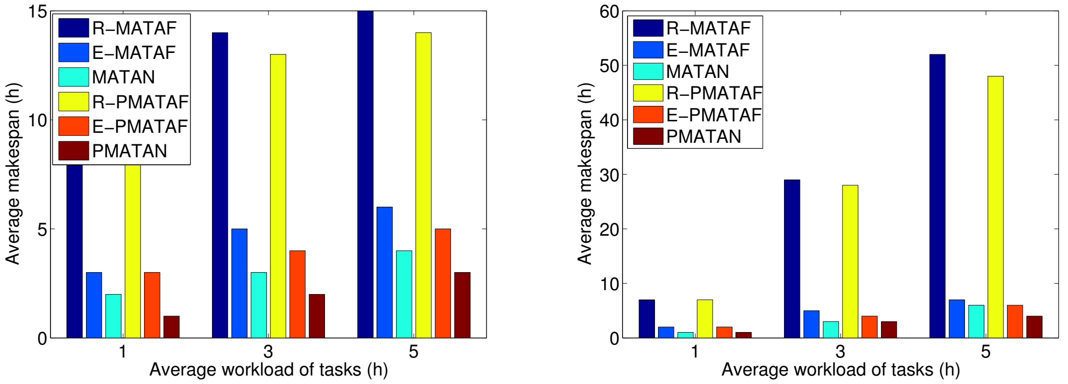
Figure 4.4: CDF of assigned tasks with $m = 20$ for each scheme and trace

second the average workload. However, it is worth mentioning that we computed the *Real* makespan for the offline modes. That is, we sum the workload of all tasks for each participant then recompute the makespan of each requester as detailed in Theorem 4.4 for the MATA variant and Theorem 4.6 for the P-MATA variant. This is because, in the offline mode, the requester assigns tasks based on his only Expected Sensing Time and without considering the fact that a participant may have received other assignments from other requesters. Results are shown in Figure 4.5 for both traces. Note that we denote By "R-x" the real makespan of a scheme and by "E-x" the estimated one.

Clearly, the average makespan of all algorithms increases as function of the number of tasks and the average workload as observed in Figures 4.5(a), 4.5(b), 4.5(c) and 4.5(d), respectively. We notice also that the variance between the estimated "E-x" and the real "R-x" makespan is relatively important. For both real traces, the estimated makespan of the two variants MATA and P-MATA are rather close to the online achieved average makespan. This was justified to be wrong by the high values described in the real makespan measures. Consequently, the online mode is proved to realize better performance by considering an updated Expected Sensing Time for all encountered participants. Furthermore, we observe that the online preference and mobility aware assignment scheme, P-MATAN, realizes the



(a) NCSU (b) KAIST



(c) NCSU (d) KAIST

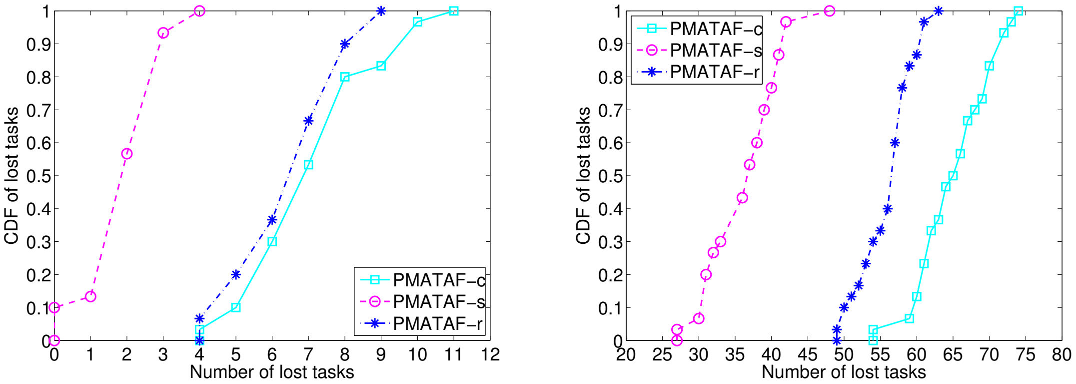
Figure 4.5: Average achieved makespan

best performance among all schemes by achieving the smallest values of average makespan. In other words, P-MATAN assigns tasks in a better way. This is due to the fact that in this variant mode, we account for an updated IST as well as a real time acceptance probability of each participant, which improves the re-assignment of rejected tasks in time. These results are shown to be valid for both cases, while varying the number of tasks and the average tasks workload.

4.7.4.3 The Number of Lost Tasks

In order to better investigate the results observed above, we study the distribution of the number of *lost* tasks. As we stated earlier, requesters assign tasks in P-MATAF respecting the generated Γ_{PF} strategy, however, without waiting for a response from the encountered participant which may result in a non perceived rejection and consequently some lost tasks.

In Figure 4.6, we plot the cumulative distribution function (cdf) of the number of these lost tasks for both real traces KAIST and NCSU and for different selection types of requesters. Note that since we have 7 requesters for the NCSU trace, we measure the



(a) cdf of lost tasks for NCSU (b) cdf of lost tasks for KAIST

Figure 4.6: CDF of lost tasks with $m = 20$ and $\tau = 1$h

number of lost tasks among the total $7 \times 20 = 140$ tasks to be assigned. Similarly, for the KAIST trace we designated 20 requesters, i.e., 400 tasks. For the former trace, the number of lost tasks does not exceed 10% while for the latter one goes up to 18%. Particularly, for both traces, we observe that the *by-speed* selected requesters assign better the held tasks by a percentage of only 2% of non perceived rejections for the NCSU trace and 11% for the KAIST trace. This conforms with our previous evaluation metrics. Besides, we emphasize the fact that P-MATAN assignment scheme is proved to perform better among all variants with no lost tasks. Indeed, requesters adopting this assignment strategy are able to detect real time rejections thereby hold back rejected tasks and assign them when encountering other participants.

4.8 Discussion

In this chapter, we seized the *hybrid* crowdsensing scenario, in which the basic idea consists of adding a distributed “support” phase to the centralized traditional one to overcome the various issues identified in the previous chapter. Motivated by such context, we focus on tasks assignment in a participatory sensing paradigm with the aim of minimizing the overall sensing and processing time of tasks.

First, we propose that the central platform identifies requesters, among registered users, to delegate sensing campaigns on the move. Therefore, we investigated in this chapter some of the possible requesters selection policies and chose to adopt the *by-speed* one as it was shown to identify the most efficient users. Moreover, we studied users arrival and preference models to estimate the inter-meeting time between each selected requester and the rest of participants. This is mainly to assign tasks to participants who potentially can reduce the average makespan of sensing. To this purpose, we advocated two variants of distributed assignment schemes; a mobility-aware variant and a joint preference and mobility-aware one.

For the first variant, we estimated the inter-encounter time of users based only on their mobility whereas the second variant introduces a recently discussed issue in crowdsensing, the ability to reject the proposed assignment. For both variants, we developed offline and online greedy-based algorithms. The former mode consists of generating at a glance the list of participants to whom to delegate tasks based on their expected sensing time. The second mode enhances a more dynamic interaction between users by inquiring encountered participants on their instant sensing time which accounts for current performed tasks and results in a more accurate assignment strategy.

We assessed the performance of all proposed distributed assignment schemes through real traces based simulations. Results show that the preference and mobility-aware variant outperforms the other variant, particularly, by its online mode P-MATAN which achieved the minimum average makespan with zero lost tasks. This is due to the principle followed by P-MATAN which examines an updated expected time of sensing for all participants. Indeed, P-MATAN generates the assignment strategy only after receiving the encountered user acceptance which enhances the detection of any rejection and avoids the problem of lost tasks.

To conclude, the proposed assignment schemes in this chapter achieve with competitive results and the P-MATAN mode is observed to be the more promising one. Nevertheless, the associated preference model is based on estimated historical records of users' acceptance probabilities. This can be perceived as a very simple model which should tackle the reasons behind a participant rejection. Moreover, respecting the P-MATA policy, we may encounter the issue of selfish participants who may repetitively reject their assignments given that their response is considered generated as a static Bernoulli variable. Therefore, we need to extend the preference model presented here to a more dynamic one depending on sensing attributes that may guide users choice in a crowdsensing process. These attributes can include, but are not limited to, participants energetic concerns and potential proposed rewards by requesters.

Chapter 5

Extended Preference-aware Task Assignment with Incentive Mechanisms

Contents

5.1 Introduction 69
5.2 Motivations and Context 70
5.3 Preference-aware and Incentivizing MCS 71
5.4 Problem Statement 73
5.5 Extended Preference-aware Task Assignment: P-MATA$^+$ . . . 76
5.6 Incentives-based Preference-aware Task Assignment: IP-MATA$^+$ 78
5.7 Performance Evaluation 82
5.8 Discussion 86

5.1 Introduction

In the previous chapter, we introduced a *hybrid* crowdsensing architecture. The principle consists of supporting the centralized management of participants with a distributed phase. More precisely, we are interested in minimizing the average time of processing sensing campaigns. Among the proposed assignment variants, the one based on users preferences, mobility-awareness and online, P-MATAN, has been proved to be the most promising.

In this chapter, we continue investigating the distributed assignment issue while targeting the minimization of the average makespan of all sensing tasks. However, we propose to pursue users preferences more vigorously. Hence, we introduce an extended Preference and Mobility-aware Task Assignment scheme that we name as P-MATA$^+$. This task allocation method considers an interactive preferences model where participants are capable to select their assignment based on different sensing attributes such as their energetic constraints, the task type or the associated workload to name a few. Such attributes are incorporated into a regression choice model [90,91] in order to design the dependency of users acceptance/rejection to these criteria. Furthermore, we seize the issue of incentivizing participants in MCS campaigns and introduce rewards as an attribute in users choice model. Accordingly, we present an Incentives-based Preferences and Mobility-aware Task Assignment method, IP-MATA$^+$, and we investigate the pricing policy in order to identify an efficient rewarding model where both requesters and participants realize a satisfying crowdsensing process.

The remainder of this chapter is organized as follows. We first highlight the different motivations behind our extended preference model without/with incentives consideration in Section 5.2. Section 5.3 enumerates the related work to the proposed assignment method in this chapter. Moreover, we present the problem illustrating this chapter aim in Section 5.4. Based on this model, we develop the two assignment variants P-MATA$^+$ and IP-MATA$^+$ in Sections 5.5 and 5.6, respectively. We evaluate the efficiency of the proposed allocation schemes on real-traces in Section 5.7 and we discuss our contributions assessment in Section 5.8.

5.2 Motivations and Context

To study the open issue of participants recruitment in mobile crowdsensing, we propose through this dissertation different assignment schemes. First, we targeted optimizing data quality (QoI) level. Along with this purpose, we developed fair and energy-aware task allocation schemes. Then, we studied sharing the workload among participants to minimize the processing time and we introduced users ability to accept/reject the proposed assignment. Nonetheless, the developed preferences model in the previous chapter computes a participant's acceptance/rejection probability towards an assignment based on historical records. On the contrary, this factor is not a static entity but depends on many sensing attributes. Therefore, we present in this chapter a choice model for participants while taking into account sensing tasks characteristics such as task type, workload and/or associated rewards if there is any.

As a matter of fact, we are assuming so far that participants are performing sensing tasks "voluntarily" as in the citizen sensing paradigm [3]. However, mobile crowdsensing, especially the participatory sensing paradigm, is undoubtedly consuming users devices physical resources such as energy (battery), computing and storage [1]. In addition, participants are dedicating their time and even human intelligence to contribute data. Therefore, they may be reluctant to participate to sensing campaigns unless for perceiving certain utility or being rewarded. Rewards are defined as "incentives" [46,47] and presented in different workers recruitment methods as well as in MCS to ensure participants commitment and realize an efficient sensing process. Hence, we incorporate rewards in user choice model and we study their impact. For the incentivizing mechanism, we assume that the MCS platform dedicates a budget to each requester for the distributed assignment and investigate two possible rewarding policies: task-priority based and data-quality based. This is to determine how to assign tasks while managing better the provided budget as described in details in Section 5.4.

In summary, we propose in this work to assign tasks when considering the evolution of participants preferences. We introduce two variants of the assignment scheme; without and with incentives. For both variants, we propose greedy-based offline and online algorithms as advocated in Sections 5.5 and 5.6. More specifically, our major contributions here are as follows:

1. We develop a new preferences model based on the regression logit model [91]. This elaborates the fact that users acceptance towards a certain assignment is mainly dependent on a certain utility function measured by different attributes.
2. We adopt this new preferences model to extend our previous assignment policy as the no-incentives based solution, P-MATA$^+$, where the choice model depends mainly on the current workload of users.
3. We advocate an incentives-based assignment variant, IP-MATA$^+$, while empowering requesters by a budget B to enhance participants' contributions.
4. We investigate two different incentive policies: task-priority based and data-quality based. The former accounts for tasks heterogeneity while the latter targets rewarding users function of the quality of their contributions.

5.3 Preference-aware and Incentivizing MCS

The second contribution of this dissertation is positioned among distributed assignment schemes in crowdsensing systems. Therefore, we recall the literature work 4.3 on distributed MCS illustrated in previous chapter. Besides, we opt for a preference-aware assignment without/with incentives of which we investigate the related work hereafter.

5.3.1 Preference-aware Crowdsensing

To the best of our knowledge, few works have considered the ability of registered participants in a crowdsensing system to reject their assignment. Literature work generally assumes that participants are willing to contribute data as long as the assignment method is energy-aware and/or offers rewarding services. However, as stated in the previous section, participants may be reluctant to perform tasks and need to reject the proposed assignment for various reasons. Hence, MCS systems must consider this ability and behave according its estimation. In this context, the Crowdlet [66] worker recruitment paradigm in Crowdsourcing systems has presented two proactive probabilities. The first is the probability of a worker acceptance towards the proposed task and depends only on the associated reward. The second considers the ability of a worker to fulfill this task according to his familiarity degree with its characteristics. The presented acceptance probability of this work is derived from research on pricing workers to finish tasks on time [90, 92]. Similarly, we suggest to extend our previous preference-aware scheme, P-MATA [18], to consider comparable choice model [90, 92]. Nevertheless, we do not consider only rewards but also other sensing tasks attributes such as the type and the workload. In order to propose appropriate rewards, we investigate incentivizing mechanisms developed for MCS systems in the following section.

5.3.2 Incentivizing Mechanisms in MCS

Several incentive mechanisms [65, 66, 84, 93–97] have been introduced for MCS. Zhang et al. [46] and Jaimes et al. [47] have surveyed these mechanisms and distinguished monetary and non-monetary incentives. The latter can be offered in form of services or entertaining games as detailed in our Literature review chapter. Yet, monetary rewards are the main intuitive incentives form [47] in participatory sensing and are designed usually using Game Theory or Auction-based models.

Yang et al. [98, 99] introduced a platform-centric incentive model where the reward is proportionally shared by participants in a Stackelberg game, and a user-centric incentive model where participants in the auction bid for tasks and get paid no lower than the submitted bids. Following this approach, a kind of auctions, named reverse auction, used in the negotiating phase between the MCS platform and participants is also widely developed in the literature [94, 97, 100]. Lee and Hoh [94] designed a Reverse Auction based Dynamic Price incentives mechanism with Virtual Participation Credit (RADP-VPC) that aims at minimizing the MCS platform cost. The idea of these works is to select among bidders the set who maximizes the social welfare. However, this may imply selecting participants who set low bids and who are usually those with low contributions quality which may result in non accurate data sensing process.

In contrast, Koutsopoulos [100] introduced a quality-aware incentive mechanism based on Vickery Clarck Groves reverse auction where the platform estimates the users participation level based on their posted costs then selects those who minimize the overall payment. Similarly, authors in [65, 95–97] presented quality-aware incentivizing mechanisms. Jin et al. [95, 96] introduced QoI as a metric into the design of reverse auction mechanism for MCS systems while considering also privacy-preserving mechanisms. Other works introduced rewards as function of data quality [65, 97]. That is, the platform publishes tasks and offers rewards to users based on their contributions quality.

The incentives mechanisms discussed above are usually implemented in the central MCS platform which requires a global knowledge of participants bids and potential data quality, which in most of the cases results in an important communication overhead. Thus, we propose in this work to offer incentives among participants in a distributed way as adopted by authors in [66, 84]. Notably, we introduce rewards as an attribute in a preference model which estimates participants' preferences towards a proposed assignment. A further detailed description of this model is illustrated in the next section.

5.4 Problem Statement

In the following, we give necessary preliminaries that describe the crowdsensing system we seize in this chapter. Accordingly, we formulate the problem and state our design objectives.

5.4.1 Preliminaries

5.4.1.1 System Overview

We recall the *hybrid* crowdsensing scenario illustrated in Figure 4.1 of Chapter 4 and we refer to the same notations to describe the system elements. That is, we consider the same scenario of a requester r_i carrying m sensing tasks to be assigned to encountered participants. However, the set of sensing tasks $S = \{s_1, s_2, \dots, s_m\}$ is not only heterogeneous in terms of average workload but also in terms of type. This new assumption can be designed by associating different tasks with different weights. Let $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ be the set of possible weights depending on tasks type, i.e., sensing application. These tasks are assigned respecting users mobility and preferences as detailed in the following.

5.4.1.2 Users Probabilistic Arrival

Given the ability of rejecting their assignment, requesters need to estimate the mean time to meet participants who may accept to perform the proposed tasks. This is computed by the probabilistic mean time detailed in previous chapter by Lemma 4.2. Yet, for this chapter, we propose to investigate two variants of preferences model: without and with incentives. Correspondingly, we distinguish two type of potential met participants as follows.

Definition 5.1. A "volunteer-positive" participant is a potential encountered user who is estimated to accept the proposed assignment with no perceived rewards.

Definition 5.2. A "positive" participant is a potential encountered user who requires rewards for performing sensing campaigns.

The aforementioned participants are capable of selecting among their assignment the tasks they like to perform. They can reject also tasks if they estimate their current workload exceeds their devices processing capacities. Respectively, we detail hereafter the new proposed user preferences model.

5.4.1.3 Discrete Choice Model

Different from the previous chapter, we do not assume knowing users acceptance probabilities p_a from their historical records. Thus, we present a Discrete Choice Model to characterize on which basis participants can select tasks among the proposed assignment. This model has been proposed first by Faridani et al. [90] to estimate workers selection of different tasks in economic marketplace according to a perceived utility. Naturally, each user targets to maximize his utility which may differ among others based on their perceptions to the different attributes of tasks such as the hourly load, reward or type. Under this model, we design the user acceptance probability p_a as the probability that the utility of the current task s_j exceeds all other assigned tasks utilities.

$$p_a(s_j) = Pr(U_j > \max_{j \neq i} U_i). \tag{5.1}$$

The utility, U_i, perceived when performing a sensing task can be expressed under the *Conditional Logit Model* [90, 91] as follows:

$$U_j = \beta_j z_j + \epsilon_j. \tag{5.2}$$

It is worth noting that z_j designate all observable attributes of the task s_j and ϵ_j accounts for non observable ones. In this model [90], a task utility is assumed to be linearly correlated with all observed attributes by a shared coefficient vector β, while parameters ϵ_j are assumed to be independent from each other and follow the Gumbel distribution [101]. Based on such assumptions, we derive the expression of a task s_j acceptance probability as a *Multinomial Logit Distribution* [91]:

$$p_a(s_j) = \frac{\exp(\beta z_j)}{\sum_{i=1}^{m} \exp(\beta z_i)}. \tag{5.3}$$

In this work, we set positive coefficients β_j for desirable task attributes such as reward or type and negative ones for undesirable attributes like task workload and the number of already accepted tasks. This is to highlight the *attractiveness* of the first two attributes since they maximize the utility of a task. On the contrary, the last attributes are associated with negative coefficients to model the burden of carrying heavy workload for a participant. The new predefined probability p_a of Equation (5.3) is necessary to compute the inter-meeting time between any requester r_i and a participant p_j in a compound k. Recall that we utilize this estimated time to select among users those who minimize the average time of processing. A more detailed study of our design objectives is presented below.

5.4.2 Problem Formulation

Let us now focus on the participatory sensing process. More precisely, we target to minimize the total necessary time to perform submitted tasks described as the average makespan in Definition 4.3. This aims to improve devices energy consumption of all participants and lowers the burden on the ones with a high number of tasks. We do so by advocating two variants for the assignment process: without and with incentives.

5.4.2.1 No-incentives-based Assignment

First, we suppose that all encountered participants in a certain compound are willing to perform sensing campaigns with no perceived rewards as presented by Definition 5.1. However, participants can accept or reject their assignment. This depends on their estimation of a task utility in terms of the number of already accepted tasks, n_{acc}, and their associated workloads τ. Accordingly, we derive a "volunteer-positive" participant acceptance probability from Equation (5.3) as follows:

$$p_a(s_j) = \frac{\exp(\beta_1 \tau_j + \beta_2 n_{acc})}{\sum_{i=1}^{m} \exp(\beta_1 \tau_i + \beta_2 n_{acc})} \tag{5.4}$$

where β_1 and β_2 are negative coefficients associated with no desirable attributes.

We exploit this new preferences model to extend our task assignment variant, P-MATA, as the no-incentives-based variant, named as P-MATA$^+$, and we investigate it in two different modes: offline and online assignment strategies as described later in this chapter, i.e., in Section5.5.

5.4.2.2 Incentives-based Assignment

As a second variant of assignment schemes, we introduce incentivizing rewards in order to study their impact on users commitment to participatory sensing campaigns. Therefore, we incorporate rewards as a third attribute in the formulation of a task acceptance probability. The task reward, denoted as R_j, is a desirable attribute thereby associated with a positive coefficient, $\beta_3 > 0$. The corresponding acceptance probability is then expressed as follows:

$$p_a(s_j) = \frac{\exp(\beta_1 \tau_j + \beta_2 n_{acc} + \beta_3 R_j)}{\sum_{i=1}^{m} \exp(\beta_1 \tau_i + \beta_2 n_{acc} + \beta_3 R_i)} \tag{5.5}$$

Algorithm 5 P-MATAF$^+$ Assignment Algorithm

```
Require: Set of sensing tasks $S = \{s_1, s_2, \dots, s_m : \tau_1 \leq \tau_2 \leq \dots \leq \tau_m\}$, Participants
    $P = \{p_1, p_2, \dots, p_{n_p}\}$, Matrix of Expected Sensing Time $EST$.
Ensure: Assignment strategy $\Gamma_{PF^+} = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$
for $s_l \in S$ do
    for $p_j \in P$ do
        $p_a(s_l) = \frac{exp(\beta_1 \tau_{s_l} + \beta_2 n_{acc})}{\sum_{t=1}^{m} exp(\beta_1 \tau_{s_t} + \beta_2 n_{acc})}$.
        $\Pi_{i,j,k} = \left( \sum_{x=1}^{n} (1 - p_a(s_l))^x p_a(s_l) \frac{1}{q_k(r_i) q_k(p_j) \lambda_{p_j}} \right)'$
        $EST_{k,j} = \Pi_{i,j,k} + \tau_{s_l}$
    end for
    Proceed as in Algorithm 3
end for
Return $\Gamma_{PF^+}$
```

We define this variant as the Incentives-based Preference and Mobility-aware Task Assignment, IP-MATA$^+$. As for the first variant, we develop here also online and offline assignment strategies that we denote by Γ_{IPN^+} and Γ_{IPF^+}, respectively. It is worth noting that we propose two different incentivizing policies throughout this work, task priority-based and data quality-based. A more detailed study of these two policies is illustrated in Section 5.6 in order to highlight the most efficient one.

5.5 Extended Preference-aware Task Assignment: P-MATA$^+$

In this section, we present the no-incentives based assignment scheme, P-MATA$^+$. This scheme refers mainly to the introduced discrete choice model expressed by Equation (5.3) to seize users preferences in terms of assigned tasks workloads and number. Accordingly, P-MATA$^+$ foresees "volunteer-positive" participants encountering mean time as detailed in Lemma (4.2) and decides to whom to delegate tasks. Similar to our preceding work, we investigate this assignment variant considering two operation modes: offline and online.

5.5.1 Offline Mode: P-MATAF$^+$

The first mode, named as P-MATAF$^+$, suggests that each requester needs to run the assignment phase only in the beginning of a sensing period. The resulted assignment, denoted as Γ_{PF^+}, is to be delegated to the selected participants when encountering them. Each requester estimates the next time slot in which he will meet a "volunteer-positive" participant based on the computed acceptance probabilities p_a derived from the no-incentives preferences model of Equation (5.4). Accordingly, the average makespan is determined as the sum of workloads to each participant plus the elapsed time before the first acceptance as in Theorem (4.6). This offline no-incentives based solution is depicted by Algorithm 5 which reuses common steps from Algorithm 3.

5.5.2 Online Mode: P-MATAN$^+$

The second mode, P-MATAN$^+$, is a more interactive assignment strategy that can be divided into three main phases.

Pre-selection: This step is conducted by each requester r_i encountering a participant p_j in a certain compound k. A comparable algorithm to Algorithm 4 is run to determine the list of users who potentially minimize the average makespan of the online strategy, $AM(\Gamma_{PN^+})$. If the encountered participant p_j is among this list, he receives his assignment γ_j.

Participant Choice: Based on the proposed tasks γ_j and the previously assigned tasks, a participant p_j computes his acceptance probability p_a function of two main attributes: the current proposed workload τ_l and the number of already accepted tasks n_{acc} as expressed in Equation (5.4). This is done for every task separately, then answers are generated as a Bernoulli process $B(n, p_a)$. For each task $s_l \in \gamma_j$, if the selected variable $X = 1$, then the participant will accept to preform the corresponding task. Otherwise, the response is a rejection. The participant choice is modeled then as a boolean vector which contains answers to all proposed tasks; $V_{ans} = [\, X_1 \; X_2 \; ... \; X_m \,]$ and sent to the requester.

Final Selection: The requester receives the participant vector of answers and starts a process of verification of tasks confirmation. If the answer (element of vector V_{ans}) $X = 1$, the requester removes the task with the corresponding index from his list of tasks S and consider it as assigned given he receives the participant's confirmation. If $X = 0$, the requester holds back the corresponding task, reassigns it in the next meeting to other participants and updates the assignment strategy Γ_{PN^+}. This is performed with no additional exchange with the encountered participant to avoid communication overhead.

5.5.3 After Thoughts

In the extended Preference and Mobility-aware Task Assignment variant, P-MATA$^+$, we restrict tasks attributes to two undesirable ones. As a result, the discrete choice model is an exponential function of two negative associated coefficients which may yield to a *sharp* regression of users preferences. That means, a participant acceptance probability is rapidly decreasing especially when receiving many overlapping assignments from encountered requesters. In order to avoid resulting repetitive rejections, we propose to enhance participants commitment by introducing in a second variant of the assignment scheme desirable attributes such as rewards.

5.6 Incentives-based Preference-aware Task Assignment: IP-MATA$^+$

The prior presented no-incentives based assignment scheme delegates tasks to participants with no reward in return. Though, participants may be unwilling to perform participatory sensing campaigns, especially if they receive an important workload. Thus, we introduce, in this section, incentives such as monetary rewards in the aim of encouraging users to contribute their data. We present first the different incentive policies and we detail further the offline and online modes of IP-MATA$^+$.

5.6.1 Incentive Policies

We suppose that the sensing platform empowers each selected requester with a certain budget B to encourage encountered users to perform proposed sensing tasks. Nevertheless, there are several policies based on which a requester can manage the available budget and proposes accordingly rewards. We identify, in the following, two different incentivizing policies: task-priority based and data quality-based.

5.6.1.1 Priority-based Incentives

Recall that we consider in this part of work a set of heterogeneous sensing tasks S in terms of type or involved sensors. As a consequence, certain type of tasks can be perceived as more important or primary to be performed. Hence, depending on the type of a task s_l, a requester may set different rewards. For instance, a requester may prioritize video streaming tasks rather than localization one. To describe this prioritization, we associate each task with a weight α_l. The higher is the value of α_l, the more prior the task is. In this context, the pay-off (reward) offered can be proportional to the task weight compared to other proposed tasks. We introduce then an incentivizing policy that defines a reward/incentive as follows:

$$I_p(s_l) = \frac{\alpha_l}{\sum_{k \in \gamma_j} \alpha_k} B(t) \tag{5.6}$$

where $B(t)$ is the current residual budget initialized at B.

This incentivizing policy may enhance participants to perform harder sensing tasks, particularly by setting their associated weights to important values.

5.6.1.2 Quality-based Incentives

The second incentivizing policy is a data-quality based since we believe that prioritizing tasks may not be the only criteria to optimize the task assignment in MCS systems. Indeed, for the same type of tasks, participants may have different quality of data samples depending on their sensor characteristics, location accuracy and/or sensing time-efficiency. Thus, we propose to account for the estimated quality of collected data. In this contribution, we define as quality criteria the "timeliness" of the measurement since we aim to minimize the average makespan. Therefore, we recall the utility function of [71] to evaluate the time-quality of the contributed data, by normalizing the Expected Sensing Time (EST) of the current participant p_j. The expression of the data-quality based reward is as follows:

$$I_q(s_l) = U(EST_j)\frac{\alpha_l}{\sum_{k\in\gamma_j}\alpha_k}B(t) \tag{5.7}$$

where $U(EST_j)$ is the utility of the current Expected Sensing Time compared to the minimum and the maximum EST estimated for the rest of participants.

This incentivizing policy jointly takes into account the prioritization and the data-quality of sensing tasks, which may positively impact the overall average makespan. Particularly, the reward allocated to each task is proportional to the execution quality of the participant which aims at attracting good quality data contributes, i.e., lower EST, by offering to them higher amount of incentives.

5.6.2 Offline Mode: IP-MATAF$^+$

Regardless of whether we incentivize participants on a task-priority basis or a data quality one, we develop hereafter our assignment modes for both rewarding policies. First, we adopt the same policy as in P-MATAF$^+$ while incorporating the reward as a third attribute in the discrete choice model of Equation (5.5) to compute each participant acceptance probability. Accordingly, we derive the mean arrival time of all "positive" participants and the corresponding average makespan $AM(\Gamma_{IPF^+})$ from Theorem 4.6, with Γ_{IPF^+} is the incentives-based assignment strategy while adopting the offline mode. We proceed to look for users who minimize the average makespan $AM(\Gamma_{IPF^+})$ of all tasks in a crowdsensing support phase as described by the steps of Algorithm 6.

We start by estimating the current task reward R_l based on the selected incentive policy then we incorporate it in the acceptance probability defined by Equation (5.5). This to compute the inter-encounter time $\Pi_{i,j,k}$, as in Lemma (4.2), and generate the list of possible "positive" participants. Further, we determine the EST of all participants and look for the smallest EST_j as introduced in P-MATAF. Finally, we update the selected participant assignment γ_j, his EST_j, the set of assigned tasks S as well as the residual budget $B(t)$. We continue the selection until we assign all tasks carried by a requester r_i.

It is worth noting that the stop condition of Algorithm 6 is only the assignment of all tasks. This is due to the fact that our incenitvizing policies do not exhaust the available budget B. In fact, the budget B can be initialized to a multiple of a unit budget b depending

Algorithm 6 IP-MATAF$^+$ Assignment Algorithm

```
Require: Set of sensing tasks $S = \{s_1, s_2, \dots, s_m : \tau_1 \leq \tau_2 \leq \dots \leq \tau_m; \alpha_1, \alpha_2 \dots \alpha_m\}$
    Participants $P = \{p_1, p_2, \dots, p_{n_p}\}$, Budget $B$, Matrix of Expected Sensing Time $EST$
Ensure: Assignment strategy $\Gamma_{IPF^+} = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$
$B(t) \leftarrow B$
for $s_l \in S$ do
    for $p_j \in P$ do
        if $I = I_p$ then
            $R_l = \frac{\alpha_l}{\sum_{k \in \gamma_j} \alpha_k} B(t)$
        else if $I = I_q$ then
            $R_l = U(EST_j) \frac{\alpha_l}{\sum_{k \in \gamma_j} \alpha_k} B(t)$
        end if
        $p_a(s_l) = \frac{\exp(\beta_1 \tau_l + \beta_2 n_{acc} + \beta_3 R_l)}{\sum_{i=1}^{m} \exp(\beta_1 \tau_i + \beta_2 n_{acc} + \beta_3 R_i)}$
        $\Pi_{i,j,k} = \left( \sum_{x=1}^{n} (1 - p_a(s_l))^x p_a(s_l) \frac{1}{q_k(r_i) q_k(p_j) \lambda_{p_j}} \right)'$
        $EST_{k,j} = \Pi_{i,j,k} + \tau_{s_l}$
    end for
    Proceed as in Algorithm 3
    $B(t) \leftarrow B(t) - R_l$
end for
Return $\Gamma_{IPF^+}$
```

on the number of tasks; $B = m \times b$. Additionally, we share B respecting a proportional method. Thus, the current budget is always lower or equal to the necessary budget for the rest of the tasks, i.e., $B(t) \leq m(t) \times b$ with $m(t)$ is the number of non-assigned tasks. Consequently, the maximum dedicated budget to an encountered participant is equal to the unit budget b. This is reached for example in the case the participant affords the best expected sensing time among all potential candidates.

5.6.3 Online Mode: IP-MATAN$^+$

IP-MATAN$^+$ is the online mode for the Incentives-based task assignment presented in this chapter. We respect the previous detailed assignment phases of Section 5.5. That means, requesters inquire encountered participants to update their Instant Sensing Time (IST) and generate accordingly the list of selected users. If the current participant is selected, he receives his assignment. However, different from P-MATA$^+$, the requester needs to send both tasks workloads and weights. As a consequence, a participant acceptance probability depends mainly on the perceived reward that is computed from the proposed task type (weight) and/or the participant IST. For instance, if the incentives are priority-based and the proposed tasks are with high weights values, α_l, it is rather more probable that the assignment will be accepted. Especially, if the current participant has few completed tasks in the past. On the other hand, if incentives are quality-based and the participant is currently performing "heavy" processing load, the estimated "timeliness", i.e., $U(IST_j)$,

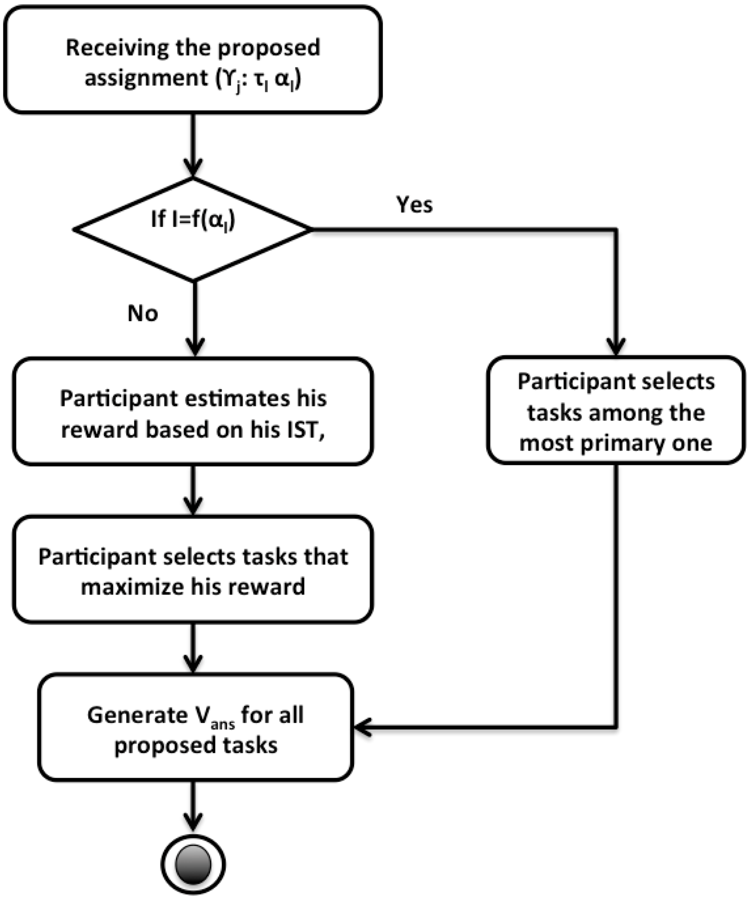


Figure 5.1: Participant choice model

can be low. Hence, the corresponding offered pay-off is low and the participant may reject any proposed assignment even if it is associated with high priority, α_l. The aforementioned cases are summarized in the activity diagram of Figure 5.1. After receiving the participant choice, the requester proceeds to the last phase of final selection. That is, for each task s_l the requester verifies if the corresponding answer $X = 1$ or it is a rejection. Accordingly, the set of tasks S is updated and so the available budget $B(t)$. The above steps are repeatedly executed whenever encountering a participant until assigning all tasks.

5.6.4 After Thoughts

In this section, we suggest *paying* participants for dedicating their devices resources during sensing campaigns. We also study varying the method of estimating such pay-off. Nevertheless, the reward estimation does not only depend on the incentivizing policy as detailed before but also on the assignment mode.

- IP-MATAF^{+}: In this mode, it is up to the requester to compute a possible reward to each participant based on the estimated EST or the current task weight α_l.

- IP-MATAN$^+$: In the online mode, the requester only sends tasks weights along with their workloads. Each participant computes based on the announced incentives policy, the correct reward and decides, after computing his acceptance probability p_a, to accept or reject the proposed assignment.

In the following section, we conduct simulations to determine which assignment mode is the most accurate. *Did incentives encourage participants to process tasks in time-efficient method?* and *Which incentives policy is the more adept to utilize the available budget while achieving the objective of minimizing the average makespan?*

5.7 Performance Evaluation

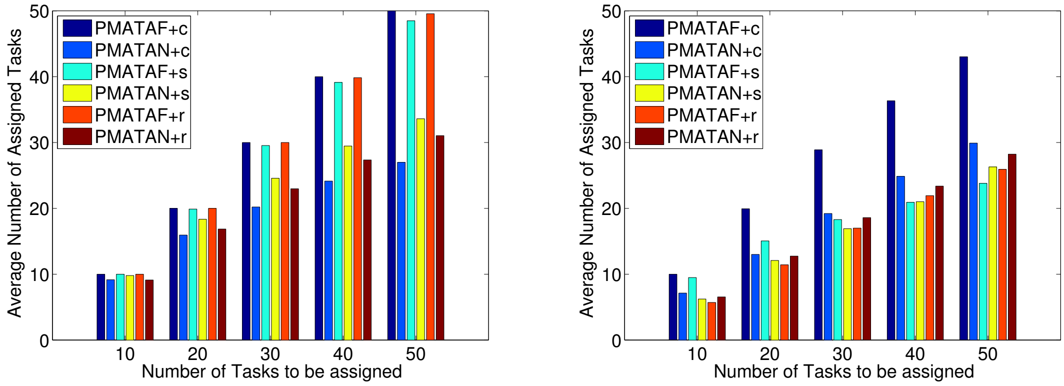
In order to evaluate the performance of the proposed schemes, P-MATA$^+$ and IP-MATA$^+$, and their different modes, offline and online, we run extensive simulations. Hereafter, we detail the simulation settings, the evaluation metrics and the obtained results.

5.7.1 Simulation Settings

To simulate the proposed distributed crowdsensing schemes, we utilize the real-mobility traces of NCSU and KAIST campuses [89] described in the previous chapter. This is to compare the different schemes on a same basis. Similarly, we dedicate a first phase to check if the prior identified requesters selection method is valid. Therefore, we vary in the beginning the requesters selection *randomly*, *by-contact* and *by-speed*, then decides which method to adopt for the rest of the evaluation. Furthermore, we consider a set of sensing tasks S to be assigned by each requester to encountered participants of which we vary the number and associated workloads to observe their impact on our proposed assignment schemes. The number of tasks is selected from $\{10, 20, 30, 40, 50\}$ and the average workload of all tasks τ is selected from $\{1, 3, 5\}$ (hours). Moreover, we consider that sensing tasks are heterogeneous and assign to them different weights, $\alpha_l \in [1, 10]$, and associate their attributes z_l in $\{load, number, reward\}$ with random coefficient β_l. Simulations are conducted also under the network simulator ns-3.24 while varying first the number of tasks m and second the average workload of tasks τ. Within each group, 30 runs are performed for each trace (NCSU/ KAIST). Results obtained for all assignment schemes, P-MATAF$^+$, P-MATAN$^+$, IP-MATAF$^+$ and IP-MATAN$^+$, are illustrated in Figures 5.2 to 5.6.

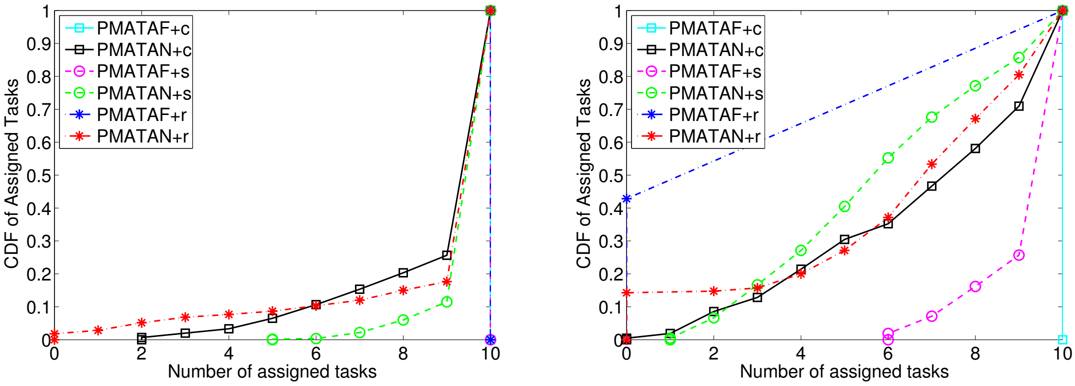
5.7.2 Performance Analysis

In the following, we first compare the number of assigned tasks achieved by different requesters selection methods. Then, we evaluate the realized makespan by each assignment scheme. Finally, we investigate the effectiveness of our incentive-based policies.



(a) P-MATA$^+$ for KAIST (b) P-MATA$^+$ for NCSU

Figure 5.2: The average number of assigned tasks by different requesters



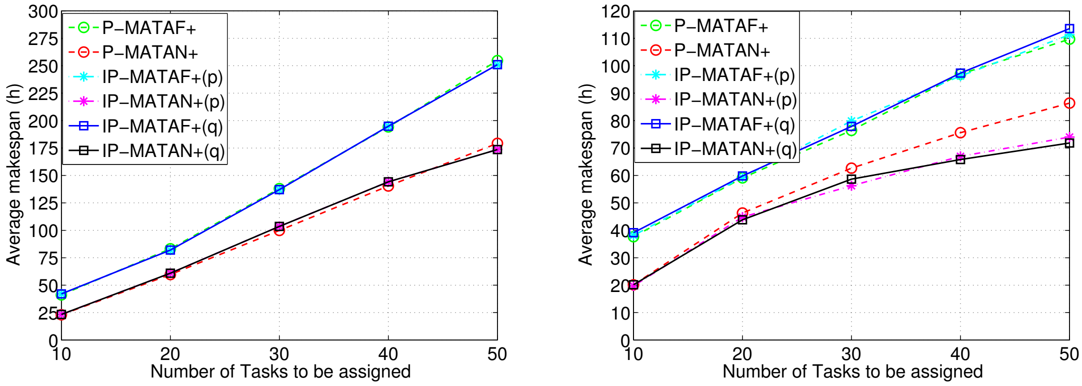
(a) P-MATA$^+$ for KAIST (b) P-MATA$^+$ for NCSU

Figure 5.3: The cdf of assigned tasks for each requesters selection policy

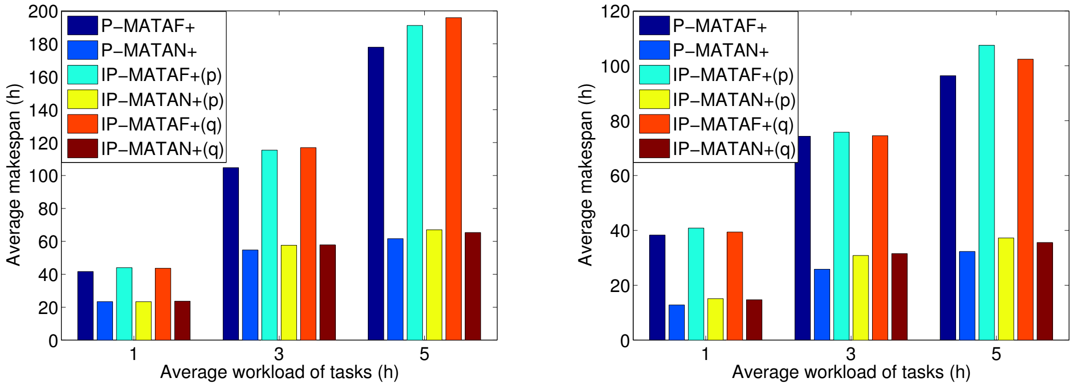
5.7.2.1 Average Number of Assigned Tasks

We run simulations with different number of tasks and an average workload $\tau = 1$h on the two considered traces, KAIST and NCSU [89]. As a first step, we plot only the no-incentives-based scheme, P-MATA$^+$, results. Nevertheless, we test it for different requesters selection strategies. Results illustrated in Figure 5.2 conform with the prior evaluated preference and mobility-aware task assignment scheme. That means, the offline mode P-MATAF$^+$ achieves the highest values of assigned number of tasks for both traces. Furthermore, we investigate the distribution of this number for different selection types while setting the number of tasks $m = 10$. Similarly, the results in Figure 5.3 confirm our observations. The offline modes for both traces outperform the online ones by assigning all and more than 65% of sensing tasks for KAIST trace and NCSU trace, respectively.

Though, due to the adopted Discrete Choice Model of Equation (5.3) the efficiency of requesters selection method is slightly different from prior results, i.e., those obtained in Chapter 4. For instance, we remark that for NCSU, selected requesters *by-contact* are



(a) KAIST (b) NCSU



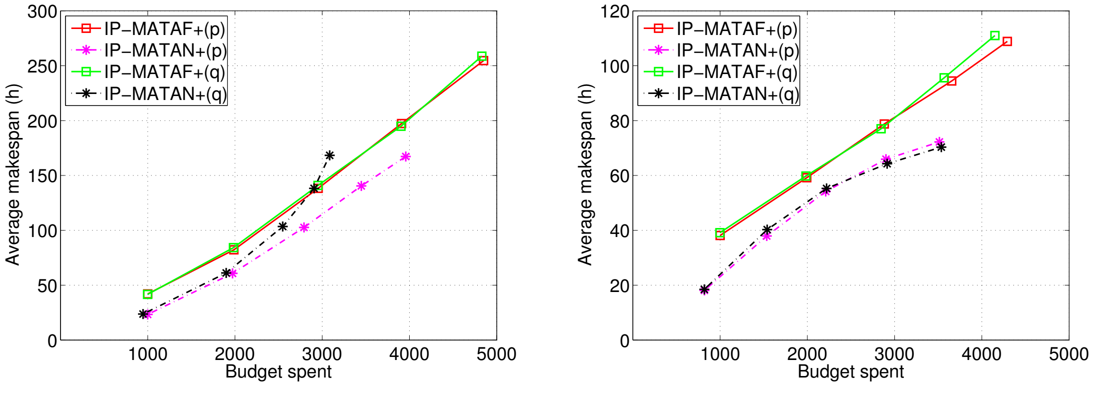
(c) KAIST (d) NCSU

Figure 5.4: Average achieved makespan

the ones with the most competitive results, especially in the online mode. Requesters *by-contact* have successfully assigned more than 80% of tasks by P-MATAF$^+$ and around 60% by P-MATAN$^+$. This difference is due to the evolving p_a function of the accepted tasks, loads and rewards. For the KAIST trace, the selected requesters *by-speed* perform better as observed in the preceding chapter. Indeed, requesters identified by this selection strategy assign at least 98% of all tasks by P-MATAF$^+$ and more than 60% by the online mode, P-MATAN$^+$. These observations are confirmed by the cumulative distributed function (cdf) plot of Figure 5.3 where both the offline and the online modes perform better when associated to the *by-speed* requesters selection strategy for KAIST trace and *by-contact* requesters for NCSU. Motivated by these results, we adopt for the rest of evaluations the *by-contact* identified requesters for NCSU trace and the *by-speed* ones for the KAIST trace.

5.7.2.2 Average Makespan

In this part of the evaluation, we plot the average makespan values realized by both no-incentives and incentives-based assignment schemes, P-MATA$^+$ and IP-MATA$^+$, respectively. In this aim, we vary first the number of tasks and second the average workload. The



(a) KAIST (b) NCSU

Figure 5.5: Expenditure efficiency of makespan over spent budget for both traces

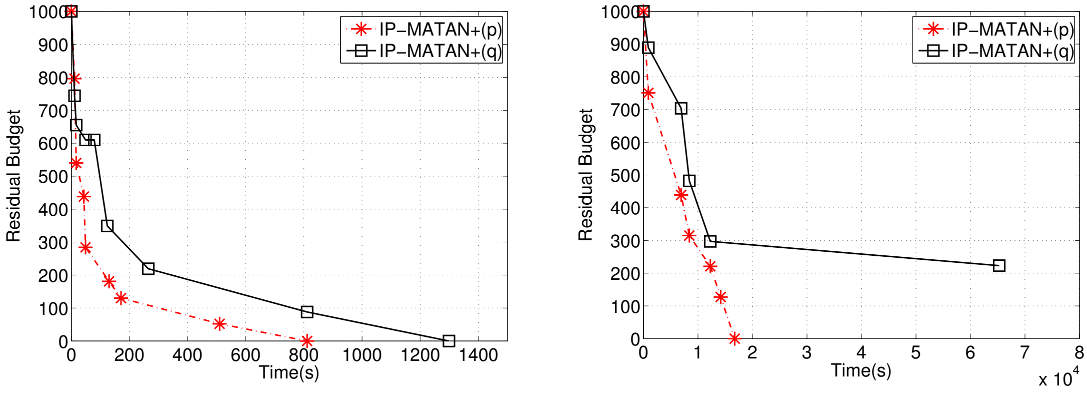
results are illustrated in Figure 5.4 for both traces.

As stated in the preceding chapter, the average makespan is an increasing function of the number of tasks or the average workload. The online modes achieved better results since they consider the instant response of the encountered participants and try accordingly to assign the maximum of tasks, especially in case of rejection. Moreover, we observe that for both real traces the realized makespan values of the two incentivizing policies of IP-MATA$^+$, i.e., the priority-based (p) and the quality-based (q) are the lowest. Particularly, for the NCSU trace, the incentives-based scheme IP-MATA$^+$ has realized lower values of makespan by enhancing the limited number of available participants ($n_p = 27$) even with high workload ($m = 50$). However, for the KAIST trace, all schemes perform similarly since there are more available users ($n_p = 72$), and even with no-incentives all tasks are assigned and all schemes achieve good makespan values.

5.7.2.3 Incentives Policies Performance

The performance of the different incentives policies, the priority-based reward (p) and the quality-based reward (q), is depicted in Figures 5.5 and 5.6.

The expenditure efficiency of the achieved makespan over the budget spent is shown in Figure 5.5. We observe that the offline mode, IP-MATAF$^+$, realizes comparable results for KAIST and NCSU traces. Indeed, in this mode, the reward is based on estimated acceptance probability and not the current updated one, which may not be accurate enough. However, for the online modes, the incentives policies persuade participants to gradually perform tasks by adapting the offered reward which minimizes the overall makespan and exploits the residual budget in a better way. This can be clearly observed in Figure 5.5.a, as the behavior of IP-MATAN$^+$ under the quality-based reward policy and KAIST trace. In fact, both online incentives policies reach comparable makespan values. Yet, the quality-based one achieves the same value with lower budget spent. For instance, IP-MATAN$^+$(p)



(a) KAIST

(b) NCSU

Figure 5.6: Budget evolution function of time for both traces

realizes around 150h of makespan by spending 4000 units while IP-MATAN$^+$(q) achieves the same makespan when spending only 3000 units. This difference is slightly remarked with NCSU trace due to the limited number of participants which obliges IP-MATAN$^+$(q) to offer rewards to modest-quality contributors in order to assign all tasks.

Moreover, we plot the evolution of the residual budget over the time under the online incentives scheme, IP-MATAN$^+$, in Figure 5.6 for both traces. Thus, we select one requester and compare the budget exploitation under the two different incentive policies. It is shown that the curve is more sharp for the priority-based incentives policy, IP-MATAN$^+$(p), than the quality-based, IP-MATAN$^+$(q). The budget is expended in a faster rate and is used up after a small period of time in contrast to IP-MATAN$^+$(q). This is explained by the fact that quality-proportional rewards are rather lower especially for participants offering high Expected Sensing Time values.

To conclude, the incentives-based scheme outperforms the no-incentives one by encouraging more participants to accept extra sensing tasks, especially in online modes. More precisely, IP-MATA$^+$ realizes comparable makespan values with its two incentivizing policies. Nevertheless, the quality-based one reaches such values while reducing the spent budget and improving the achieved quality.

5.8 Discussion

Hybrid assignment in crowdsensing systems needs to recruit participants by the central platform and the “supportive” requesters. The latter phase has been developed in this chapter while recalling common scenarios from the preceding one. Nevertheless, we investigated here user preferences as a proactive regression logit model. That is, we suggested to model the dependence between participants acceptance probabilities and their devices current workload besides proposed task type or reward. This is to emphasize the fact that participants usually reject heavy sensing campaigns that exhaust their devices batteries. Also, they may be attracted to certain type of sensing tasks, which leads to rejecting all others. Finally, the most common attribute that may affect users commitment is whether they are receiving incentives for their contributions or not.

In the aim of studying these attributes impact, we developed in this chapter two different assignment schemes, one without and the other with incentives. Both variants target the minimization of the average makespan. Moreover, we introduced offline and online solutions for each variant. We elaborated the different estimations of users' acceptance probabilities and consequently their expected time of sensing. In other terms, we let in the first mode each requester assign tasks according to his perceived EST. While in the online mode, the encountered participant is responsible of providing a more accurate result. Furthermore, we advocate two rewarding policies for the incentives-based task assignment, IP-MATA$^+$. The first is developed on the basis of task priority described by its associated weight. The second is rather depending on the potential contribution quality in terms of “timeliness”. This is to highlight the difference of user preferences on one hand and to investigate the efficiency of both methods in terms of budget allocation on the other hand. All proposed solutions were evaluated using real mobility traces simulations. As a consequence, the incentives-based preference and mobility-aware variant, IP-MATA$^+$, has been observed to outperform the other variants. Specifically, the online mode has achieved the minimum average makespan. In addition, when comparing the two incentives policies, the quality-based one has been proved to be more efficient in terms of budget expenditure.

This work has shown the effectiveness of proposing rewards to enhance participants commitment in participatory sensing campaigns. Particularly, in low density areas as in the case of the NCSU trace, incentives have reduced the number of rejections observed within the no-incentives assignment, P-MATA$^+$. P-MATA$^+$ and IP-MATA$^+$ assignment strategies can be generalized to the centralized phase as well but with a communication cost among the different entities. Similarly, the quality-based rewarding policy can be extended to other data quality criteria in order to attract participants with good contributions. To this extent, we aimed to satisfy both requesters and participants by introducing prior-measurement during the task assignment and processing phases. Nonetheless, participants may be anxious during the uploading phase as well. This is due to the fact that collected measurements may contain or be coupled with sensitive information such as location, user ID or context of collection. In the next chapter, we propose to investigate this issue in order to protect users privacy in crowdsensing systems.

Chapter 6

Privacy Preserving Utility-aware Mechanism in Data Uploading

Contents

6.1 Introduction 89
6.2 Motivations and Context 90
6.3 Research Work on Privacy-preserving Mechanisms 92
6.4 System Model 94
6.5 Problem Formulation 96
6.6 Proposed Solution: PRUM 102
6.7 Performance Evaluation 104
6.8 Discussion 112

6.1 Introduction

Previously, we focused on the *Task Assignment* and the *Sensing and Processing* phases of crowdsensing systems. In such a context, we presented three main contributions of this dissertation. First, we targeted to seize the energy cost along with data-quality issues. Then, we investigated the minimization of the processing time of sensing campaigns by estimating users behavior in a distributed fashion. And finally, we introduced necessary rewards to encourage participants contributions. These schemes handle major participants and requesters concerns yet not the privacy related one. Indeed, users privacy concerns are claimed to be the most significant barriers to their participation in sensing campaigns [54, 102]. Usually, participants upload data publicly in participatory tasks which accentuates the risk of inferring their sensitive information from reported measurements.

In this chapter, we tackle the privacy issue in the *Data Uploading* phase of a MCS process and mainly while using the *participatory* paradigm. We develop a PRivacy-preserving

Utility-aware Mechanism, PRUM [19], which aims at protecting participants private information while accounting for the potential degradation of data quality when applying privacy preserving techniques. Therefore, we first study two different uploading scenarios of collected data. Furthermore, we examine the potential adversary elements and models to better design our protection mechanism. Hence, an adversary could be internal or external and with/ without prior knowledge on users. According to the different assumptions, we formulate our objective of minimizing the privacy leakage level on the basis of information theory tools as a modified rate-distortion problem. More specifically, we look for a probabilistic privacy-preserving mechanism that we evaluate through simulating real collected MCS applications data.

The rest of this chapter is organized as follows. First, we discuss the motivations behind privacy preserving mechanisms in participatory sensing systems. Section 6.3 gives an insight into the related work to our proposal. We describe the different scenarios and measures studied in our system in Section 6.4. Correspondingly, we outline the mathematical formulation and the proposed algorithm in Sections 6.5 and 6.6, respectively. The performance evaluation of our fourth and last contribution, PRUM, is illustrated in Section 6.7. Conclusions and future work are withdrawn in Section 6.8.

6.2 Motivations and Context

Undeniably, reporting data in a community-scale task may incur private information leakage for participants. On one hand, most of participatory applications require location and sensing time tagged data for accuracy reasons. Furthermore, an adversary (malicious user) can derive sensitive information (e.g. a participant residence, income level, political affiliation, medical condition, etc.) by only observing the multiple reports in the system [103]. On the other hand, other applications do not access directly private information but can collect sensor readings that may be correlated with. For instance, by sending the energy consumption reports of different smart home equipment, users may release *unintentionally* their household activities to a service provider. As a result, users seem to be reluctant to participate to sensing campaigns unless with privacy protection guarantees.

To answer such concerns, a variety of privacy-preserving schemes have been proposed in the literature. Privacy preserving mechanisms inherit mainly from security and obfuscation techniques. Hence, different works [27, 53, 103–109] adopt pseudonymisation, anonymization, encryption and obfuscation methods in order to protect participants identities and personal data from being inferred. Nevertheless, some of these techniques come with an important computation complexity which may use up users devices batteries. Other methods, mainly anonymization and obfuscation, are based on reducing the accuracy of the reported information to a “general” value to make it common among a set of participants or perturbing data by adding random noise. Unfortunately, this impacts the usefulness of

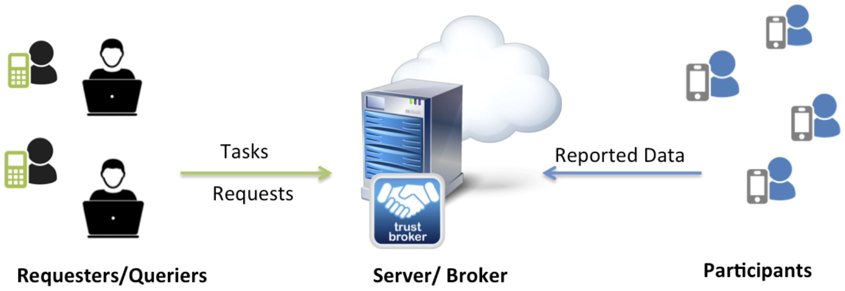


Figure 6.1: Participatory system with trusted entity

the collected data, a major criterion in sensing systems. Indeed, data utility is necessary to queriers (requesters) to offer services that meet user needs. Thus, any approach that gives priority only to the information privacy aspect while overlooking the resultant reduction in utility is not likely to be practically usable. To address utility and privacy competing goals, we propose in this work a privacy-preserving model which conserves requesters data utility requirements.

Last but not least, a participant may be registered within many crowdsensing platforms to contribute different data. These contributions do not reveal any sensitive information when analyzed separately. However, if an adversary collects side information from a previous sensing process or sensing platform and infers a current one, he can divulge private information about participants in common sensing campaigns. This scenario is considered as an extreme adversary model which we propose to solve with additional measures.

Based on these assumptions, we suggest to introduce first a trusted entity in the centralized participatory sensing system as described in Figure 6.1. This entity, denoted as the trust broker, could be merged with the central server, i.e., it receives queriers specific requests and collects data from participants. More specifically here, we propose to run, on this broker, a general obfuscation-based mechanism that aims to achieve a utility-privacy trade-off. The basic idea is to obfuscate the reported data by participants in order to minimize the privacy leakage with their private information without degrading the final released data. Our main contributions in this chapter can be thus summarized as follows:

- We present privacy and utility metrics to quantify the requirements of each part of the participatory sensing system: requesters and participants.
- We propose a privacy-preserving utility-aware mechanism that achieves a trade-off between the queriers requirements and participants privacy concerns while considering different uploading scenarios and adversary models.

6.3 Research Work on Privacy-preserving Mechanisms

In this section, we summarize the most relevant research work on privacy preserving techniques in participatory sensing systems. Also, we investigate the literature review on privacy-utility trade-off proposals in different domains.

6.3.1 Privacy-preserving Mechanisms in Participatory Sensing

Privacy-preserving mechanisms vary based on their adopted techniques. Jaimes et al. [102] and Christin et al. [54] have presented the taxonomy of such mechanisms into four major classes: cryptography, anonymization, data aggregation and obfuscation/perturbation.

6.3.1.1 Cryptography

This class relies on encryption methods as in the case of PEPSI [105], a privacy-enhanced architecture for crowdsensing applications. In this work, authors aim to hide sensing reports from unauthorized entities. Thus, each participant has to obtain an encryption key to cipher his collected data and a decryption key to be known by the end-user (service querier) to decipher it. This process requires high computation and energy resources which makes it unsuitable for most crowdsensing applications. Moreover, queriers can be curious to gather information about participants and hence the collected data should be distorted before being reported to them in order to protect participants sensitive information.

6.3.1.2 Anonymization Techniques

Another widely adopted privacy-aware technique in participatory sensing systems is the anonymization. This method aims to avoid the association between participants identities and their private information [102]. The simpler method is the pseudonymity [53,104] where participants share their sensors readings associated with pseudonyms instead of their real identities. However, this does not guarantee necessarily privacy given that a continuous inferring on reported data can divulge users identities. Similarly, other anonymization techniques [27, 106, 107] have been implemented especially for location-privacy. In this context, authors in [27] utilize the *Tesselation* technique to generalize one user location in a *Tile* containing k other users locations to guarantee *k-anonymity*. This results in a so-called spatial cloaking to prevent participants trajectories tracking [106]. Hence, anonymization is generally proposed to protect users location and is slightly studied in data reporting [107]. In addition, this technique can not be used in many crowdsensing applications that require accurate location tagged data.

6.3.1.3 Data Aggregation

Shi et al. [108] focus on privacy-preserving data aggregation in participatory sensing and propose a distributed-scheme called PriSense. The basic idea is not to rely on a central entity to provide protection to participants. However, participants tend to distribute their collected data among their neighbors. Upon receiving a request from the aggregation server, each participant returns his data and the remaining data of his neighbors. This reduces the probability to successfully attribute each sensor reading to its corresponding mobile user. Nevertheless, this scheme ignores the inter-participants threat. In fact, participants may be curious to collect information about each-other. Particularly, in competitive sensing campaigns, disclosing the bid of other users may help the participant to select an adequate price and get the proposed task [96].

6.3.1.4 Obfuscation/ Perturbation

Researchers adopt also data perturbation/obfuscation techniques which consist of perturbing the data by adding noise or by hiding some of its features [54]. Poolview [109] is a novel data perturbation mechanism that generates a noise model with similar characteristics to the phenomenon measured by the crowdsensing application. Then, this model is distributed to all participants so they can generate the noise locally and modify the configuration parameters regularly in order to enhance their privacy protection against historical attacks. The main drawback of these methods is that the data utility is slightly studied. Though, different privacy protection mechanisms incur an important data degradation and information loss, essentially, when modifying the attributes of the reported measurement as in the case of perturbation.

6.3.2 Privacy-utility Trade-off

In order to address the privacy-utility trade-off, the research community tends to apply information theory tools [110–113]. Particularly, using rate-distortion theory, authors in [111] have developed a utility-privacy trade-off region for databases. Their model is based on Shannon entropy which may be inadequate for some applications where individual anonymity guarantees are required. Calmon and Fawaz [112] have introduced a general framework for privacy against statistical inference and formulated a convex program to find privacy-preserving mappings for minimizing the information leakage from a user data with utility constraints. Both works address collective privacy in database systems while considering the data utility constraint. However, they present only theoretic metrics with no evaluation on real applications. Further, authors in [113] have extended the work of [112] with the aim of solving the privacy-utility issue in rating web applications. In this trend, we propose a general PRivacy-preserving Utility-aware Mechanism, PRUM [19], in participatory sensing systems for which we describe the analytic model in the following sections.

6.4 System Model

In this section, we introduce the different entities involved in a participatory sensing system with privacy protection policy. We define also the attacker as well as the different data reporting models.

6.4.1 Participatory Sensing Involved Entities

We consider, in this work, the system architecture of Figure 6.1 where two parties are communicating via a trusted entity as detailed below.

6.4.1.1 Participants

Set of smart-devices equipped users registered within a participatory sensing platform to contribute data. Thus, they are more likely to report a good quality contribution in order to get interesting “rewards”. However, they have privacy concerns and may require that a set of their contributed measurements should remain private. For instance, by reporting periodic temperature measurements along with collection points, participants may be exposing their trajectories to a service provider. Consequently, the latter can disclose their behavior and daily-life related information such as frequented places.

6.4.1.2 Queriers/Requesters

Set of service providers or individual users looking for a specific data and requiring a predefined quality level under which the information may be perceived as *useless*. That is, each requester sets a utility threshold value based on a quality metric that takes into account the specificity of the data collecting purpose. This value is communicated to the broker to be considered as the minimum accepted quality level for reported data.

6.4.1.3 Server/Broker

A central entity in the cloud that receives sensing requests from queriers and assigns the corresponding tasks to the available participants based on the adopted task assignment strategy. For simplicity reasons, we consider that the server plays the role of the trusted entity (broker) which collects the participants contributed data and receives the utility threshold values from queriers. The broker’s goal is to answer participants privacy concerns by minimizing the probability of inferring their private information from publicly reported measurements while respecting the data utility threshold value set by the queriers.

6.4.2 Adversary Model

In this work, we consider both partial-internal and external adversaries [114]. That is, an attacker might be part or not of the participatory sensing system. First, we assume that the server is the only trusted entity who collects data from participants with no curiosity or intentions of compromising their privacy. Second, we consider that other participants and queriers are honest but curious. Thus, they honestly report their data or their utility metrics, but they may try to learn about other users behaviors from their periodically released sensing data. Also, we consider that external adversaries may be eavesdropping as false queriers to the different collected measurements in order to disclose participants private information. Finally, we assume that adversaries are passive. That means, they can only read and observe data but they do not modify it. Nevertheless, we investigate two different cases as follows:

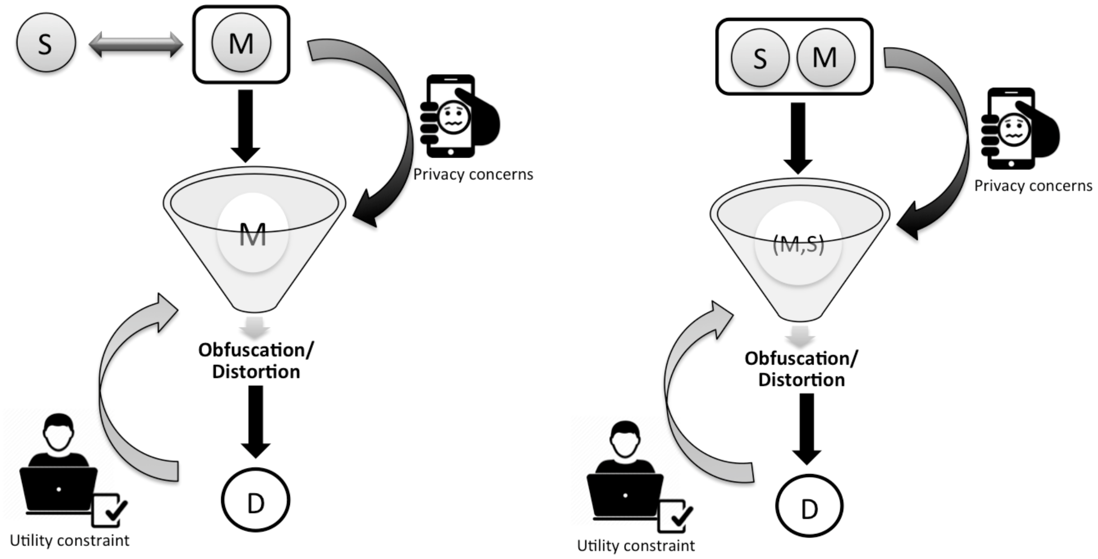
- **Adversary without side information**: the attacker is inferring only the current sensing process and do not have any additional information about participants.
- **Adversary with side information**: the attacker has already inferred reported measurements of participants in other sensing campaigns and is trying to correlate them with current reported data.

6.4.3 Scenarios of Data Reporting Model

In the following, we first define the necessary notations for our system elements and detail the two possible scenarios of data uploading in participatory sensing campaigns.

In the beginning, we define by $M \in \mathcal{M}$ a sensing measurement collected by a participant where $\mathcal{M}$ represents the set of all possible measurements. Indeed, M can be correlated with a participant private information denoted by $S \in \mathcal{S}$, where $\mathcal{S}$ denotes the set of all possible secrets detained by participants. The correlation between M and S can be expressed through the jointly random distribution $P_{M,S}(m, s)$, where $(m, s) \in \mathcal{M} \times \mathcal{S}$ are two realizations of the random variables M and S. This may compromise participants privacy since observing M may result in inferring S. Therefore, a participant would rather send his collected data M to the broker which should generate a distorted version $D \in \mathcal{D}$ to be reported to the end users such as queriers. Nevertheless, we distinguish throughout this work two data reporting scenarios:

- First, we tackle the case of a measurement M *correlated* with a secret S. For instance, a turned-on room light, considered as a measurement M, indicates its occupancy and thus a user indoor position considered as a secret S. In this case, only the measurement M should be obfuscated.



(a) S is correlated with M (b) S is reported with M

Figure 6.2: The different scenarios of data reporting and obfuscation

- Second, we consider the case of a secret S *jointly reported* with a measurement M. For example, in location tagged sensing tasks, the location S and the measurement M (e.g Temperature) are reported together to the MCS platform. Thus, both variables (M, S) should be distorted.

The aforementioned scenarios are illustrated in details in Figure 6.2. In the following, we design the objective of this contribution while accounting for different adversary models and reporting scenarios.

6.5 Problem Formulation

We aim to develop a privacy preserving mechanism in participatory sensing which conforms with data utility requirements. Therefore, we define adequate evaluation metrics for privacy leakage and data utility. Then, we formulate the corresponding optimization problem that models the trade-off we target.

6.5.1 Privacy Leakage Metric

We opt for data distortion techniques to modify a measurement M before reporting it to the querier. In order to do so, we remind the data distortion model adopted by Fawaz et al. [112] and we define our privacy mechanism as follows.

Definition 6.1 (Privacy-mapping [112]). A privacy-preserving model is a probabilistic mapping $f : \mathcal{M} \rightarrow \mathcal{D}$ characterized by the conditional probability $p_{D|M}(d|m)$ that minimizes the privacy risk to infer a private information S, jointly distributed with a public data M, from the released data D.

In addition to the privacy mapping $p_{D|M}$ presented above, we need to introduce a privacy leakage metric which quantifies the degree of information that one adversary can get about a secret S when observing the reported distorted data D. To this purpose, we should consider the adversary estimation model. According to [114], an adversary may perform a statistical inference on different measurements reported by participants to make a first estimation of the secret S. Further, when inferring the reported data D, the attacker estimates again the possible set of users secrets. The two estimations come with costs. Therefore, after observing the collected data, the adversary obtains an average cost gain defined by authors in [112] as:

$$\Delta C = C_0 - C_D \tag{6.1}$$

where C_0 and C_D are the estimation costs with no prior knowledge and after observing the released data, respectively.

With respect to this model and while considering a log-loss cost function, the adversary average cost gain expression turns into the *Mutual Information* [112] between the secret S and the released data D expressed as:

$$\begin{aligned} I(S, D) &= H(S) - H(S|D) \\ &= \sum_{s \in \mathcal{S}} \sum_{d \in \mathcal{D}} p_{S,D}(s, d) \log \left(\frac{p_{S,D}(s, d)}{p_S(s) p_D(d)} \right) \end{aligned} \tag{6.2}$$

where $H(S)$ is the entropy of S defined as: $H(S) = -\sum_{s \in \mathcal{S}} p_S(s) \log(p_S(s))$.

The mutual information quantifies how much information is shared between two random variables [114]. In our case, we measure by $I(S, D)$ the quantity of information shared between S and D, which represents the amount of information leaked from a privacy mechanism. This metric designs the privacy leakage function to be minimized.

6.5.2 Data Utility Metric

Hereafter, we take into consideration queriers requirements in terms of Quality of Information (QoI). Thus, we recall the utility function proposed by authors in [71] and utilized as a QoI metric in Chapter 3. This function is claimed to be appropriate given its general formulation that can be flexible for different sensor data. Nevertheless, any other less complex utility metric can be used such as the Euclidean distance in case of location-privacy or the Hamming distance in case of binary reported information.

Furthermore, we propose to set a threshold value on the utility of the reported data D, denoted as U_δ, under which the measurement is rejected. In other terms, the broker must report data D with utility $U_D \geq U_\delta$. Correspondingly, we define the utility regression level as the difference between the utility of a measurement M and the utility of the distorted data D, $U_M - U_D$. The maximum utility regression level is defined thus as follows.

Definition 6.2 (Maximum utility-regression level)**.** The maximum utility regression level tolerated by a requester (querier) is the difference between the utility of the sensing data M, U_M, and the utility threshold level, U_δ: $\delta = U_M - U_\delta$.

It is worth noting that the data utility metric is rather a constraint to be respected while looking for the privacy mapping $p_{D|M}(d|m)$. Thus, we formulate the corresponding constraint in Equation (6.3). We set the maximum accepted utility regression level for a querier U_δ, introduced in Definition 6.2, as an upper bound of the average distance between the utilities of the original measurements M and their distorted versions D.

$$\mathbb{E}_{M,D}[d(U_M, U_D)] \leq \delta \tag{6.3}$$

where $d(x, y)$ is a distance function between two variables x and y that depends on the obfuscation type.

Particularly, if we recall the second scenario of data reporting described in the preceding section, the utility of the temperature measurement M would be a measure of the accuracy of the measurement point (location) while U_D is the quantification of how accurate is the distorted version D after cloaking the location of M.

Let us now utilize the two defined privacy and utility metrics in order to formulate the corresponding optimization problem.

6.5.3 Optimization Problem

Our goal is to minimize participants incurred privacy leakage expressed by Equation (6.2) while respecting the utility constraint described by Equation (6.3). In the following, we model the *general* optimization where we target a privacy-utility trade-off:

$$\begin{aligned} \underset{p_{D|M}}{\text{Minimize:}} \quad & I(S, D) \\ \text{subject to:} \quad & \mathbb{E}_{M,D}[d(U_M, U_D)] \leq \delta \end{aligned} \tag{6.4}$$

This formulation illustrates our main target in this dissertation contribution. Essentially, we target to minimize the privacy leakage metric illustrated by the mutual information between a secret S and the released distorted data D. This should be conducted while considering the utility constraint quantified by the distance between a measurement M utility and its corresponding distorted data D utility.

In the following, we present a more detailed mathematical analysis that accounts for the two adversary models: without and with side information. Moreover, we investigate according to each adversary model the two different cases of data reporting scenarios.

6.5.3.1 Adversary without Side Information

We consider an adversary eavesdropping on an MCS data uploading/reporting phase in order to infer participants sensitive information S from their reported measurements M. Though, we investigate separately the case where S is only jointly distributed with M and the case where S is reported along with a measurement.

S correlated with M: First, we study the general model where a participant private information S is correlated with a collected data M. As a consequence, the three variables S, M and D can form a Markov chain: $S \to M \to D$. Thus, the joint distribution of S and D can be expressed as a function of $p_{S,M}$ and $p_{D|M}$ as follows:

$$p_{S,D}(s,d) = \sum_{m \in \mathcal{M}} p_{D|M}(d|m) p_{S,M}(s,m) \tag{6.5}$$

The expression of the joint distribution $p_{S,D}$ is introduced to elaborate the objective function of the optimization Problem (7.2). More precisely, we write this objective as a function of the conditional probability $p_{D|M}$ representing our privacy mapping and the joint probability $p_{S,M}$ representing the prior knowledge that the broker can statistically compute from the different historical reported data by participants.

$$\begin{aligned} I(S,D) = &\sum_{s,m,d} p_{D|M}(d|m) p_{S,M}(s,m) \times \\ &\log\left(\frac{\sum_{m"} p_{D|M}(d|m") p_{M|S}(m"|s)}{\sum_{s',m'} p_{D|M}(d|m') p_{S,M}(s',m')}\right) \\ &= g(p_{D|M}, p_{S,M}). \end{aligned} \tag{6.6}$$

Note that p_S, p_M and $p_{M|S}$ could be derived from the joint probability $p_{S,M}$. Also, the obtained expression $g(p_{D|M}, p_{S,M})$ responds to the form $ax\log(\frac{x}{z})$ which proves its convexity as stated in [112].

Furthermore, we approximate the distance between the utility of a measurement M and the distorted data D to any other distance depending on the obfuscation type, i.e., $\mathbb{E}_{M,D}[d(U_M, U_D)] \sim \mathbb{E}_{M,D}[d(M,D)]$. Then, we express the data utility constraint as a function of $p_{D|M}$ and $p_{S,M}$ by replacing $p_M(m) = \sum_{s \in \mathcal{S}} p_{S,M}(s,m)$ in the expectancy expression as follows.

$$\begin{aligned}\mathbb{E}_{M,D}[d(M,D)] &= \sum_{m,d} p_{M,D}(m,d)d(M,D) \\ &= \sum_{m,d} p_{D|M}(d|m)p_M(m)d(M,D) \\ &= \sum_{s,m,d} p_{D|M}(d|m)p_{S,M}(s,m)d(M,D) \\ &= h(p_{D|M}, p_{S,M}, d(M,D))\end{aligned} \tag{6.7}$$

Finally, we reformulate Problem (7.2) in order to enhance its dependency on the probabilistic privacy mapping $p_{D|M}$ and the prior knowledge $p_{S,M}$. The final formulation is expressed as:

$$\begin{aligned}&\underset{P_{D|M}}{\text{Minimize:}} && g(p_{D|M}, p_{S,M}) \\ &\text{subject to:} && h(p_{D|M}, p_{S,M}, d(M,D)) \leq \delta\end{aligned} \tag{6.8}$$

S reported with M: Problem (6.8) models the privacy-utility trade-off in participatory sensing for any reported measurement that may be correlated with a participant secret. Differently, we seize here the scenario of a secret S reported along with a measurement M illustrated in Figure 6.2(b). In this context, the prior knowledge is reduced to the marginal distribution of one measurement $\tilde{M} \sim (M,S)$. As a result, the formulation of our objective and constraint are simplified to the next following functions: $I(\tilde{M},D) = g'(p_{D|\tilde{M}}, p_{\tilde{M}})$ and $\mathbb{E}_{\tilde{M},D}[d(U_{\tilde{M}}, U_D)] = h'(p_{D|\tilde{M}}, p_{\tilde{M}}, d(\tilde{M},D))$.

6.5.3.2 Adversary with Side Information

Generally, participatory sensing tasks are performed periodically. Besides, a sensing platform may be responsible of collecting different data from participants and hence associated with different sensing applications on their smart handsets. Consequently, malicious or curious users eavesdropping on the data uploading phase can collect data from previous tasks and use it as additional information to infer current ones. For instance, an adversary can infer first a reporting phase of sensing temperature, $\mathcal{M}_1$. Further, he eavesdrops on humidity level measurements $\mathcal{M}_2$ reporting. Thus, when associating the two sets, $\mathcal{M}_1$ and $\mathcal{M}_2$, the attacker can have more information and hence limits the set of possible secrets S of corresponding participants in the two sensing tasks.

We propose to investigate such inference attack as an extreme case that our privacy preserving mapping should consider. Nonetheless, we only account for the case of a secret S correlated with a reported measurement M to avoid complexity of analytic formulation.

S correlated with M: Let $Y \in \mathcal{Y}$ be the set of side information an adversary has collected during previous sensing campaigns. That is, y is also associated with joint distribution to the secret S. In order to highlight the general dependency among all entities, we adapt the previous Markov chain as follows: $S \rightarrow (M, Y) \rightarrow D$. Consequently, the joint distribution of S and D can be expressed as function of $p_{S,M,Y}$ and $p_{D|(M,Y)}$ as follows.

$$p_{S,D}(s,d) = \sum_{m \in \mathcal{M}, y \in \mathcal{Y}} p_{D|(M,Y)}(d|(m,y)) p_{S,M,Y}(s,m,y) \tag{6.9}$$

We utilize the above formulation to assess the objective function of the general optimization problem defined by Equation (7.2) while considering the side information condition. This formulation depends mainly on the conditional probability $p_{D|(M,Y)}$ representing our extended privacy mapping and the joint probability between a secret S and the prior and current measurements, $p_{S,M,Y}$.

$$\begin{aligned} I(S,D) = \sum_{s,m,y,d} & p_{D|(M,Y)}(d|(m,y)) p_{S,M,Y}(s,m,y) \times \\ & \log\Big(\frac{\sum_{m',y'} p_{D|(M,Y)}(d|(m',y')) p_{S,M,Y}(s,m',y')}{p_S(s) \sum_{m",y"} p_{D|(M,Y)}(d|(m",y")) p_{M,Y}(m",y")}\Big) \\ & = f(p_{D|(M,Y)}, p_{S,M,Y}). \end{aligned} \tag{6.10}$$

The expression of Equation (6.10) respects the convexity condition as well, i.e., $ax \log(\frac{x}{z})$ [112]. Though, it is based on a joint distribution among the three variables S, M and Y which represents the prior knowledge from which the broker needs to derive the probabilities of p_S and $p_{M,Y}$ and computes the corresponding privacy preserving mapping $p_{D|(M,Y)}$.

Additionally, we require in this work to respect the maximum utility regression level introduced in Definition 6.2 between the current reported measurement M and its corresponding distorted version D. Therefore, we utilize the utility constraint developed in (6.7) while introducing the side information Y to obtain the new formulation below.

$$\begin{aligned} \mathbb{E}_{M,D}[d(M,D)] &= \sum_{m,d} p_{M,D}(m,d) d(M,D) \\ &= \sum_{m,y,d} p_{D|(M,Y)}(d|(m,y)) p_{M,Y}(m,y) d(M,D) \\ &= \sum_{s,m,y,d} p_{D|(M,Y)}(d|(m,y)) p_{S,M,Y}(s,m,y) d(M,D) \\ &= L(p_{D|(M,Y)}, p_{S,M,Y}, d(M,D)) \end{aligned} \tag{6.11}$$

Finally, we model the problem of looking for a privacy preserving mechanism while considering utility constraints and adversary with side information as follows:

$$\begin{aligned} \underset{p_{D|(M,Y)}}{\text{Minimize:}} \quad & f(p_{D|(M,Y)}, p_{S,M,Y}) \\ \text{subject to:} \quad & L(p_{D|(M,Y)}, p_{S,M,Y}, d(M,D)) \leq \delta \end{aligned} \tag{6.12}$$

Based on these assumptions, we introduce a Privacy-preserving Utility-aware Mechanism for data uploading in participatory sensing which account for previous formulations. PRUM [19] represents then our fourth contribution in this dissertation and aims at solving the optimization problems presented above while focusing first on the privacy-utility trade-off issue for both scenarios of data uploading. Further, we extend this algorithm to seize the case of adversary with side information. The proposed solution is designed in details in the following section.

6.6 Proposed Solution: PRUM

In the previous section, we have formulated the aim of this work and analyzed the associated use cases. Accordingly, we develop hereafter the PRUM algorithm to achieve a privacy-utility trade-off in participatory sensing systems. Moreover, we extend the proposed mechanism to tackle the different issues that may be encountered due to MCS data characteristics such as the mapping of distorted data and large size data alphabets.

6.6.1 Proposed Algorithm

Problems (6.8) and (6.12) are convex and bear some approximation to modified rate-distortion problems [112] and rate-distortion problem with side information, respectively. Such problems are widely studied in information theory with the aim of computing the bounds of a specific distortion. However, few applications have been developed based on this model given the necessity of using a dual minimization procedure analogous to the Arimoto-Balhut algorithm [115]. Nevertheless, by using the formulation of (6.8) and (6.12), we may rely on standard and efficient algorithms for solving convex optimization such as the Interior-Point Method (IPM) [116] or the Successive Quadratic Programming (SQP) [117].

With all this in mind, we introduce a first algorithm to study the two scenarios of Figure 6.2 associated with the case of an adversary with no prior information. These data uploading scenarios differ mainly in the formulation and in the prior knowledge but not in the solving method. Without loss of generality, we assume that the joint probability distribution between a private information S and a collected data M, $p_{S,M}$, is known by the broker by analyzing statistically the multiple sensing reports and thus can be the input of our mechanism. Besides, for simplicity reasons, we suppose that all queriers for a same type of data have the same maximum utility regression level, δ.

Algorithm 7 Privacy-Utility Trade-off Algorithm

```
Require: Joint probability $p_{S,M}(s,m)$, Maximum utility regression level $\delta$,
1: Obfuscation type: Generate the corresponding $\mathcal{D}$
2: Choose the distortion/utility metric $U$ and compute the distance $d(U_M, U_D) \sim d(M, D)$

3: Set the vector of variables $X = p_{D|M}$
4: Interior-point method:
   $\underset{X}{\textbf{Minimize}} : g(X, p_{S,M})$
   subject to:
       1. $h(X, p_{S,M}, d(M,D)) \leq \delta$
       2. $AX = b$
       3. $\sum X = 1, \forall D \in \mathcal{D}$
       4. $0 \leq X \leq 1, \forall D \in \mathcal{D}, M \in \mathcal{M}$
5: Return $p_{D|M}$
```

Additionally, we design the targeted privacy preserving mechanism as a probabilistic privacy mapping $p_{D|M}$. Therefore, we add two extra constraints to Problem (6.8) to limit the set of possible solutions. First, we remind that the sum of probabilities must be equal to 1 as follows: $\sum_{D\in\mathcal{D}} p_{D|M} = 1, \forall M \in \mathcal{M}$. Also, we force the achieved values to be between 0 and 1, $0 \leq p_{D|M} \leq 1, \forall D \in \mathcal{D}, M \in \mathcal{M}$. PRUM is illustrated in details by Algorithm 7.

Finally, we are interested to solve the minimization of privacy leakage while the adversary has side information Y collected from previous sensing tasks. The main objective is then to minimize with best effort the mutual information representing the privacy leakage, $I(S, D)$. Yet, we always limit our distortion technique by the utility requirements set by data queriers. The steps of the extended PRUM algorithm are as follows:

1. We compute the prior knowledge of the broker as the joint distribution between the secret and the two observable measurements $p_{S,M,Y}$.
2. We derive necessary probabilities: $P_{M,Y}$ and p_S.
3. We select an obfuscation technique based on the data type.
4. We generate the set of possible distorted data $\mathcal{D}$ and compute the distance $d(U_M, U_D)$.
5. We run the IPM method to solve Problem (6.12) and generate $p_{D|(M,Y)}$.

In order to minimize the complexity of solving Algorithm 7, we study, in the following, the potential encountered challenges specific to participatory sensing applications.

6.6.2 Mapping of Distorted Data

This issue is mainly related to the obfuscation technique we are adopting in this work. In fact, the set of the distorted data $\mathcal{D}$ generated by a broker might be larger/smaller than or equal to the set of measurements $\mathcal{M}$ collected by participants depending on the type of obfuscation used. Hence, some mapping from M to D may not be possible, i.e., $p_{D|M} = 0$ or $p_{D|(M,Y)} = 0$. To avoid time complexity issues, we set such probabilities to be null. Therefore, we form a boolean matrix A with ones in the positions of potential zeros probabilities and zeros otherwise. Besides, we add the following condition to our optimization problem: $AX = b$, where $X = p_{D|M}$ or $X = p_{D|(M,Y)}$ depending on the use case and b is a zeros vector with the same size of X.

6.6.3 Data with Large Size Alphabets

Most of participatory sensing collected data is with large size alphabets. Consequently, the size of the set of measurements $\mathcal{M}$ can be very important which results in a challenging estimation of the prior knowledge $p_{S,M}$ or $p_{S,M,Y}$. Furthermore, the set of distorted data $\mathcal{D}$ can be, depending on the selected obfuscation type, as large as $\mathcal{M}$. Thus, the solving of the predefined optimization Problems (6.8) or (6.12) could be complex and time consuming. Also, regarding the non-linear nature of the objective functions $g(p_{D|M}, p_{S,M})$ and $f(p_{D|(M,Y)}, p_{S,M,Y})$, we should minimize the size of data to be distorted in order to utilize the standard convex solvers.

To deal with such an issue, we opt for the well-known quantization techniques [118]. We differ the quantization method based on sensing application type. For instance, if the sensing task targets collecting location-tagged data, we can opt for clustering the sensing area into small cells. Though, we must highlight that the quantization step represents an additional distortion source which may yield to sub-optimal privacy-utility trade-offs. Based on this, we generate a new set of data alphabets that we denote by $\mathcal{Q}$ and we compute the corresponding prior knowledge $p_{S,Q}$ or $p_{S,Q,Y}$ which represents the new input of Algorithm 7. The resulted privacy mapping is defined as $p_{D_q|Q}$ and $p_{D_q|(Q,Y)}$, respectively, with $D_q \in \mathcal{N}$ is the set of distorted quantized-data. Finally, we map the obtained probabilities values to their corresponding original alphabets.

6.7 Performance Evaluation

We validate our fourth and last contribution while relying on three different real sensing datasets to be designed as crowd-sensed data. In this section, we describe the dataset, the simulations settings and the obtained results.

6.7.1 Sensing Datasets

6.7.1.1 Occupancy Detection Data

This dataset was generated by authors in [119] to evaluate the accuracy of different classification Machine Learning algorithms. The experiment consists of detecting if a room is occupied or not based on environment sensors readings. The sensors collect ambient temperature, humidity, light and CO_2 level measurements. For PRUM evaluation, we assume that such readings are measurements M reported via a crowdsensing application to a service provider in order to monitor smart-house connected objects. However, when reporting these measurements, a participant can expose his position which is considered as the secret S since it reveals his *indoor* lifestyle.

6.7.1.2 GPS Trajectory Data

This data has been collected from the GO_Track mobile application [120] and has been studied by authors in [121] to identify similar trajectories for carpooling purpose. The readings include rating of the traffic, weather and transportation besides collecting the visited geographical points, the vehicle speed and the total realized distance. Accordingly, we consider such traffic rating crowdsensing application which reports periodically these measurements. Also, we assume that the private information S can be participants driving behavior or transport mode. The former secret information can be inferred from reported speed, weather and traffic status values. While the latter can be directly derived from reported data about transportation mode or from correlated reported measurements such as weather and transportation ratings.

6.7.1.3 Crowd Temperature Data

This dataset is contributed by Mohannad et al. [122] as collected outdoor temperature values by taxis in Rome. Taxicabs are equipped with temperature sensors attached to their vehicles which report their readings along with the corresponding geographical position to a central server every 6 hours. Here, we suppose that such measurements are gathered via a participatory sensing application. Note that participants private information (location) is sent along with measurements, which represents the scenario of Figure 6.2(b).

6.7.2 Simulations Settings

We run our privacy-preserving utility-aware mechanism, PRUM, using the convex solver Interior Point Method built in Matlab on the datasets presented above while studying different use cases. Therefore, we conduct two different groups of simulations. First, we

Table 6.1: Occupancy data characteristics

Datasets	Data Distribution	
	0 (non occupied)	1 (occupied)
Dataset1 (closed door)	0.64	0.36
Dataset2 (open door)	0.79	0.21

tackle the case of an adversary with no side information and investigate the trade-off region between privacy and data utility in participatory sensing. Further, we look into the case of an adversary with side information and plot the new achieved privacy level. Additionally, we recall the utility metric defined by Equation (3.1) to measure achieved data utility values, and the mutual information defined by Equation (6.2) to quantify the privacy leakage. Besides, we set an inversely proportional privacy score as $P_s = 1 - I(S, D)$.

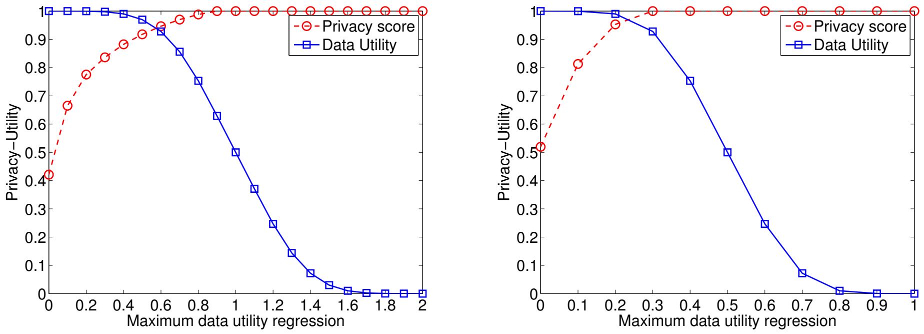
6.7.3 Privacy-Utility Trade-off

In this part of evaluation, we assume that adversaries are with no additional information. Simulation results of different dataset and data uploading scenarios are illustrated in Figures 6.3 - 6.5. Note that we vary the factors that may impact the trade-off region among the different applications to analyze our solution efficiency.

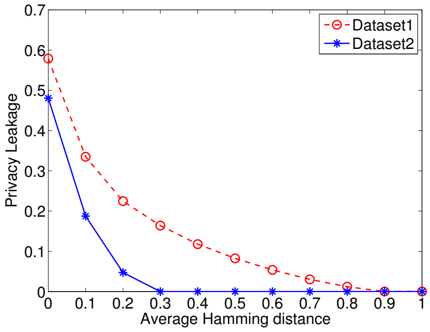
6.7.3.1 Smart-house Monitoring Scenario

We assume that the occupancy detection data [119] is generated by a smart-home monitoring application and utilize the two datasets collected with room door open and closed as detailed in Table 6.1. Let $S \in \{0, 1\}$ be the room occupancy status where $S = 1$ for an occupied room and 0 otherwise. Also, we denote by $M = \{T, H, L, CO_2\}$ and $D = \{\tilde{T}, \tilde{H}, \tilde{L}, \tilde{CO_2}\}$ the vector of measurements reported by a participant to the broker and by the broker to the requesters, respectively. These sensors readings are large size alphabets data. Therefore, we apply a quantization step that clusters the values into significant intervals based on identified thresholds in [119]. Given the important number of features in M, we opt for an *exchange-distortion* technique to generate a set of distorted data $\mathcal{D}$ of a same size as $\mathcal{M}$. That is, we perturb the measurements vector M by exchanging its elements values with others in the set $\mathcal{M}$.

Figures 6.3(a) and 6.3(b) show the achieved utility-privacy trade-off for the two datasets. The utility metric is naturally a decreasing function of the data regression level δ which implies the amount of distortion applied to M. That is, the more the data is perturbed, the less useful it is. Nevertheless, the proposed method has reached a utility-privacy trade-off with 90% of privacy protection level and less than 10% of data-utility loss for both datasets. Figure 6.3(c) shows the evolution of the privacy leakage, $I(S, D)$, as function of the average obfuscation level measured by the Hamming distance for the two datasets.



(a) Privacy-Utility Tradeoff for Dataset1 (b) Privacy-Utility Tradeoff for Dataset2



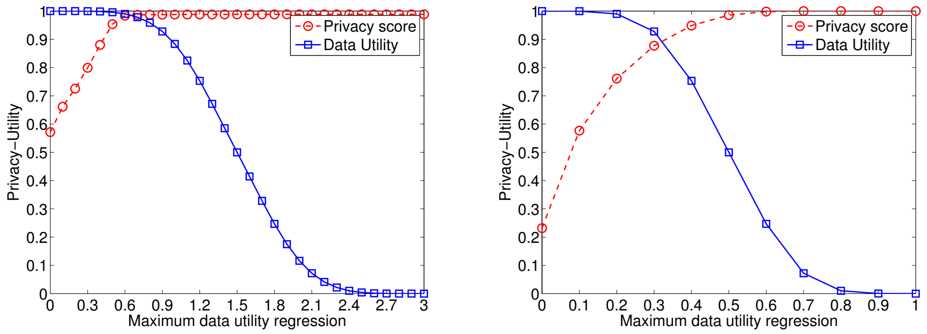
(c) Privacy leakage-Dataset1 & Dataset2

Figure 6.3: Privacy-Utility plots for Occupancy Data

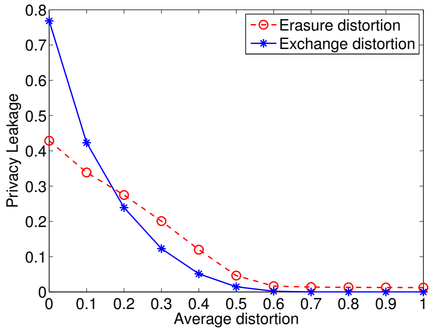
Clearly, the privacy leakage is minimized when data perturbation level increases until we achieve a perfect privacy, $I(S, D) = 0$. We observe that, for Dataset1, the perfect privacy requires distorting at least one feature of the measurements vector M whereas an average of 30% distorted data was enough to reach the same privacy level for Dataset2. This is due to the difference of the data distribution as detailed in Table 6.1. Dataset1 has a balanced distribution of occupancy measurements which makes distinguishing the secret harder while Dataset2 consists of more non occupied-room reported readings. Indeed, this distribution presents the prior knowledge of the broker about the participants' data. Hence, the more balanced the joint distribution $p_{S,M}$ (by reporting different readings) is, the harder to compromise one's private information.

6.7.3.2 Traffic Rating Application

For the traffic rating participatory sensing application, we consider as a secret S the driving behavior of a participant and by $M = \{speed, R, W\}$ the reported values for the vehicle speed, the traffic and the weather rating, respectively. Similar to the occupancy data, we notice that the reported speed measurements are large size alphabets data and we quantize it by setting three different intervals: $([0, 30], [30, 50], [\geq 50])km.h^{-1}$. Besides, we vary the type of obfuscation to generate D and we plot the corresponding results in Figure 6.4.



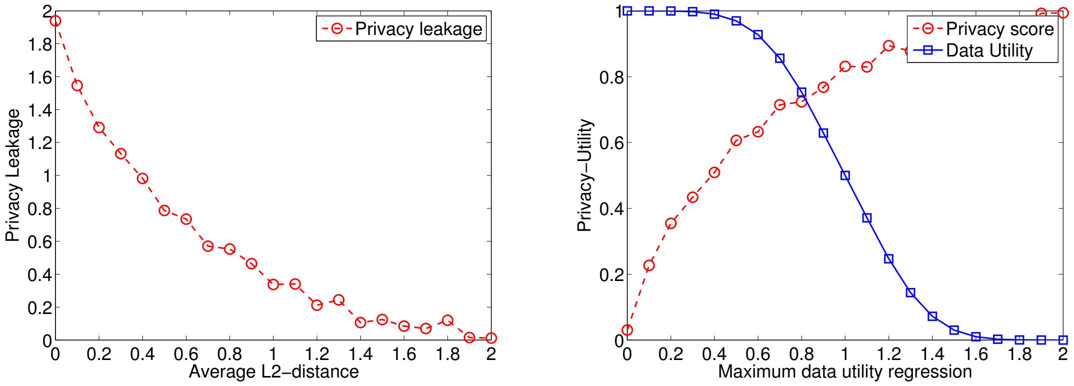
(a) Privacy-Utility Tradeoff for erasure-distortion (b) Privacy-Utility Tradeoff for exchange-distortion



(c) Privacy leakage exchange Vs erasure distortion

Figure 6.4: Privacy-Utility plots for GPS Data

First, we set the obfuscation technique as the *erasure-distortion*, i.e., we hide one or more features of the measurements vector M. The obfuscation level is quantified by the average number of erasures and varies from 0 to 3. Figure 6.4(a) shows the achieved privacy and utility values for different data-utility regression levels. We notice that hiding one feature among 60% of the collected data is enough to obtain the maximum privacy level. Moreover, PRUM realizes such privacy protection while maintaining more than 95% of data utility. Further, we set the obfuscation type as the *exchange-distortion* measured by the Hamming distance. Similarly, we show in Figure 6.4(b) that, differently from the *erasure-distortion* simulation, the privacy-utility trade-off is obtained for only 35% of data obfuscated while realizing an important data quality level. Finally, we compare the impact of varying the obfuscation type in Figure 6.4(c). We observe that an average of 0.1 *erasure-distortion* data ensures lower privacy leakage values than the *exchange-distortion*. This confirms that hiding a feature of the vector M has an interesting impact on the privacy risk whereas exchanging it with another value affects mainly the data joint distribution. Nevertheless, by exchanging more values, $\delta \geq 0.2$, we obtain better privacy levels. Note also that the minimum achieved privacy leakage level by *erasure-distortion* is $I(S, D) = 0.0117$ while the *exchange-distortion* realizes a perfect privacy. This highlights the importance of the obfuscation type selection in order to ensure better privacy for similar data regression levels.



(a) Privacy leakage (b) Privacy-Utility Trade-off

Figure 6.5: Privacy-Utility for temperature crowdsensing app

6.7.3.3 Crowd-Temperature Application

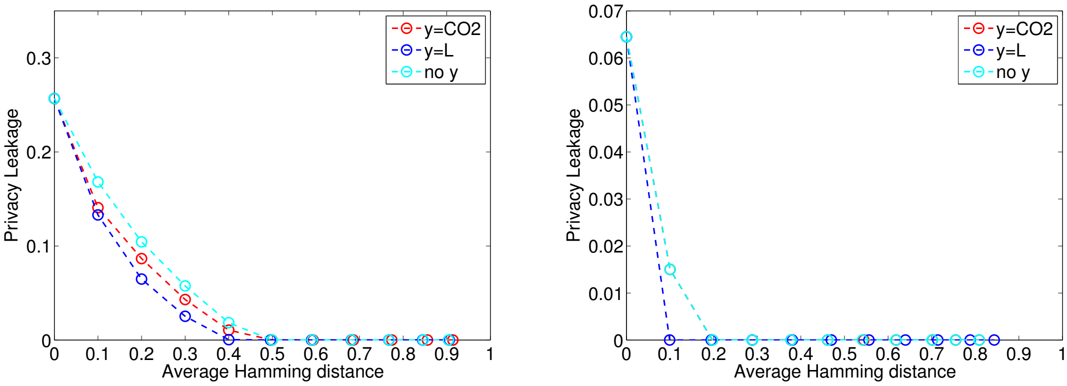
Hereafter, we investigate the scenario of Figure 6.2(b) for location tagged reported temperature measurements. Hence, we study the efficiency of our method while using the marginal distribution of $\tilde{M} = (M, S)$ instead of the joint one. To do so, we first cluster the sensing area into 9 sub-areas and we consider the *exchange-distortion* on participants locations. That is, users may report close by sub-areas tags along with their sensors readings rather than their real positions. Figure 6.5 plots the obtained privacy and utility levels for different L2-distance (Euclidean) values. We observe that the privacy leakage is decreasing slowly and reaches its minimum level for a distance ≥ 2 blocs. The privacy-utility trade-off is scaled by changing 80% of the location data to the cost of minimizing the utility of data to 70%. This slightly more important cost, compared to the first scenario, is due to the fact that location is distorted, however necessary for the accuracy of the reported temperature readings. Indeed, we achieve by our privacy mechanism a better privacy level for participants, but lower data quality.

6.7.4 Side Information Impact on Privacy Leakage

In this second evaluation phase, we simulate participatory sensing applications datasets while considering adversaries with side information. Hence, we modify the different measurement vectors M and introduce side information variables y for the smart-house monitoring app (occupancy detection data) [119] and the traffic rating app [121].

6.7.4.1 Smart-house Monitoring Scenario

We recall the occupancy detection data [119] and maintain the use of secret S as a participant indoor position. Though, we assume that the reported measurement vector is



(a) Privacy leakage for Dataset1

(b) Privacy leakage for Dataset2

Figure 6.6: Privacy leakage with side information for Occupancy Data

$M = \{T, H\}$. Furthermore, we suppose that adversaries have inferred previous uploading phases of CO_2 and light L measurements. Hence, we vary $y \in \{CO_2, L\}$ and observe its impact on the privacy leakage level measured by $I(S, D)$ when compared to the scenario of no side information, "no y". Results are illustrated in Figure 6.6.

First of all, we observe that all line graphs start from the same level of privacy leakage. This is explained by the fact that, if no distortion is applied to the reported data then $D = M$ and consequently $I(S, D) = I(S, M)$. This level of privacy leakage is then obtained without or with side information equally. Nevertheless, we remark that the lowest privacy leakage is obtained when $y = L$, i.e., the side information is the light measurement. Generally, the light L is the most indicative factor of a room occupancy. Therefore, if the adversary detains such information, then a high distortion level to the measurement $M = \{T, H\}$ will only incur lower data quality and do not guarantee higher privacy level. In other terms, the less indicative is y, the more distortion we need on the current M. The above observations are valid for both datasets.

6.7.4.2 Traffic Rating Application

For this application, we run two simulations scenarios. First, we consider a participant driving behavior as a secret, $S = bh$. Second, we assume that a participant considers his transportation mode as a sensitive information, i.e., $S = Tm$. For the two use cases, the reported measurement vector consists of speed and traffic rating values, $M = \{speed, R\}$. However, we vary the side information $y \in \{R_w, R_b, cb\}$ where R_w is the rating of weather, R_b is the bus rating (crowded or not) and cb is a boolean variable to describe the transportation mode of users, i.e., car or bus. It is worth mentioning that in this evaluation we simulated only the *exchange-distortion* technique given the high complexity of the *erasure-distortion*, especially when associated with the solving process of Problem (6.12). This is observed by the computation time shown in Figure 6.7.

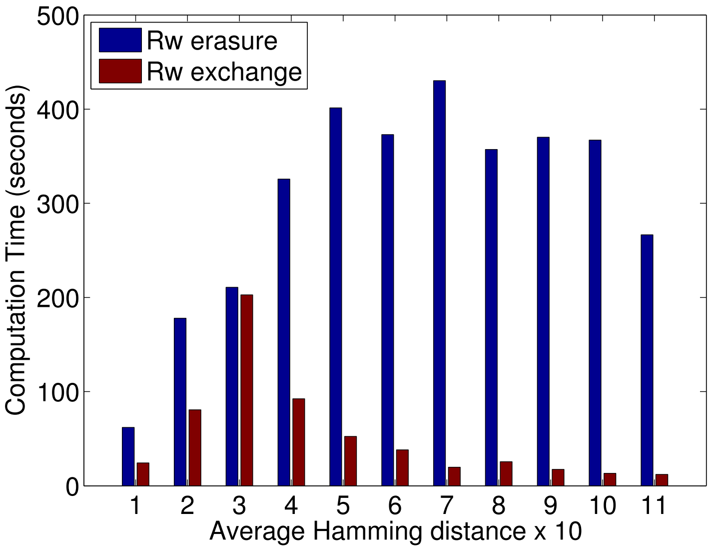
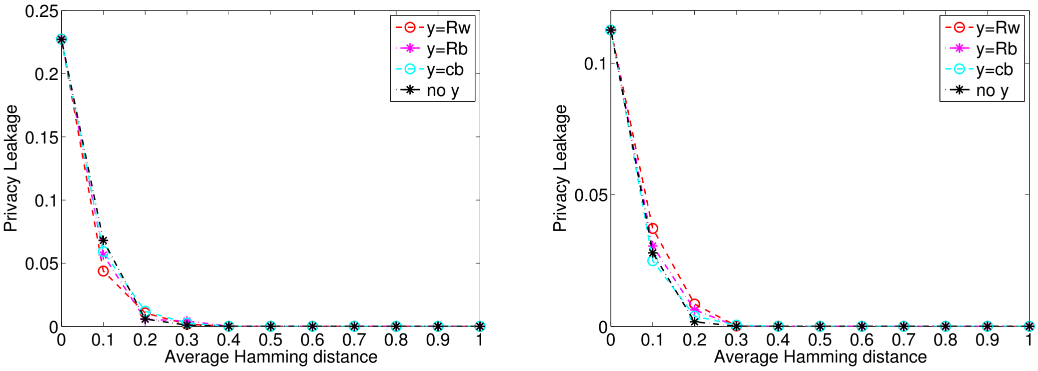


Figure 6.7: Computation time with exchange Vs erasure distortion techniques



(a) Privacy leakage for Secret= Driving behavior (b) Privacy leakage for Secret= Transport Mode

Figure 6.8: Privacy leakage with side information for Traffic Rating

Secret= Driving behavior: In this case, the measurement vector $M = \{speed, R\}$ includes indicative features about a user driving behavior. Particularly, when the speed is correlated with the rating of traffic and weather Rw, an adversary can easily infer this sensitive information. Therefore, we observe in Figure 6.8 that all line graphs for different side information y achieve comparable privacy leakage even when compared to the case of no side information. That means that the privacy level highly depends on the distortion applied to M. As a consequence, the perfect privacy, $I(S, D) = 0$, is reached for all cases with an average *exchange-distortion* of 0.4, i.e., 40% of reported data must be obfuscated.

Secret= Transport Mode: Here, the secret $S = Tm$ depends mainly on the reported data linked to transportation mode. As a result, the most indicative features are the rating of bus Rb and especially the boolean variable cb which is in this case the secret S. Hence, we observe that without side information or with $y = cb$ our method achieves the same privacy leakage. This is explained as in the case of occupancy data [119]. Thus, if the adversary has the closest feature to the secret, our algorithm does not *heavily* distort the current measurement respecting the utility constraint.

To conclude, our method, PRUM, is initially designed to look for a privacy-preserving mapping of a measurement M into a distorted version D in order to minimize the privacy leakage $I(S, D)$. Accordingly, knowing that the adversary detains a side information y only impacts the level of distortion to be applied to guarantee better data quality but does not prevent the leakage induced from y itself.

6.8 Discussion

In this chapter, we focus on participants privacy concerns in crowdsensing. Particularly, we tackle the question of data protection during the *data uploading* phase using obfuscation techniques. However, we consider data quality requirements studied in previous chapters as a major criterion for a successful sensing process. Therefore, we proposed to develop a privacy preserving mechanism while considering the data utility condition.

In this context, we presented first adequate metrics to quantify the privacy protection level of participants and the data utility threshold set by queriers. Furthermore, we investigated the different adversary models coupled with data uploading scenarios. Accordingly, we propose to rely on a trusted entity aiming at minimizing the privacy leakage while respecting the data utility constraint. The proposed algorithm, PRUM, runs a convex solving method, the Interior Point Method, to generate a probabilistic privacy mapping that describes the level of distortion necessary to achieve privacy-utility trade-off.

We validate this solution on real sensing datasets with different use cases. PRUM has been proved to be efficient while studying various factors. For instance, we achieved the trade-off region for different data distributions and obfuscation techniques. Nevertheless, we identified lower data quality when reporting sensitive information along with measurements. Besides, we remark that the privacy leakage induced in the adversary with side information case is minimized for a same distortion level when compared to the general scenario.

In summary, a privacy-preserving must consider many factors such as the obfuscation type as well as the prior knowledge of the trusted entity and the attacker. Additionally, recent work on crowdsensing suggests to compensate participants privacy leakage with incentives. Hence, participants may accept to get lower protection level when being rewarded. Such a condition may enlarge the region of the trade-off investigated in this work and enhance contributions with less distortion levels which guarantees higher quality level of the reported sensory data.

Chapter 7

Conclusion and Perspectives

Contents

7.1 Contributions 113
7.2 Perspectives 114

Mobile CrowdSensing (MCS) is envisioned to be an efficient paradigm of human-centric sensing by assessing various applications such as environment measurements, traffic control and health monitoring. Nevertheless, it is due to such human involvement that crowdsensing systems need to seize different challenges that can be distinguished as requesters concerns and participants concerns. Accordingly, we developed in this dissertation appropriate scheduling schemes that answer the general question of: *How to design efficient crowdsensing phases that respect both participants and requesters requirements?*

We summarize in this final chapter the main contributions of this dissertation. Besides, we discuss the possible future work.

7.1 Contributions

Throughout this dissertation, we tackled three phases of MCS: *Task Assignment*, *Sensing and Processing* and *Data Uploading*. For each of these steps, we listed potential encountered issues for participants and/or requesters then proposed adequate measures and solutions.

Initially, we were interested in the *Task Assignment* phase to set collaborative sensing among users in the aim of cutting off MCS cost and providing better data quality contributions. Undeniably, this phase predetermines who will sense? when? and where? Consequently, an efficient scheduling of sensing tasks among participants reduces their individual energetic consumption devoted to MCS campaigns. As well, it prohibits collecting data samples from poor sensing regions to avoid accuracy issues. The proposed Quality and Energy-aware Mobile Sensing Schemes (QEMSS and F-QEMSS) were proved to reduce the overall cost of sensing and respect individual use of participants devices besides achieving important data quality levels when compared to state-of-the-art methods. Particularly, our first contribution methods have enhanced sensing within low dense areas and with participants short of resources by varying the schedules which answered both participants energetic concerns and requesters data quality ones.

Our second and third contributions are presented to assess assignment methods discussed above in a distributed fashion. Specifically, we first focused on minimizing the necessary time for the *Sensing and Processing* phase of MCS. Thereupon, we addressed participants concerns about performing *heavy* load of sensing that may use up their devices along with the necessity of requesters to get *fresh* data. The introduced distributed assignment conducted by some users designated as *requesters* has studied users mobility and preferences in terms of tasks acceptance in the aim of reducing tasks *loss* and non accurate measurements. Participants' commitment was then investigated when introducing incentivizing mechanisms to a users discrete choice model which depends on proposed tasks attributes (number, load , reward..). Incentive concerns for participants raise budget issues within requesters. Hence, we proposed two different incentive policies to observe the most efficient one in terms of budget allocation. The proposed schemes through Chapters 4 and 5 have proved to achieve their main goal of minimizing the average sensing and processing time in MCS tasks. In addition, the Incentive-based preference -aware scheme was proved to be the most promising especially when coupled with quality-based rewarding.

Finally, we studied users privacy concerns during the MCS *Data Uploading* phase. Privacy protection is a primary issue tackled for MCS given that users are sharing data in a community scale which accentuates the risk of inferring their sensitive information. Nevertheless, major proposed privacy preserving schemes addressed this challenge when focusing mainly on participants concerns and slightly analyzing side impact on data quality. Therefore, we targeted a trade-off among the two competing goals of distorting data to minimize induced privacy leakage and maintaining a good data quality level defined as a utility. The proposed Privacy-preserving Utility-aware Mechanism (PRUM) determines the distortion level necessary to realize such trade-off under different scenarios. As a result, the proposed algorithm evaluated on many sensing applications dataset has scaled queriers utility requirements while guaranteeing an important privacy protection level for participants.

In summary, the key contributions presented in this thesis have targeted different optimization objectives within MCS systems. Yet, different from most research work on MCS, we developed solutions tackling simultaneously participants and requesters concerns.

7.2 Perspectives

This dissertation has enabled us to study the promising MCS paradigm and advance new research work on its different challenges besides leveraging new reflections for future work:

Quality of Information: Through our first contribution, we defined a general Quality of Information utility metric. However, we selected three data quality attributes: "completeness", "timeliness" and "affordability". Hence, one can extend the proposed quality measure to any other data attributes such as reliability, re-usability, etc. Thereupon, we can intensify the proposed quality-aware mobile sensing schemes to include as much data quality metrics as necessary to answer the targeted sensing application characteristics.

General MCS Framework: The presented assignment schemes in both centralized and distributed fashions were investigated separately but claimed to be complementary. Accordingly, a future advance on our work can tackle the *fusion* of the two first contributions of this dissertation. It is worth noting that we had worked on a general prototype of MCS platform, Sensarena [29] in terms of collecting various sensing measurements. Hence, this framework can be a first testbed to validate the potential coexistence of QEMSS [16] and IP-MATA schemes and investigate their impact on recruiting participants in MCS systems.

Privacy Measures: Our last contribution focused on privacy leakage concerns during the MCS data uploading phase. The proposed architecture enables a trust entity to distort participants collected data before being reported to requesters while respecting data quality settings. That is, the measurement is reported clear from participants to the trust broker, then distorted to be reported to queriers. Yet, such architecture did not seize the issue of *curious* participants who can infer their neighbors information during the reporting phase to the trust entity. Indeed, in competitive sensing campaigns, users should submit high data quality samples or low bids to be selected. As a result, inferring other participants information enables a user to improve his own offer and get the task. An open research question is how to move data obfuscation techniques from the cloud to the device itself or other computation edge techniques to reduce the privacy risk on users and do not incur additional energetic one.

Incentives Vs Privacy: Another reflection on introducing privacy to MCS systems is to which extent a user can divulge his private information? Indeed, MCS relies essentially on the *open* commitment of participants to provide sensory data. Moreover, smart devices users are usually accepting charts of downloaded applications that ask to access their information (contacts, photos, microphone, location, etc.) in order to take advantage of their services. Therefore, we suppose that participants may accept to provide sensing samples correlated with their private information if receiving incentives in return. This assumption can be coupled with our fourth contribution to emphasize the targeted trade-off between data quality and privacy satisfying levels.

List of Publications

International Journals

- **Rim Ben Messaoud**, Yacine Ghamri-Doudane and Dimitri Botvich, “Incentives-based Preference and Mobility-aware Task Assignment in Participatory Sensing Systems”. Under Review to Special Issue of Computer Communications Journal.
- **Rim Ben Messaoud**, Nouha Sghaier, Mohamed Ali Moussa and Yacine Ghamri-Doudane,“Privacy Preserving Utility-aware Mechanism for Data Uploading phase in Participatory Sensing”. Submitted to IEEE Transactions on Mobile Computing, June 2017.

International Conferences

- **Rim Ben Messaoud**, Nouha Sghaier, Mohamed Ali Moussa and Yacine Ghamri-Doudane,“On the privacy-utility tradeoff in participatory sensing systems”. 15th IEEE International Symposium on Network Computing and Applications (NCA), Cambridge, MA, USA, pp.294-301, 2016.
- **Rim Ben Messaoud**, Yacine Ghamri-Doudane and Dimitri Botvich,“Preference and Mobility-Aware Task Assignment in Participatory Sensing”. 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Malta, pp.93-101, 2016.
- **Rim Ben Messaoud** and Yacine Ghamri-Doudane,“Fair QoI and Energy-aware Task Allocation in Participatory Sensing”. IEEE Wireless Communications and Networking Conference (WCNC), Doha, Qatar, pp.1-6, 2016.
- **Rim Ben Messaoud**, Zeineb Rejiba and Yacine Ghamri-Doudane, An Energy-aware End-to-End crowdsensing platform: Sensarena”. 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, pp.284-285, 2016.
- **Rim Ben Messaoud** and Yacine Ghamri-Doudane,“QoI and Energy-aware Mobile Sensing Scheme: A tabu-search approach”. 82nd IEEE Vehicular Technology Conference: VTC 2015-Fall, Boston, US, pp.1-6, 2015.

National Conferences

- **Rim Ben Messaoud** and Yacine Ghamri-Doudane, “QEMSS: A selection scheme for participatory sensing tasks”. International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS), Paris, France, pp.1-6, 2015.

- **Rim Ben Messaoud** and Yacine Ghamri-Doudane “Allocation équitable des tâches de collecte participative de données” Rencontres Francophones sur la Conception de protocoles, lévaluation de performance et lexpérimentation des réseaux de communication (CORES) 2016, Bayonne, France

References

[1] N.D. Lane, E. Miluzzo, Hong Lu, D. Peebles, T. Choudhury, and A.T. Campbell. A survey of mobile phone sensing. *IEEE Communications Magazine*, 48(9):140–150, 2010.

[2] The statistical portal. `https://www.statista.com/topics/840/smartphones/`. Accessed: 2017-03-20.

[3] R. K. Ganti, F. Ye, and H. Lei. Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine*, 49(11):32–39, 2011.

[4] Prashanth Mohan, Venkata N. Padmanabhan, and Ramachandran Ramjee. Nericell: Rich monitoring of road and traffic conditions using mobile smartphones. In *ACM conf. on Embedded Network Sensor Systems*, pages 323–336, 2008.

[5] Prashanth Mohan, Venkata N. Padmanabhan, Ramachandran Ramjee, and Venkat Padmanabhan. Trafficsense: Rich monitoring of road and traffic conditions using mobile smartphones. Technical report, 2008.

[6] Arvind Thiagarajan, Lenin Ravindranath, Katrina LaCurts, Samuel Madden, Hari Balakrishnan, Sivan Toledo, and Jakob Eriksson. Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 85–98, 2009.

[7] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. In *WSW: Mobile Device Centric Sensor Networks and Applications*, pages 117–134, 2006.

[8] Emiliano Miluzzo, Nicholas D Lane, Shane B Eisenman, and Andrew T Campbell. Cenceme–injecting sensing presence into social networking applications. In *European Conference on Smart Sensing and Context*, pages 1–28, 2007.

[9] Rodrigo de Oliveira and Nuria Oliver. Triplebeat: Enhancing exercise performance with persuasion. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 255–264, 2008.

[10] Suhas Mathur, Tong Jin, Nikhil Kasturirangan, Janani Chandrasekaran, Wenzhi Xue, Marco Gruteser, and Wade Trappe. Parknet: Drive-by sensing of road-side parking statistics. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, pages 123–136, 2010.

[11] Rajib Kumar Rana, Chun Tung Chou, Salil S. Kanhere, Nirupama Bulusu, and Wen Hu. Ear-phone: An end-to-end participatory urban noise mapping system. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pages 105–116, 2010.

[12] Nicolas Maisonneuve, Matthias Stevens, Maria E Niessen, and Luc Steels. Noisetube: Measuring and mapping noise pollution with mobile phones. In *Information technologies in environmental engineering*, pages 215–228. 2009.

[13] Min Mun, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen, Eric Howard, Ruth West, and Péter Boda. PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In *Mobile Systems, Applications and Services*, pages 55–68, 2009.

[14] Prabal Dutta, Paul M. Aoki, Neil Kumar, Alan Mainwaring, Chris Myers, Wesley Willett, and Allison Woodruff. Common sense: Participatory urban sensing using a network of handheld air quality monitors. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 349–350, 2009.

[15] Shigeya Morishita, Shogo Maenaka, Daichi Nagata, Morihiko Tamai, Keiichi Yasumoto, Toshinobu Fukukura, and Keita Sato. Sakurasensor: Quasi-realtime cherry-lined roads detection through participatory video sensing by cars. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 695–705, 2015.

[16] R. Ben Messaoud and Y. Ghamri-Doudane. QoI and energy-aware mobile sensing scheme: A tabu-search approach. In *IEEE 82nd VTC Fall*, pages 1–6, 2015.

[17] R. Ben Messaoud and Y. Ghamri-Doudane. Fair QoI and energy-aware task allocation in participatory sensing. In *IEEE WCNC*, pages 1–6, 2016.

[18] Rim Ben Messaoud, Yacine Ghamri-Doudane, and Dmitri Botvich. Preference and Mobility-Aware Task Assignment in Participatory Sensing. In *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 93–101, 2016.

[19] R. Ben Messaoud, N. Sghaier, M. A. Moussa, and Y. Ghamri-Doudane. On the privacy-utility tradeoff in participatory sensing systems. In *2016 IEEE 15th International Symposium on Network Computing and Applications (NCA)*, pages 294–301, 2016.

[20] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1797–1806, 2008.

[21] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad. Mobile phone sensing systems: A survey. *IEEE Communications Surveys Tutorials*, 15(1):402–427, 2013.

[22] Shane B. Eisenman, Emiliano Miluzzo, Nicholas D. Lane, Ronald A. Peterson, Gahng-Seop Ahn, and Andrew T. Campbell. Bikenet: A mobile sensing system for cyclist experience mapping. *ACM Trans. Sen. Netw.*, 6(1):1–39, 2010.

[23] Sasank Reddy, Andrew Parker, Josh Hyman, Jeff Burke, Deborah Estrin, and Mark Hansen. Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype. In *Proceedings of the 4th workshop on Embedded networked sensors*, pages 13–17, 2007.

[24] Moo-Ryong Ra, Bin Liu, Tom F. La Porta, and Ramesh Govindan. Medusa: A programming framework for crowd-sensing applications. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, pages 337–350, 2012.

[25] 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*, 51(6):112–119, 2013.

[26] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung. Vita: A crowdsensing-oriented mobile cyber-physical system. *IEEE Transactions on Emerging Topics in Computing*, 1(1):148–165, 2013.

[27] Cory Cornelius, Apu Kapadia, David Kotz, Dan Peebles, Minho Shin, and Nikos Triandopoulos. AnonySense: Privacy-aware people-centric sensing. In *Mobile Systems, Applications and Services*, pages 211–224, 2008.

[28] Tathagata Das, Prashanth Mohan, Venkata N. Padmanabhan, Ramachandran Ramjee, and Asankhaya Sharma. Prism: Platform for remote sensing using smartphones. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, pages 63–76, 2010.

[29] Rim Ben Messaoud, Zeineb Rejiba, and Yacine Ghamri-Doudane. An energy-aware end-to-end crowdsensing platform: Sensarena. In *Consumer Communications & Networking Conference (CCNC), 2016 13th IEEE Annual*, pages 284–285, 2016.

[30] S. S. Kanhere. Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. In *2011 IEEE 12th International Conference on Mobile Data Management*, pages 3–6, 2011.

[31] Fehmi Ben Abdesslem, Andrew Phillips, and Tristan Henderson. Less is more: Energy-efficient mobile sensing with senseless. In *Proceedings of the 1st ACM Workshop on Networking, Systems, and Applications for Mobile Handhelds*, pages 61–62, 2009.

[32] B. Priyantha, D. Lymberopoulos, and J. Liu. Littlerock: Enabling energy-efficient continuous sensing on mobile phones. *IEEE Pervasive Computing*, 10(2):12–15, 2011.

[33] I. Constandache, S. Gaonkar, M. Sayler, R. R. Choudhury, and L. Cox. Enloc: Energy-efficient localization for mobile phones. In *IEEE INFOCOM 2009*, pages 2716–2720, 2009.

[34] Mikkel Baun Kjærgaard, Jakob Langdal, Torben Godsk, and Thomas Toftkjær. Entracked: Energy-efficient robust position tracking for mobile devices. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, pages 221–234, 2009.

[35] Mikkel Baun Kjærgaard, Sourav Bhattacharya, Henrik Blunck, and Petteri Nurmi. Energy-efficient trajectory tracking for mobile devices. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, pages 307–320, 2011.

[36] Byung-Gon Chun, Sunghwan Ihm, Petros Maniatis, Mayur Naik, and Ashwin Patti. Clonecloud: elastic execution between mobile device and cloud. In *Proceedings of the sixth conference on Computer systems*, pages 301–314, 2011.

[37] S. Kosta, A. Aucinas, Pan Hui, R. Mortier, and Xinwen Zhang. Thinkair: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading. In *2012 Proceedings IEEE INFOCOM*, pages 945–953, 2012.

[38] Huadong Ma, Dong Zhao, and Peiyan Yuan. Opportunities in mobile crowd sensing. *IEEE Communications Magazine*, 52(8):29–35, 2014.

[39] Jukka K. Nurminen. Parallel connections and their effect on the battery consumption of a mobile phone. In *Proceedings of the 7th IEEE Conference on Consumer Communications and Networking Conference*, pages 385–389, 2010.

[40] Nicholas D. Lane, Yohan Chon, Lin Zhou, Yongzhe Zhang, Fan Li, Dongwon Kim, Guanzhong Ding, Feng Zhao, and Hojung Cha. Piggyback crowdsensing (pcs): Energy efficient crowdsourcing of mobile sensor data by exploiting smartphone app opportunities. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, pages 1–14, 2013.

[41] G. P. Perrucci, F. H. P. Fitzek, and J. Widmer. Survey on energy consumption entities on the smartphone platform. In *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pages 1–6, 2011.

[42] S. Nirjon, A. Nicoara, C. H. Hsu, J. Singh, and J. Stankovic. Multinets: Policy oriented real-time switching of wireless interfaces on mobile devices. In *2012 IEEE 18th Real Time and Embedded Technology and Applications Symposium*, pages 251–260, 2012.

[43] H. Liu, S. Hu, W. Zheng, Z. Xie, S. Wang, P. Hui, and T. Abdelzaher. Efficient 3g budget utilization in mobile participatory sensing applications. In *2013 Proceedings IEEE INFOCOM*, pages 1411–1419, 2013.

[44] X. Sun, S. Hu, L. Su, T. F. Abdelzaher, P. Hui, W. Zheng, H. Liu, and J. A. Stankovic. Participatory sensing meets opportunistic sharing: Automatic phone-to-phone communication in vehicles. *IEEE Transactions on Mobile Computing*, 15(10):2550–2563, 2016.

[45] L. Wang, D. Zhang, H. Xiong, J. P. Gibson, C. Chen, and B. Xie. ecosense: Minimize participants' total 3g data cost in mobile crowdsensing using opportunistic relays. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, PP(99):1–14, 2016.

[46] 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*, 18(1):54–67, 2016.

[47] L. G. Jaimes, I. J. Vergara-Laurens, and A. Raij. A survey of incentive techniques for mobile crowd sensing. *IEEE Internet of Things Journal*, 2(5):370–380, 2015.

[48] Louise Barkhuus, Matthew Chalmers, Paul Tennent, Malcolm Hall, Marek Bell, Scott Sherwood, and Barry Brown. Picking pockets on the lawn: the development of tactics and strategies in a mobile game. In *International Conference on Ubiquitous Computing*, pages 358–374, 2005.

[49] Identification of structural landmarks in a park using movement data collected in a location-based game. In *Proceedings of The First ACM SIGSPATIAL International Workshop on Computational Models of Place*, pages 1–8, 2013.

[50] Mingzhong Wang. Followme if you can: A study of mobile crowd sensing with pokÉmon go. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 39:1–39:9, 2017.

[51] B. Hoh, T. Yan, D. Ganesan, K. Tracton, T. Iwuchukwu, and J. S. Lee. Trucentive: A game-theoretic incentive platform for trustworthy mobile crowdsourcing parking services. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 160–166, 2012.

[52] Tie Luo and Chen-Khong Tham. Fairness and social welfare in incentivizing participatory sensing. In *IEEE Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pages 425–433, 2012.

[53] Linda Deng and Landon P. Cox. Livecompare: Grocery bargain hunting through participatory sensing. In *Proceedings of the 10th Workshop on Mobile Computing Systems and Applications*, HotMobile '09, pages 4:1–4:6, 2009.

[54] Delphine Christin, Andreas Reinhardt, Salil S. Kanhere, and Matthias Hollick. A survey on privacy in mobile participatory sensing applications. *J. Syst. Softw.*, 84, 2011.

[55] C. Bisdikian, J. Branch, K.K. Leung, and R.I. Young. A letter soup for the quality of information in sensor networks. In *Pervasive Computing and Communications (PerCom)*, pages 1–6, 2009.

[56] Vinay Sachidananda, Abdelmajid Khelil, and Neeraj Suri. Quality of information in wireless sensor networks: A survey.

[57] Zheng Song, C.H. Liu, Jie Wu, Jian Ma, and Wendong Wang. QoI-aware multitask-oriented dynamic participant selection with budget constraints. *IEEE Transactions on Vehicular Technology*, 63(9):4618–4632, 2014.

[58] Xiang Sheng, Jian Tang, and Weiyi Zhang. Energy-efficient collaborative sensing with mobile phones. In *INFOCOM*, pages 1916–1924, 2012.

[59] Xiang Sheng, Jian Tang, Xuejie Xiao, and Guoliang Xue. Leveraging GPS-less sensing scheduling for green mobile crowd sensing. *IEEE Internet of Things Journal*, 1(4):328–336, 2014.

[60] H. Xiong, D. Zhang, L. Wang, and H. Chaouchi. Emc3: Energy-efficient data transfer in mobile crowdsensing under full coverage constraint. *IEEE Transactions on Mobile Computing*, 14(7):1355–1368, 2015.

[61] Daqing Zhang, Haoyi Xiong, Leye Wang, and Guanling Chen. Crowdrecruiter: Selecting participants for piggyback crowdsensing under probabilistic coverage constraint. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 703–714, 2014.

[62] Xiaoxin Yin, Jiawei Han, and Philip S. Yu. Truth discovery with multiple conflicting information providers on the web. *IEEE Trans. on Knowl. and Data Eng.*, 20(6):796–808, 2008.

[63] Bin Liu, Yurong Jiang, Fei Sha, and Ramesh Govindan. Cloud-enabled privacy-preserving collaborative learning for mobile sensing. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 57–70, 2012.

[64] Xianling Lu, Deying Li, Biaofei Xu, Wenping Chen, and Zhiming Ding. Minimum cost collaborative sensing network with mobile phones. In *International Conference on Communications (ICC)*, pages 1816–1820, 2013.

[65] Dan Peng, Fan Wu, and Guihai Chen. Pay as how well you do: A quality based incentive mechanism for crowdsensing. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 177–186, 2015.

[66] L. Pu, X. Chen, J. Xu, and X. Fu. Crowdlet: Optimal worker recruitment for self-organized mobile crowdsourcing. In *Proc. IEEE INFOCOM*, 2016.

[67] H. Weinschrott, F. Durr, and K. Rothermel. Streamshaper: Coordination algorithms for participatory mobile urban sensing. In *Mobile Adhoc and Sensor Systems (MASS)*, pages 195–204, 2010.

[68] Qingwen Zhao, Yanmin Zhu, Hongzi Zhu, Jian Cao, Guangtao Xue, and Bo Li. Fair energy-efficient sensing task allocation in participatory sensing with smartphones. In *INFOCOM*, pages 1366–1374, 2014.

[69] J Von Neumann and O Morgenstern. Theory of games and economic behavior. 1944.

[70] Peter C Fishburn. Utility theory for decision making. Technical report, DTIC Document, 1970.

[71] Quoc-Thinh Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine. On utility models for access network selection in wireless heterogeneous networks. In *IEEE Network Operations and Management Symposium, NOMS*, pages 144–151, 2008.

[72] Fred Glover. Tabu search-Part I. *ORSA Journal on Computing*, 1(3):190–206, 1989.

[73] H. Kamal, M. Coupechoux, and P. Godlewski. A tabu search DSA algorithm for reward maximization in cellular networks. In *Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 40–45, 2010.

[74] Huaizhou Shi, R.V. Prasad, E. Onur, and I.G.M.M. Niemegeers. Fairness in wireless networks: Issues, Measures and Challenges. *IEEE Communications Surveys Tutorials*, 16(1):5–24, 2014.

[75] Raj Jain, Dah-Ming Chiu, and W. Hawe. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. *CoRR*, cs.NI/9809099, 1998.

[76] Immanuel Koenig, Abdul Qudoos Memon, and Klaus David. Energy consumption of the sensors of smartphones. In *Int. Symposium on Wireless Communication Systems (ISWCS)*, pages 1–5, 2013.

[77] Layla Pournajaf, Li Xiong, and Vaidy Sunderam. Dynamic data driven crowd sensing task assignment. *Procedia Computer Science*, 29(0):1314 – 1323, 2014.

[78] Miao Lin and Wen-Jing Hsu. Mining {GPS} data for mobility patterns: A survey. *Pervasive and Mobile Computing*, 12:1 – 16, 2014.

[79] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Comput.*, 7(5):275–286, October 2003.

[80] R.Timothy Marler and JasbirS. Arora. The weighted sum method for multi-objective optimization: new insights. *Structural and Multidisciplinary Optimization*, 41(6):853–862, 2010.

[81] Nils Aschenbruck, Raphael Ernst, Elmar Gerhards-Padilla, and Matthias Schwamborn. Bonnmotion: a mobility scenario generation and analysis tool. In *Int. Conf. on Simulation Tools and Techniques (ICST)*, 2010.

[82] Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, pages 107–113, 1993.

[83] Hong Lu, Nicholas D. Lane, Shane B. Eisenman, and Andrew T. Campbell. Bubble-sensing: Binding sensing tasks to the physical world. *Pervasive and Mobile Computing*, 6(1):58 – 71, 2010.

[84] Man Hon Cheung, Richard Southwell, Fen Hou, and Jianwei Huang. Distributed time-sensitive task selection in mobile crowdsensing. In *Proc. ACM MobiHoc*, pages 157–166, 2015.

[85] M. Xiao, J. Wu, L. Huang, Y. Wang, and C. Liu. Multi-task assignment for crowdsensing in mobile social networks. In *Proc. IEEE INFOCOM*, pages 2227–2235, 2015.

[86] J. Wu, M. Xiao, and L. Huang. Homing spread: Community home-based multi-copy routing in mobile social networks. In *Proc. IEEE INFOCOM*, pages 2319–2327, 2013.

[87] Wei Gao, Qinghua Li, Bo Zhao, and Guohong Cao. Multicasting in delay tolerant networks: A social network perspective. In *Proc. ACM MobiHoc*, pages 299–308, 2009.

[88] K. F. Riley. *Mathematical methods for physics and engineering: a comprehensive guide.* Cambridge University Press, 2006.

[89] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Seongjoon Kim, and Song Chong. CRAWDAD dataset ncsu/mobilitymodels (v. 2009-07-23). Downloaded from http://crawdad.org/ncsu/mobilitymodels/20090723/GPS, 2009.

[90] Siamak Faridani, Björn Hartmann, and Panagiotis G. Ipeirotis. What's the right price? pricing tasks for finishing on time. In *Proceedings of the 11th AAAI Conference on Human Computation*, pages 26–31, 2011.

[91] D. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*, pages 105–142, 1974.

[92] Yihan Gao and Aditya Parameswaran. Finish them!: Pricing algorithms for human computation. *Proc. VLDB Endow.*, 7(14):1965–1976, 2014.

[93] Shuang Chen, Min Liu, and Xiao Chen. A truthful double auction for two-sided heterogeneous mobile crowdsensing markets. *Computer Communications*, 81:31 – 42, 2016.

[94] Juong-Sik Lee and Baik Hoh. Dynamic pricing incentive for participatory sensing. *Pervasive and Mobile Computing*, 6(6):693 – 708, 2010.

[95] Haiming Jin, Lu Su, Danyang Chen, Klara Nahrstedt, and Jinhui Xu. Quality of information aware incentive mechanisms for mobile crowd sensing systems. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 167–176, 2015.

[96] Haiming Jin, Lu Su, Houping Xiao, and Klara Nahrstedt. Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems. In *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 341–350, 2016.

[97] Y. Wen, J. Shi, Q. Zhang, X. Tian, Z. Huang, H. Yu, Y. Cheng, and X. Shen. Quality-driven auction-based incentive mechanism for mobile crowd sensing. *IEEE Transactions on Vehicular Technology*, 64(9):4203–4214, 2015.

[98] Dejun Yang, Guoliang Xue, Xi Fang, and Jian Tang. Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing. In *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, pages 173–184, 2012.

[99] D. Yang, G. Xue, X. Fang, and J. Tang. Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones. *IEEE/ACM Transactions on Networking*, 24(3):1732–1744, 2016.

[100] I. Koutsopoulos. Optimal incentive-driven design of participatory sensing systems. In *2013 Proceedings IEEE INFOCOM*, pages 1402–1410, 2013.

[101] EJ Gumbel. Les valeurs extrêmes des distributions statistiques. In *Annales de l'institut Henri Poincaré*, pages 115–158, 1935.

[102] I. J. Vergara-Laurens, L. G. Jaimes, and M. A. Labrador. Privacy-preserving mechanisms for crowdsensing: Survey and research challenges. *IEEE Internet of Things Journal*, 2016.

[103] Marius Wernke, Pavel Skvortsov, Frank Dürr, and Kurt Rothermel. A classification of location privacy attacks and approaches. *Personal Ubiquitous Comput.*, 18:163–175, 2014.

[104] Katie Shilton. Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection. *ACM Communications*, 52:48–53, 2009.

[105] E. De Cristofaro and C. Soriente. Participatory privacy: Enabling privacy in participatory sensing. *IEEE Network*, 27, 2013.

[106] Khuong Vu, Rong Zheng, and Jie Gao. Efficient algorithms for k-anonymous location privacy in participatory sensing. In *2012 Proceedings IEEE INFOCOM*, pages 2399–2407, 2012.

[107] Y. Yao, L. T. Yang, and N. N. Xiong. Anonymity-based privacy-preserving data reporting for participatory sensing. *IEEE Internet of Things Journal*, 2:381–390, 2015.

[108] J. Shi, R. Zhang, Y. Liu, and Y. Zhang. Prisense: Privacy-preserving data aggregation in people-centric urban sensing systems. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, 2010.

[109] Raghu K. Ganti, Nam Pham, Yu-En Tsai, and Tarek F. Abdelzaher. Poolview: Stream privacy for grassroots participatory sensing. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, pages 281–294, 2008.

[110] I. S. Reed. Information theory and privacy in data banks. In *Proceedings of the June 4-8, 1973, National Computer Conference and Exposition*, pages 581–587, 1973.

[111] S. Rajagopalan L. Sankar and H. V. Poor. A theory of privacy and utility in databases. In *ArXiv e-prints, Feb. 2011*, 2011.

[112] F. du Pin Calmon and N. Fawaz. Privacy against statistical inference. In *50th Annual Allerton Conference on Communication, Control, and Computing*, pages 1401–1408, 2012.

[113] S. Salamatian, A. Zhang, F. du Pin Calmon, S. Bhamidipati, N. Fawaz, B. Kveton, P. Oliveira, and N. Taft. Managing your private and public data: Bringing down inference attacks against your privacy. *IEEE Journal of Selected Topics in Signal Processing*, 9:1240–1255, 2015.

[114] I.Wagner and D.Eckhoff. Technical privacy metrics: a systematic survey. In *eprint arXiv:1512.00327*, 2015.

[115] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley-Interscience, 2006.

[116] Arkadi S. Nemirovski and Michael J. Todd. Interior-point methods for optimization. *Acta Numerica*, 17:191–234, 2008.

[117] Roger Fletcher. *The Sequential Quadratic Programming Method*, pages 165–214. Springer Berlin Heidelberg, 2010.

[118] S. A. Kassam and H. V. Poor. Robust techniques for signal processing: A survey. *Proceedings of the IEEE*, 73(3):433–481, 1985.

[119] Luis M. Candanedo and Vronique Feldheim. Accurate occupancy detection of an office room from light, temperature, humidity and {CO2} measurements using statistical learning models. *Energy and Buildings*, 112:28 – 39, 2016.

[120] GO TRACK application. `https://play.google.com/store/apps/gotrackfree`. Accessed: 2016-08-16.

[121] M. O. Cruz, H. Macedo, and A. Guimares. Grouping similar trajectories for carpooling purposes. In *2015 Brazilian Conference on Intelligent Systems (BRACIS)*, pages 234–239, 2015.

[122] Mohannad A. Alswailim, Hossam S. Hassanein, and Mohammad Zulkernine. CRAWDAD dataset queensu crowd temperature (v. 2015-11-20). Downloaded from `http://crawdad.org/queensu/crowd_temperature/20151120`, nov 2015.

Vers une capture participative mobile efficace : affectation des tâches et téléchargement des données

Introduction Générale

L'ubiquité des terminaux intelligents équipés de capteurs a donné naissance à un nouveau paradigme de collecte participative des données appelé *Crowdsensing.* Ce modèle de collecte émergent a attiré beaucoup d'attention tant du monde académique qu'industriel grâce aux avantages significatifs qu'il présente en comparaison avec les réseaux de collecte traditionnels (WSNs). En premier lieu, le Crowdsensing surmonte les limitations de coût des capteurs statiques déployés et collecte des données dans des zones non économiquement faisables auparavant comme dans le cas des applications de contrôle de congestion trafic routier. De plus, ce paradigme élargit la zone de couverture vu la mobilité des participants. En conséquence, diverses applications de collecte, conduites d'une façon opportuniste [1] ou participative [7], ont été développées s'étendant du suivi des activités personnelles [8,9] au contrôle d'infrastructure urbaine [4–6,10] et d'environnement [11–15].

Le Crowdsensing s'appuie sur deux acteurs essentiels : la plateforme de collecte et les utilisateurs équipés de leurs terminaux intelligents. Ces derniers peuvent jouer le rôle de demandeurs de services, i.e., ceux qui soumettent les tâches de collecte, ou des participants, i.e., ceux qui dédient leurs ressources pour effectuer ces tâches. En outre, nous distinguons quatre phases majeures de la collecte participative : la soumission des tâches, l'attribution ou affectation de tâches, la collecte et le traitement des données et enfin le téléchargement. En observant les phases du Crowdsensing, les chercheurs revendiquent qu'un tel paradigme prometteur soulève aussi de nouveaux défis [1] vu sa dépendance à l'égart du facteur humain. En effet, les participants consacrent leurs ressources énergétiques pour exécuter des campagnes de collecte et peuvent être activement impliqués dans certaines tâches comme la prise de photos ou l'enregistrement de vidéos. Par conséquent, ils sont généralement recrutés sur la base de récompenses, définies comme "incitations". Par ailleurs, le téléchargement public de données suscite des préoccupations relatives la protection de la vie privée des utilisateurs en plus de leur réticence à allouer un quota de leur budget de connexions données mobile. En même temps, les demandeurs de service exigent des contributions de qualité mais doivent dédier un budget pour *payer* les participants pour en profiter. Pour mener à bien les tâches de collecte, ces divers défis relatifs à l'implication des participants et des demandeurs de services doivent être levés.

Dans cette thèse, nous proposons d'aborder ces différents défis qui affectent à la fois les demandeurs et les participants. Particulièrement, nous introduisons à travers les trois contributions majeures de cette thèse des protocoles d'affectation de tâches et de téléchargement de données qui visent à organiser les différentes phases du Crowdsensing en abordant les questions suivantes : *Comment affecter les tâches de collecte afin de maximiser la qualité des données d'une façon éco-énergétique? Comment minimiser le temps nécessaire à la* *collecte et au traitement des tâches?* et *Comment protéger la vie privée des participants tout en préservant la qualité des données reportées?*

Contributions

Cette thèse détaille, à travers sept chapitres, nos approches pour apprhender les défis identifiés autour des systèmes de collecte de données en introduisant des protocoles qui organisent les diverses phases du Crowdsensing. Les différentes contributions de ce travail sont détaillées comme suit :

I- QEMSS/F-QEMSS : QoI and Energy-aware Mobile Sensing Scheme

Tout d'abord, nous nous intéressons à la phase d'affectation de tâches et nous étudions le modèle d'allocation centralisé. Plus précisément, nous mettons l'accent sur le fait que d'une part les ressources énergétiques des terminaux mobiles restent limitées. D'autre part, les demandeurs de services requièrent un niveau de qualité de contributions en dessous duquel les données seront perçues comme *inutiles*. Ainsi, nous introduisons des modèles de déploiement de tâches qui visent à maximiser la qualité des données reportées tout en minimisant le coût énergétique global de la collecte. Le modèle d'allocation proposé, QEMSS, définit des métriques de qualité de données et cherche à les maximiser en se basant sur des heuristiques de la recherche taboue.

Pour ce faire, nous examinons la notion de qualité de données désignée aussi par le terme QoI (Quality of Information). Cette caractèristique consiste en plusieurs attributs comme la précison, l'exactitude, la validité et la fiabilité [56]. Dans notre travail, nous considérons trois attributs pour qualifier les données extraites : la *completeness*, la *timeliness* et l'*affordability*. Le premier terme mesure le degr (niveau) avec lequel la donnée extraite donne une vision complète sur la réalité. Nous étudions la *completeness* spatiale en estimant la zone potentiellement couverte par un participant. La *timeliness* exige que la collecte et le partage des données se déroulent dans un intervalle de temps précis. Le dernier critère, l'*affordability*, permet de repérer et d'exclure de la collecte les individus munis de dispositifs de faible batterie. Afin d'estimer la qualité des données collectées que des participants peuvent fournir, nous quantifions les différents attributs mentionnés ci-dessus par la fonction d'utilité de [71].

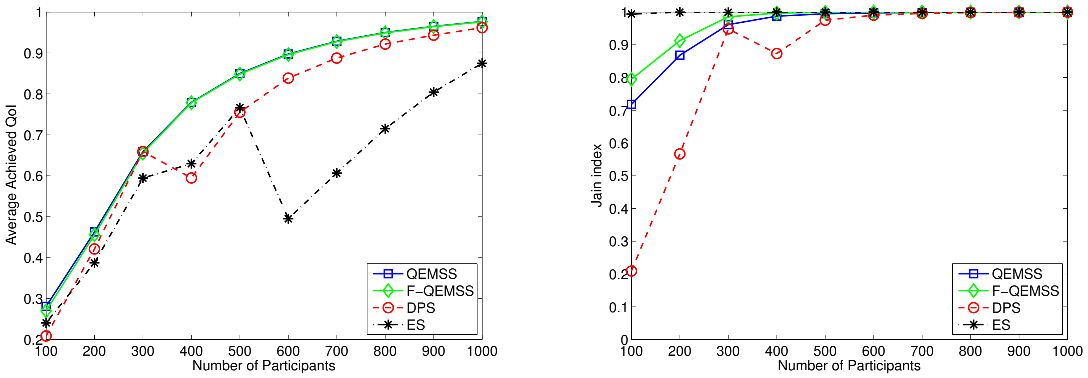
Sur la base de cette modélisation, nous formulons l'objectif de cette première contribution QEMSS qui consiste à maximiser la qualité des données collectées tout en diminuant le coût de leur acquisition en termes de ressources énergétiques. Cette formulation s'est traduite par un problème d'optimisation avec fonction objectif la maximisation du produit des métriques de qualité.

$$\begin{aligned}
&\underset{X}{\text{Maximize}} \quad && \sum_{i=1}^{n_p}\sum_{j=1}^{m} x_{ij} \times U_{QoI}(QoI_{ij}) && \\
&\text{Subject to} \quad && U_e(e_i - e_c\sum_{j=1}^{m} x_{ij}) \geq 0 && \forall p_i \in P \\
&\text{and} \quad && \sum_{i=1}^{n_p} x_{ij} \leq 1 && \forall a_j \in A
\end{aligned} \tag{7.1}$$

La première contrainte exprime pour chaque terminal l'énergie nécessaire à ses activités quotidiennes tandis que la deuxième contrainte interdit la redondance de l'information en limitant le nombre d'utilisateurs sélectionnés par sous-région à 1. Pour résoudre ce type de problème, nous cherchons la “combinaison” des participants qui répond aux contraintes et maximise la fonction objectif. Néanmoins, les problèmes d'optimisation combinatoire sont difficiles voire impossibles à résoudre lorsqu'il s'agit d'un nombre important d'éléments impliqués comme dans notre cadre de campagnes de collecte à grande échelle. En conséquence, il serait plus efficace de faire appel à un algorithme méta-heuristique tel que la recherche taboue (Tabu Search) pour générer une solution sous-optimale. L'algorithme proposé consiste en quatre étapes : initialisation, formation du voisinage, sélection du meilleur voisin et mise à jour de la liste taboue.

- **Initialisation** : cette phase vise à générer une solution initiale qui répond aux contraintes du problème d'optimisation. La solution X est une matrice dont les lignes sont les participants et les colonnes reprsentent les rgions de collecte, où chaque élément $x_{i,j}$ est mis à 1 si et seulement si le participant p_i est sélectionné pour la collecte dans la zone a_j.
- **Formation du voisinage** : à partir d'une solution X, nous pouvons générer tout son voisinage $N(X)$ en appliquant un seul mouvement m. Nous considérons comme mouvement l'échange d'affectation d'une tâche de collecte dans une zone déterminée entre deux utilisateurs présents.
- **Sélection de la solution** : pour chaque solution générée dans $N(X)$, nous calculons la valeur de la fonction objectif pour choisir la solution avec la valeur maximale.
- **Mise à jour de la liste taboue (TL)** : nous ajoutons les attributs des solutions visitées à la TL pour éviter de retourner à ce voisinage de solutions.

Nous itérons les étapes précédentes jusqu'à affectation complète des tâches ou expiration du nombre d'itérations maximum de l'algorithme QEMSS basé sur la recherche taboue. La solution générée est ainsi l'affectation adéquate qui assure maximum de QoI.



(a) Le taux moyen de QoI

(b) Le taux d'équité mesuré par l'indice de Jain

Figure 7.1: Évaluation de QEMSS et F-QEMSS en termes de QoI et d'équité

Par ailleurs, nous évaluons l'équité de l'allocation résultante en faisant appel à un deuxième algorithme, F-QEMSS. Le but de ce modèle est de trouver le sous-ensemble de participants qui réalise un maximum de qualité des données collectées et d'équité tout en minimisant la consommation des ressources dédiées. Nous formulons cette vision en un problème multi-objectif et nous procédons à sa résolution par la méthodologie de la recherche taboue tout en respectant la nouvelle métrique d'efficacité système.

Les deux solutions proposées ont été évaluées par rapport aux méthodes d'affectation de l'état de l'art. Les résultats des simulations démontrent que nos modèles QEMSS et F-QEMSS peuvent atteindre une performance équivalente en termes de qualité de données dans les régions de forte densité de population. Toutefois, nous réalisons un gain significatif en termes de QoI et de précision spatiale et temporelle notamment dans les scénarios difficiles tels que les zones à faible densité ou encore les participants munis de terminaux à faibles ressources énergétiques. De plus, F-QEMSS assure un niveau important d'équité entre les participants tout en préservant le même niveau de qualité de données que QEMSS. Par conséquent, nos solutions d'affectation de tâches réalisent un compromis de satisfaction entre les demandeurs et les participants lors de la collecte participative; ce qui assure leur engagement dans ce type de campagnes.

II- P-MATA/IP-MATA^{+} : Preferences and Mobility-aware Task Assignment

En seconde phase, nous considérons les limitations du modèle centralisé en termes de gestion globale de la mobilité des participants dans le cadre du Crowdsensing à grande échelle. Nous proposons d'assister l'affectation des tâches proposée dans la première contribution par une phase d'allocation locale distribuée. L'idée est de désigner des participants parmi ceux sélectionnés et leur attribuer des tâches supplémentaires à déléguer en rencontrant d'autres utilisateurs. La finalité de cette deuxième phase d'allocation est de minimiser le temps moyen de collecte et de traitement des données.

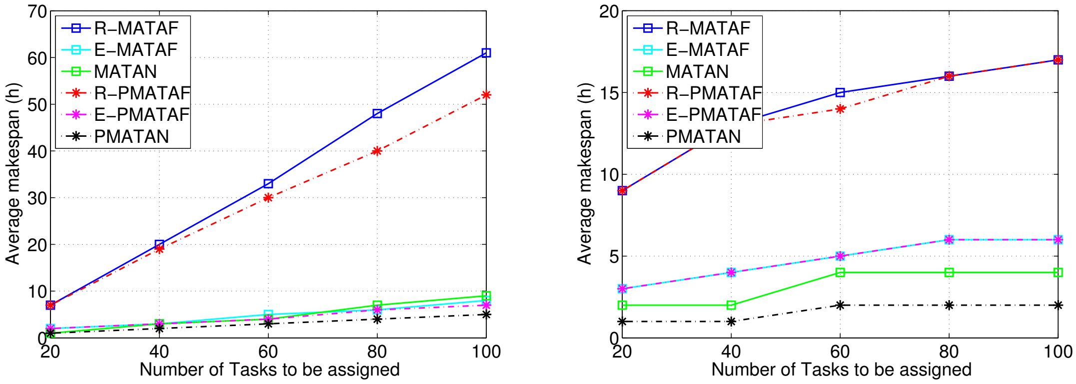
Nous avons détaillé cette deuxième contribution dans deux chapitres vu que les modèles d'affectation proposés peuvent être classés en deux catégories différentes. La première catégorie est un ensemble de méthodes (MATA et P-MATA) qui modélisent la mobilité des utilisateurs et leurs préférences de collecte en termes d'acceptation ou de rejet selon l'historique des rencontres et d'affectations précédentes. La deuxième catégorie introduit un modèle de préférences plus sophistiqué en faisant appel au modèle de choix de la régression logit. Particulièrement, nous développons deux autres méthodes d'affectation distribuée; P-MATA$^+$ et IP-MATA$^+$ qui intègrent le nouveau modèle de choix et étudient l'engagement des participants aux campagnes de collecte en fonction des différents attributs comme: le nombre de tâches, la durée de collecte et les incitations associées.

Nous débutons notre étude d'affectation distribuée par la modélisation de la mobilité locale des utilisateurs dans les régions de collecte. Nous nous basons sur les modèles de mobilité des réseaux sociaux [66, 85] reposant sur le processus de Poisson. En conséquent, nous estimons le temps de rencontre entre un participant responsable de déléguation de tâches r_i et un participant ordinaire p_j, par une distribution exponentielle de paramètre $F_{(k,p_j)} = q_k(r_i)q_k(p_j)\lambda_{p_j}$, où q_k est la probabilité qu'un utilisateur soit dans une sous région de collecte k et λ_{p_j} est le taux de rencontre déduit des échanges précédents. Ensuite, nous élaborons le temps moyen de collecte et de traitement des tâches, AM, en fonction du temps de rencontre plus la somme des durées des tâches :

$$AM = \frac{1}{m}\sum_{j=1}^{n}\sum_{l\in\gamma_j}\left(\frac{1}{q_k(r_i)q_k(p_j)\lambda_{p_j}} + \tau_l\right)$$

Afin de minimiser ce terme, nous dévélopons un premier algorithme MATA en considérant seulement la mobilité des participants et leur temps de rencontre pour désigner celui à qui allouer les tâches. Cette affectation est introduite de deux façons : hors ligne et en ligne. La première méthode, MATAF, requiert qu'un participant responsable estime le temps de rencontre et par conséquent de collecte à l'avance et associe à chaque participant sélectionné sa liste de tâches à déléguer au moment de sa rencontre. En revanche, la deuxième méthode d'affectation, MATAN, ne se déclenche qu'en rencontrant un participant et adopte le même principe d'affectation que MATAF qui consiste à associer les tâches avec les durées les plus courtes aux participants avec les plus proches instants de rencontre.

La solution MATA ne considère que la mobilité des utilisateurs. Ainsi, pour éviter le risque de rejet de tâches par les participants après la phase d'affectation, nous introduisons la propriété de préférences des participants durant la collecte. Cela a été dévéloppé en premier lieu sur la base de l'historique des affectations en modélisant la probabilité d'acceptation d'une tâche de collecte p_a par une variable de Bernoulli. Par conséquent, le temps moyen avant de rencontrer un participant "positif", i.e., qui acceptera la tâche



(a) KAIST (b) NCSU

Figure 7.2: Le temps moyen de collecte réalisé par MATA Vs P-MATA

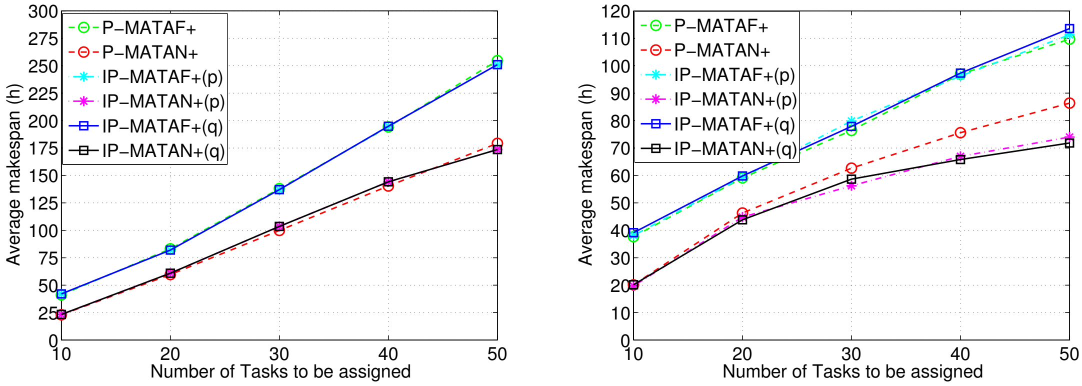
proposée, peut se reformuler comme suit :

$$\Pi_{i,j,k} = \Big[\sum_{l=1}^{n} p_r^l p_a \frac{1}{q_k(r_i)q_k(p_j)\lambda_{p_j}}\Big]'$$

Nous intégrons ce nouveau terme dans l'expression du temps moyen de collecte et nous proposons des solutions conjointement conscientes de la mobilité des participants et de leurs préférences, P-MATA. Ce type d'affectation a été développé comme son prédécesseur MATA sur deux modèles; hors ligne et en ligne. Les deux solutions présentées dans la première catégorie d'affectation distribuée ont été évaluées par simulation. La solution P-MATAN, qui considère la mobilité des participants et leurs préférences en ligne est la plus prometteuse en termes de minimisation du temps moyen de collecte avec zéro tâche rejetée comme l'illustre la figure 7.2.

En conséquence, nous proposons d'étendre la solution P-MATA pour intégrer un modèle de préférences plus élaboré. Nous faisons appel à la régression logit pour formuler la probabilité d'acceptation en tant que fonction des attributs des tâches ; $p_a = \frac{\exp(\beta z_j)}{\sum_{i=1}^{M} \exp(\beta z_i)}$. Puis, nous exploitons ce modèle pour reformuler le temps de collecte moyen défini ci-dessus et concevoir les algorithmes nécessaires pour le minimiser. La solution proposée est nommée P-MATA$^+$. Cette méthode se base sur le fait que les participants sont volontaires pour effectuer les campagnes de collecte sans incitations. Ainsi, la probabilité d'acceptation dépend essentiellement du nombre de tâches proposé et accepté ainsi que de la durée totale de la collecte. Nous établissons l'expression de AM et proposons des algorithmes hors ligne et en ligne basés sur la recherche glouton pour déterminer les participants adéquats.

De plus, nous examinons le cas où les participants exigent des récompenses en contrepartie de leur collecte. La solution correspondante est définie par IP-MATA$^+$. Cette méthode étend la formulation de la régression logit pour inclure l'attribut "reward" (récompense) et estime sur cette base le temps nécessaire pour rencontrer les bons participants qui minimisent le temps moyen de la collecte; lequel est le but de cette affectation distribuée. Dans



(a) KAIST

(b) NCSU

Figure 7.3: Le temps moyen de collecte réalisé par P-MATA$^+$ Vs IP-MATA$^+$

ce contexte, nous étudions aussi la politique de partage de récompenses entre les participants. Nous proposons de calculer le *reward* en fonction du type de la collecte (application) ou en fonction de la qualité de la contribution. La première politique d'incitation vise à établir des priorités selon l'importance et l'échéance de la collecte. Cependant, *payer* les participants selon la qualité des données remontées les encourage à bien effectuer les tâches et à télécharger les données à termes. IP-MATA$^+$ est introduite en deux algorithmes hors ligne et en ligne et selon les deux politiques de payement proposées.

Les solutions de la deuxième catégorie d'affectation distribuée sont évaluées sur les mêmes bases que MATA et P-MATA pour mieux distinguer l'impact du modèle de préférences et d'incitations. Comme l'illustre la figure 7.3, IP-MATA$^+$ a favorisé l'engagement des utilisateurs des terminaux intelligents aux campagnes de collecte par l'introduction des récompenses. Notamment, la méthode a minimisé le temps moyen de collecte et a bien exploité le budget dédié lorsqu'elle est associée à la politique de partage en fonction de la qualité des contributions.

III- PRUM : PRivacy-preserving and Utility-aware Mechanism

Notre troisième contribution se positionne dans la phase de téléchargement des données. Plus précisément, nous nous sommes intéressés aux préoccupations de la vie privée des utilisateurs durant la collecte participative. Indéniablement, le téléchargement des données à une échelle communautaire peut exposer les participants la divulgation de leurs informations sensibles. En effet, les applications participatives exigent généralement l'envoi des données associées à la position géographique des mesures pour des raisons de précision. De plus, l'envoi périodique de certain type de données peut causer la corrélation avec des informations sensibles des utilisateurs. Par exemple, l'envoi des rapports de la consommation énergétique des équipements domestiques intelligents peut exposer les activitées *indoor* d'un foyer au prestataire de services. En conséquence, les utilisateurs peuvent être réticents à participer aux campagnes de collecte à moins qu'une garantie de protection de vie privée soit assurée.

Pour répondre à de telles préoccupations, une variété de mécanismes [27, 53, 103–109] préservant la vie privée des participants a été proposée dans la littérature. Cependant, la majorité de ces travaux s'est intéressée à la modification ou au cryptage des données remontées aux plateformes de collecte sans considérer l'impact de telles techniques sur l'utilité et la qualité des données, lequel est un critère majeur pour les systèmes de collecte. En effet, l'utilité des données est nécessaire pour les demandeurs afin d'offrir des services qui correspondent aux besoins des utilisateurs. Pour répondre à ces deux objectifs conflictuels, nous proposons de développer un modèle préservant la vie privée des participants tout en conservant les exigences d'utilité des données des demandeurs.

Pour ce faire, nous proposons d'introduire une entité de confiance (broker) dans l'architecture centralisée des systèmes de collecte. Cette entité pourrait être fusionnée avec le serveur central ou se présenter comme entité intermédiaire entre la plateforme de collecte et les utilisateurs. De plus, nous proposons d'exécuter dans ce broker, un mécanisme d'offuscation général qui vise à réaliser un compromis entre la protection de la vie privée et l'utilité des données. Le principe de ce mécanisme est d'offusquer les données reportées par les participants avant d'être envoyées aux demandeurs de service pour minimiser le risque de divulgation d'informations privées sans dégrader la qualité finale des données générées.

Sur cette base, nous abordons notre modélisation de mécanisme de protection par l'étude des modèles d'attaques possibles dans un système de collecte participative. Nous considérons des nœuds malicieux (adversaires) internes et externes. D'abord, nous supposons que le serveur (broker) est la seule entité éprouvée. En revanche, les utilisateurs sont honnêtes mais curieux de collecter des informations sur d'autres participants. Enfin, nous considérons que des adversaires externes peuvent se faire passer pour de faux demandeurs pour infiltrer le téléchargement des données et divulguer des informations sensibles des utilisateurs. Selon ces hypothèses, nous suggérons l'étude de deux modèles d'attaque ; les adversaires sans information adjacente et les adversaires avec information adjacente. Le premier type d'attaque suppose que l'adversaire n'a aucune information sur les participants et essaie d'intercepter les données en cours de téléchargement tandis que le deuxième type a déjà des informations a priori sur les participants qu'il corrèle avec les données remontées.

En outre, nous distinguons deux scénarios de téléchargement de données dans les systèmes de collecte participative. Le premier scénario consiste en une mesure M corrélée avec une information sensible (secret) S. Ainsi, le mécanisme de protection devrait offusquer M seulement. Le deuxième cas de téléchargement est celui des applications participatives qui reportent des mesures M conjointement avec une information secrète S comme dans le cas des mesures géo-localisées. Dans ce cas, les deux données S et M doivent être altérérées avant d'être communiquées aux demandeurs.

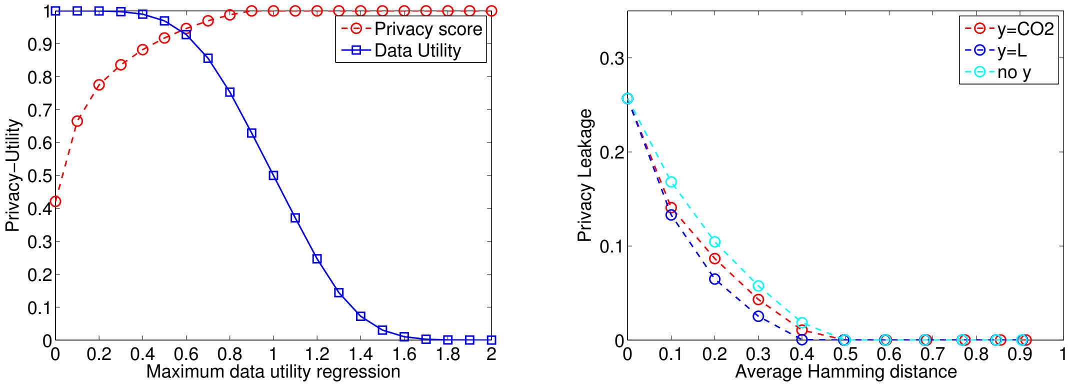
Nous considérons les différents scénarios et modèles d'attaques listés ci-dessus et nous abordons la modélisation de notre objectif. En effet, notre but est de minimiser le risque de divulgation de la vie privée des participants en respectant la contrainte d'utilité des données. Nous quantifions d'abord le risque identifié par l'information mutuelle entre une information secrète S et une mesure M. Aussi, évaluons-nous l'utilité d'une mesure altérée D par la fonction d'utilité détaillée dans le troisième chapitre. De plus, nous proposons d'adopter une approche probabiliste qui modélise le modèle de protection en tant que probabilité de *mapping* entre une mesure M et une donnée altérée D. Par conséquent, nous formulons le problème d'optimisation suivant pour illustrer la finalité de cette troisième contribution.

$$\begin{aligned} \underset{P_{D|M}}{\text{Minimize:}} \quad & I(S, D) \\ \text{subject to:} \quad & \mathbb{E}_{M,D}[d(U(M), U(D))] \leq \delta \end{aligned} \tag{7.2}$$

Cette formulation est convexe et apporte une certaine approximation aux problèmes d'optimisation débit-distorsion. La résolution de ce type de problème peut se faire avec des solveurs convexes comme l'algorithme des points intérieurs. En se basant sur cette méthode, nous dévelopons notre mécanisme de protection PRUM qui vise à générer une version altérée des données collectées minimisant la fonction de l'information mutuelle et répondant à la condition du niveau seuil de régression d'utilité. La solution de ce problème est un *mapping* probabiliste qui définit le pourcentage des données à altérer.

Pour évaluer la performance de PRUM, nous nous basons sur des données d'applications participatives. Au début, nous recherchons la région de compromis entre la vie privée et l'utilité des données. Ensuite, nous considérons le modèle d'attaque avec des informations adjacentes et nous étudions l'impact de cette condition sur les résultats préliminaires. La figure 7.4 illustre les résultats obtenus pour l'application de PRUM sur des données collectées par une application de gestion d'équipements domestiques intelligents. Les mesures reportées sont la température, le niveau d'humidité, de CO2 et la luminosité. L'information considérée comme secrète est la position de l'utilisateur qui peut être déduite de la corrélation des diffŕentes mesures reportées.

Nous observons en premier lieu que le compromis entre l'utilité des données et la protection de la vie privée est réalisable. Naturellement, l'utilité des données décroit en fonction de l'altération appliquée. Cependant, PRUM a réalisé un taux de protection de vie privée égal à 90% en dégradant seulement 10% de la qualité souhaitée. D'autre part, nous étudions l'impact de l'information adjacente sur cette région de compromis, en particulier sur le niveau de protection de la vie privée. Cet impact dépend du type d'information adjacente détenue par l'attaquant. Par exemple, dans le cas où l'information adjacente est la luminosité, un niveau de distorsion faible suffit pour minimiser le risque de divulgation de S en observant M, étant donné que M n'est pas très corrélée à S. En d'autres termes, si l'information adjacente est fortement corrélée au secret S, PRUM minimise la distorsion appliquée.



(a) Compromis Privacy-Utility

(b) l'Informtaion mutuelle avec diffŕentes information adjacentes

Figure 7.4: Évaluation de PRUM sur des données de capteurs domestiques

Conclusion Générale

Le Crowdsensing est envisagé comme paradigme efficace de collecte centrée sur les utilisateurs mobiles qui peut être un socle pour diverses applications comme les mesures environnementales, la gestion du trafic routier et le contrôle des activitées. Néanmoins, la dépendance au facteur humain nécessite une étude approfondie de nouveaux défis classés en préoccupations des demandeurs et préoccupations des participants. En conséquence, nous avons développé dans ce travail des méthodes d'organisation appropriées qui répondent à la question générale de *Comment concevoir des phases de Crowdsensing efficaces respectant les préoccupations des participants et les exigences des demandeurs?*

Pour répondre à cette question, nous avons introduit trois contributions majeures assistant les trois phases d'affectation de tâches, de collecte et de téléchargement des données. Initialement, nous avons abordé la phase d'affectation de tâches avec une approche centralisée qui incite la collaboration parmi les utilisateurs dans le but de minimiser le coût énergétique de la collecte participative et fournir des contributions de meilleure qualit de donnes. Les méthodes proposées (QEMSS et F-QEMSS) ont réduit la consommation d'énergie globale et individuelle des dispositifs participants tout en réalisant des niveaux de qualité de données importants par rapport aux méthodes de l'état de l'art. Notre deuxième contribution a pour rôle d'assister les méthodes d'attribution d'une façon distribuée. Spécifiquement, nous nous sommes concentrés sur la minimisation du temps nécessaire pour la collecte et le traitement des tâches. Le principe de cette proposition est d'étudier la mobilité des utilisateurs et leurs préférences en termes d'acceptation de tâches en présence ou non de récompenses pour sélectionner les participants qui minimisent le temps de collecte. Les mécanismes introduits ont atteint leur but principal. Particulièrement, les méthodes fondées sur les incitations sont les plus prometteuses lorsqu'elles sont associées à des politiques de récompenses en fonction de la qualité des contributions.

Finalement, nous avons étudié les préoccupations de protection de la vie privée des participants lors de téléchargement des données. Nous avons développé une méthode qui vise à réaliser un compromis de la divulgation des informations privées et le niveau de qualité des données défini comme l'utilité.

En résumé, les contributions clés présentées dans cette thèse ont visé différents objectifs d'optimisation différents des systèmes de collecte participative. En effet, nous avons développé des solutions traitant simultanèment des préoccupations de demandeurs et des participants. Ce travail nous a ouvert également la voie à des études plus approfondies pour faciliter la mise en œuvre du Crowdsensing.