Resource and crowd management in mobile networks

Sahar Hoteit

UNIVERSITÉ PIERRE ET MARIE CURIE
Laboratoire d'Informatique de Paris 6

RESOURCE AND CROWD MANAGEMENT IN MOBILE NETWORKS

Doctoral Dissertation of:
Sahar Hoteit

Advisors:
Prof. Guy Pujolle
Dr. Stefano Secci

2014

Thèse

Présentée pour obtenir le grade de docteur
de l'Université Pierre et Marie Curie
Spécialité: Informatique

Sahar HOTEIT

Gestion de ressources et de congestion dans les réseaux mobiles

Soutenue le 19 Septembre 2014 devant le jury composé de:

Rapporteurs	**Prof. Bijan JABBARI**	**George Mason University, Etats-Unis.**
	Prof. Deep MEDHI	**University of Missouri-Kansas City, Etats-Unis.**
Examinateurs	**Dr. Marceau COUPECHOUX**	**Télécom ParisTech, France.**
	Dr. Marcelo DIAS DE AMORIM	**Université Pierre et Marie Curie, France.**
	Dr. Mahmoud DOUGHAN	**Université Libanaise, Liban.**
	Dr. Walid SAAD	**Virginia Tech, Etats-Unis.**
Président	**Dr. Marcelo DIAS DE AMORIM**	**Université Pierre et Marie Curie, France.**
Directeurs de thèse	**Prof. Guy PUJOLLE**	**Université Pierre et Marie Curie, France.**
	Dr. Stefano SECCI	**Université Pierre et Marie Curie, France.**

RESOURCE AND CROWD MANAGEMENT IN MOBILE NETWORKS

Author: Sahar HOTEIT

Defended on September 19, 2014, in front of the committee composed of:

Referees: Prof. Bijan JABBARI (George Mason University, USA).
Prof. Deep MEDHI (University of Missouri-Kansas City, USA).

Examiners: Dr. Marceau COUPECHOUX (Télécom ParisTech, France).
Dr. Marcelo DIAS DE AMORIM (Université Pierre et Marie Curie, France).
Dr. Mahmoud DOUGHAN (Lebanese University, Lebanon).
Dr. Walid SAAD (Virginia Tech, USA).

President: Dr. Marcelo DIAS DE AMORIM (Université Pierre et Marie Curie, France).

Advisors: Prof. Guy PUJOLLE (Université Pierre et Marie Curie, France).
Dr. Stefano SECCI (Université Pierre et Marie Curie, France).

Acknowledgements

First and foremost I would like to express my deepest gratitude to my advisors, Dr. Stefano Secci and Prof. Guy Pujolle for their excellent guidance, patience, motivation and immense knowledge.

I would like also to thank all the jury members for reading my thesis, for offering their valuable time and for their constructive feedbacks.

I must acknowledge as well my friends and colleagues in the Phare team who have been so supportive along the way of doing my thesis.

I would especially like to thank my amazing family for the love, support, and constant encouragement I have gotten over the years. In particular, I would like to thank my parents, my brother, and my two sisters. I undoubtedly could not have done this without you.

Finally, I would like to thank my husband who has always been a constant source of love, support and encouragement. I am truly thankful for having you in my life!

Abstract

The dissertation presents different solutions to cater the increasing demand of mobile data traffic in access and backhauling networks, taking into account real usage profiling and human mobility patterns.

Using real mobile phone data, we study human trajectories and behaviors. We show that these trajectories can be estimated at good accuracy levels, using specific mobility metrics, namely the radius of gyration and the centroid. Then, motivated by the fact that crowded spot positions change as a function of time and the occurrence of special events in real networks, we propose two crowded spot estimators based on these user mobility metrics. From an empirical evaluation, we show that they appear as excellent crowd detection solutions for cellular backhauling network management. Crowded spot detection can also be of high utility in the optimization of the access network under the resulting resource contention situations.

Our interests move in the second part of the dissertation into finding good offloading solutions to adequately solve resource contention situations around crowded spots. In particular we study two offloading approaches: traffic offloading and content offloading. In the former one, we present two solutions: horizontal offloading over small-cell networks and vertical offloading over Wi-Fi certified Passpoint hotspots. We study benefits and limitations of each solution. We propose a new algorithm based on cooperative game theory for resource allocation and interference management in small-cell networks. The results show important improvements and emphasize the necessity of referring to cooperative game theory concepts in the definition of spectrum contention situations. Furthermore, we show the benefits of using the newly released WiFi Passpoint solution in increasing spectrum capacity gain and decreasing user energy consumption. Finally, we also investigate how resource contention can be solved in Information Centric Networking in terms of cache allocation to multiple content providers. Also in this context, coalitional game approaches improve legacy solutions, by decreasing the content access latency in high contention situations.

Résumé en Langue Française

La thèse présente différentes solutions qui tiennent compte de la mobité humaine, pour faire face à la croissance de la demande du trafic de données mobiles dans les réseaux d'accès et de collecte.

En utilisant des données réelles de téléphonie mobile, nous étudions les trajectoires et les comportements humains. Nous montrons que ces trajectoires peuvent être estimées à haute précision, en utilisant des indices spécifiques de mobilité: le rayon de giration et le centre de gravité. Ensuite, motivé par le fait que les positions des cellules surchargées (i.e., les points chauds) changent, dans les réseaux réels, en fonction du temps et de la survenance des événements spéciaux, nous proposons deux estimateurs de cellules surchargées basés sur les mêmes indices de mobilité. Par une étude empirique, nous montrons que nos estimateurs apparaissent comme des excellentes solutions de détection de congestion pour la gestion du réseau de collecte cellulaire. La détection de congestion peut être aussi de haute utilité pour le réseau d'accès surtout dans les situations de contention de ressources.

Nos intérêts s'orientent dans la deuxième partie de la thèse vers la recherche des solutions de déchargement pour résoudre les situations de contention de ressources dans les cellules surchargées. En particulier, nous étudions deux approches de déchargement: le déchargement de trafic et de contenu. Dans la première approche, nous présentons deux solutions: le déchargement horizontal du trafic sur les réseaux de petites cellules (femtocells par exemple) et le déchargement vertical sur les points d'accès WiFi certifiés Passpoint. Nous étudions les avantages et les limites de chaque solution. Nous proposons un nouveau algorithme basé sur la théorie de jeux coopératifs pour l'allocation de ressources et la gestion des interférences dans les réseaux de petites cellules. Les résultats montrent des améliorations importantes et soulignent la nécessité de se référer aux concepts de théorie de jeux coopératifs pour la gestion des situations de contention du spectre. En outre, nous montrons les avantages de l'utilisation de la nouvelle technologie, WiFi-certifié Passpoint, pour l'augmentation du gain de capacité du spectre et la diminution de la consommation énergétique de l'utilisateur. Enfin, nous étudions comment les conflits de ressources peuvent être résolus dans *Information Centric Networking* en termes d'allocation de mémoire cache pour plusieurs fournisseurs de service. Dans ce cadre aussi, on utilise les approches par théorie de jeux coopératifs qui améliorent les performances du réseau en minimisant la latence d'accès aux contenus dans les situations de contention.

Contents

List of Figures

List of Tables

Acronyms

ANQP: Access Network Query Protocol
C-DFP: Centralized - Dynamic Frequency Planning
C-RAN: Cloud - Radio Access Network
CAPEX: CAPital Expenditure
CDF: Cumulative Distribution Function
CDR: Call Detail Records
CG: Capacity Gain
CP: Content Provider
CST: Crowd Selection Threshold
D-DFP: Distributed - Dynamic Frequency Planning
DNS: Domain Name System
DSIC: Dominant- Strategy Incentive Compatible
DSL: Digital Subscriber Line
DTB: Distance To Borders
DtC: Distance to Centroid
EAP: Extensible Authentication Protocol
ESG: Energy Saving Gain
F-ALOHA: Frequency - ALOHA
FAP: Femtocell Access Point
FTP: File Transfer Protocol
GAS: Generic Advertisement Service
GPS: Global Positioning System
HS2.0: HotSpot2.0
HTTP: Hypertext Transfer Protocol
ICN: Information Centric Networking
IP: Internet Protocol
ISP: Internet Service Provider
LRU: Least Recently Used
LTE: Long-Term Evolution
MAC: Multiple Access Channel
MC: Marginal Contribution
MMF: Max-Min Fairness
MR: Mesh Router
NGMN: Next Generation Mobile Networks

NTU: Non-Transferable Utility
OFDMA: Orthogonal Frequency Division Multiple Access
oSCTP: offload Stream Control Transmission Protocol
P2P: Peer to Peer
PC: Personal Computer
PDF: Probability Density Function
PF: Proportional Fairness
POP: Post Office Protocol
QAM: Quadrature Amplitude Modulation
QPSK: Quadrature Phase Shift Keying
RAN: Radio Access Network
RB: Router Betweenness
RD: Router Degree
RP: Router Proximity to network edge
RRU: Remote Radio Unit
RTW: Regression Time Window
SIM: Subscriber Identity Module
SINR: Signal-to-Interference plus Noise Ratio
SMS: Short Message Service
SMTP: Simple Mail Transfer Protocol
TCP: Transport Control Protocol
TE: Trajectory Error
TU: Transferable Utility
UDP: User Datagram Protocol
VoIP: Voice over Internet Protocol
VoD: Video on Demand
VM: Virtual Machine
WiMAX: Worldwide Interoperability for Microwave Access
WMN: Wireless Mesh Network

Chapter 1

Introduction

Human mobility analysis has long been a prominent research topic for social scientists, urban planners, geographers and telecommunication researchers. It has traditionally been relying on expensive data-collection methods (such as surveys and empirical observations) to sense the way people move. This high cost typically leads to seldom data collection and too small sample sizes; for example, the United States national census produces a lot of information about where hundreds of millions of people live and work only once every 10 years [15].

Nowadays, the pervasiveness of information and communication technologies is offering ideal vehicles to study human mobility cheaply, frequently, and on a global scale. Billions of people worldwide keep a phone near them most of the time. Since cellular networks need to know the approximate location of all active phones to provide them voice and data services, location information from these networks can strongly favor the emergence of advanced device and network applications.

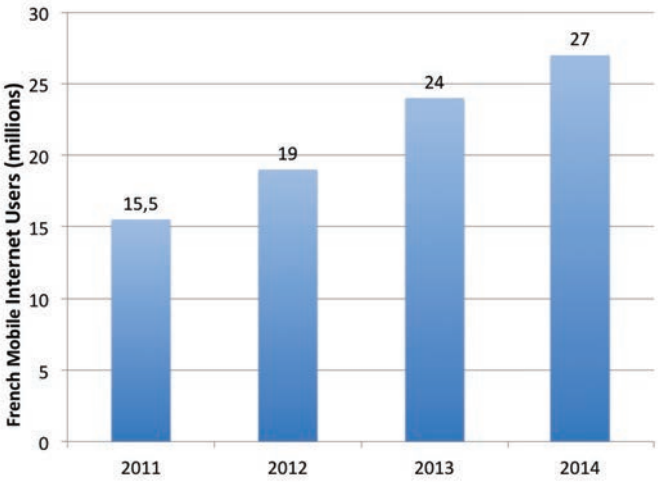


Figure 1.1: Mobile Internet users growth in France - Médiamétrie: L'audience de l'Internet mobile en France. Source: [16].

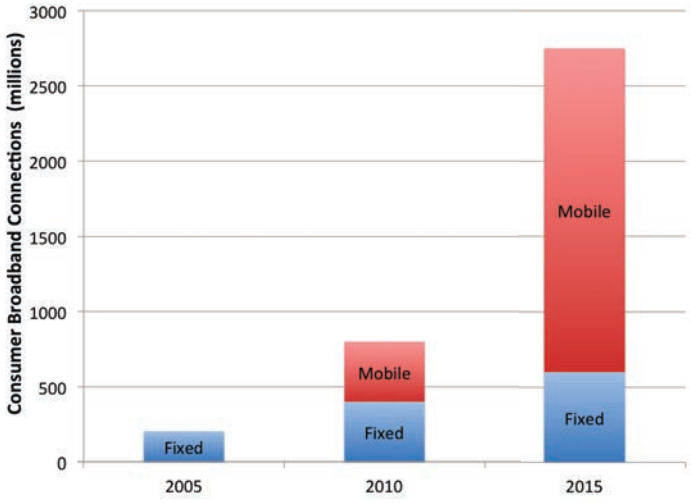


Figure 1.2: Fixed vs mobile Internet usages - Boston Consulting Group, Mary Meeker, Kleiner Perkins, Morgan Stanley Research, Berg Insight. Source: [18].

As a matter of fact, the Internet was initially conceived to serve fix and sedentary usages, while current socio-technological trends clearly show that future Internet users will be increasingly mobile and nomadic. For example, as of Institut Mediametrie [16], the number of mobile Internet users has increased by 75% from 2011 to 2014 as shown in Figure 1.1. Moreover, another research study by Morgan Stanley [17] in the United States, reports that the number of people accessing the Internet through fixed computers is expected to shrink from 240 million consumers in 2012 to 225 million in 2016. At the same time, the number of mobile users is expected to increase from 174 million to 265 million. In 2015, for the first time ever, it is expected to have more U.S. consumers accessing the Internet through mobile devices than through fixed computers. Figure 1.2 shows the scale of the change coming. As we see, mobile data usage is rising at an accelerated pace and the projections show it will soon surpass the fixed internet with an even more impressive growth in the range of mobile applications as augmented reality, video and voice recognition applications.

On the other hand, mobile phones are frequently used in everyday life for rather simple digital services, yet requiring increasing download volumes and resiliency levels [19]. At present, the rapid pace at which this evolution is taking place, and the often-inadequate management of broadband access networks, practically manifests with poor service availability, which represents a major bottleneck to the development of advanced services. According to the technical report [20], mobile data traffic will grow at a compound annual growth rate of 66% from 2012 to 2017, reaching 11.2 Exabytes per month by 2017 as shown in Figure 1.3. The exponential growth of mobile Internet usage is a relevant indicator of the overall bandwidth provisioning needed at the access networks, which have to be geographically and temporally distributed. Nevertheless, the dynamics of content consumption is very little known today, also because of the insufficient coordination between traffic en-

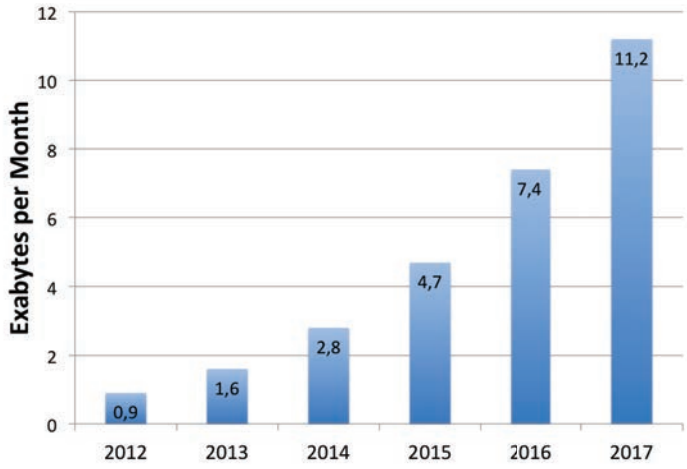


Figure 1.3: Global mobile data traffic growth - Cisco VNI global mobile data traffic forecast, 2012-2017. Source: [21].

gineering procedures and user mobility detection in nowadays telecommunication networks.

It is quite well known that the networks of many cell-phone operators recently collapsed after the release of mobile equipment software updates [22]. Moreover, cell-phone networks generally suffer during special events aggregating large masses of individuals sharing similar interests (e.g., sport events, conferences, city-wide cultural events, mass manifestations), hence accessing similar content in the same time and in the same places. In order to solve these problems and to meet mobile Internet demand, while addressing the scarcity of available mobile spectrum and minimizing capital expenditures for new infrastructure, service providers are severely challenged. They need to master the needed capacity expansion in their backhauling network, otherwise the data traffic will sooner or later clog their networks. Next-generation network deployments promise to deliver higher bandwidth and speed, but they often imply high capital and operational expenditures [23].

An alternative economically and technically viable way is represented by mobile offloading solutions. Such solutions can alleviate radio spectrum congestion and lower the operating load on base stations and the backhauling network. They can also represent a business opportunity for service providers, for example offering small cell offloading solutions, while helping customers to reduce their costs and increasing network performance. The most common offloading solutions used thus far are femtocells (i.e., used for example by Vodafone in UK and by SFR in France) and Wi-Fi hotspots representing a very significant part of the iPhone traffic at Orange France, O2 in UK and AT&T in the USA [24]. According to the technical report [20] and as shown in Figure 1.3, mobile data traffic grew 81% in 2013, and Cisco says it would have grown 98% if operators had not been able to offload traffic to Wi-Fi connections and femtocells. Globally, 45% of total mobile data traffic was offloaded onto the fixed network through Wi-Fi or femtocells in 2013 and carriers offload around 1.2 Exabytes of mobile data traffic onto the fixed network each month.

The purpose of this dissertation is, by providing a deep understanding of human mobility patterns of anonymous aggregates of individual, to propose novel mobile offloading solutions for telecom access networks. The technical scope, depicted in Figure 1.4, ranges from mobile data mining and usages cartography to wireless resource management and content distribution.

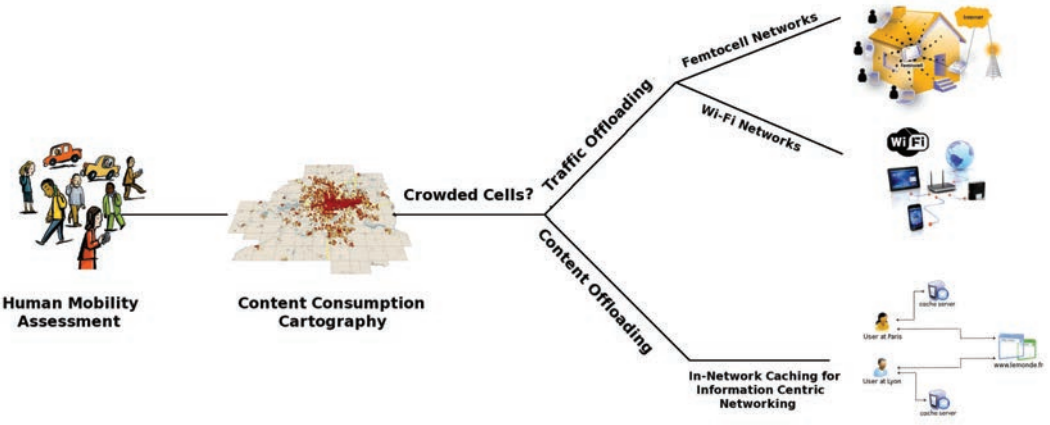


Figure 1.4: Scope of the dissertation

Structure of the dissertation

The first part of the dissertation, Chapters 2 and 3, is dedicated to human mobility analysis and content cartography distributions, while in the second part, Chapters 4, 5 and 6, we concentrate on resource contention and offloading solutions. The third part, Appendix A, contains additional complementary information on cooperative game theory.

Chapter 2 investigates the opportunity to use human trajectories as an approach to address the qualification of mobility patterns. We evaluate different interpolation methods to model user movements between two places. We qualify compact mobility metrics, such as the radius of gyration and user centroid, showing the dependency of user's trajectory accuracy with respect to these metrics.

According to the mobility metrics described in Chapter 2, we study in **Chapter 3** the content consumption cartography of users in three cities in France for different days. Moreover, we propose two estimators of crowded spots, that is, those places and groups of antennas absorbing a large portion of mobile traffic. We compare a territory-based estimator and a trajectory-based estimator. We show that the proposed methods appear as efficient crowd detection solutions for cellular and backhauling network management, their errors strictly decrease with the cell load, they become very small for highly loaded cells while following different regimes.

After determining the highly loaded cells in the network, and in order to satisfy users' demands in those cells while ensuring a good quality of service, our interest moves, in the next part of the dissertation, to finding offloading solutions to lower traffic and solve resource contention around crowded spots. This part is composed of three chapters. In **Chapter 4**, we tackle the case of small-cell offloading solution. We propose a new resource allocation and interference management solution for small-cells deployed within the macro-cell coverage zone. We rely on game-theoretical approaches for the allocation scheme. After a comparison with two state of the art allocation schemes, we show that the proposed approaches outperform the other approaches and provide a fair and efficient resource allocation for small-cell networks.

Under the same motivation to offload mobile data traffic from crowded cells to other networks, we study in **Chapter 5** the capacity gain of offloading mobile data through WiFi-certified Passpoint hotspots. As a matter of fact, the new Passpoint program aims to make the WiFi network a "true extension of service provider networks", letting users to roam from one hotspot to another with no manual action, just like cell phone networks switching seamlessly from one cell tower to another. We compare different offloading policies and hotspot placement schemes, showing that the usage of Passpoint control-plane information leads to a significant radio spectrum capacity gain.

Finally, in **Chapter 6**, our attention moves to the mobile backhauling network. We propose a content offloading algorithm for cache allocation in Information Centric Networking based on game-theoretical approaches. We show that the proposed solution improves legacy solutions, by decreasing the content access latency in high contention situations.

Chapter 7 concludes the dissertation and contains perspectives for further work.

Appendix A presents principles of cooperative game theory, some of which are recalled across the dissertation.

Part I

Mobility Estimation

Chapter 2

Human Mobility Assessment

A deep understanding of human-mobility patterns can yield interesting insights into a variety of important societal issues. As a matter of fact, evaluating the effect of human displacements on the environment maps to determining how large populations move in their daily lives. Likewise, understanding the spread of a disease requires a clear picture of how humans move and interact [25]. Other examples abound in such fields as urban planning, where knowing how people come and go can help determining where to deploy infrastructure and how to reduce traffic congestion [26]. Moreover, cloud and content delivery networks [27], and location-based recommender systems [28] [29], can all benefit from quantitative and qualitative knowledge of users' mobility patterns.

Nowadays, the huge worldwide mobile-phone penetration is increasingly turning the mobile network into a gigantic ubiquitous sensing platform, enabling large-scale analysis and applications. The objective of this chapter is to check the appropriateness of using mobile phone data in estimating the trajectory of people across urban areas and to assess the pertinence of different conceivable trajectory estimation approaches in terms of deviation from real human trajectories.[1] After reviewing related work from the literature and describing the dataset used in our study, we define an appropriate, compact and easy to compute mobility metric to qualify user mobility, the "radius of gyration". Then we present the different interpolation methods used to model user trajectories and we compare them together in terms of deviation from real trajectories. The results highlight the dependence of mobility characteristic with the radius of gyration.

2.1 Related Work on Macro-Mobility Analysis

In recent years, mobile data-based research reached important conclusions about various aspects of human characteristics, such as human mobility and calling patterns [30] [31], virus spreading [32] [33], social networks [19] [34] [35], urban and transport planning [36] [37] and network design [38].

Nevertheless, in such user displacement sampling data, a high uncertainty is related to users movements, since available samples strongly depend on the user-

[1] The contents of this chapter are presented in [1], [5] and [6].

network interaction frequency. For instance, Call Data Records CDRs (i.e., data records containing information related to phone calls, such as the origination and destination addresses of the calls, the starting time and the duration of the calls, etc.) alone do not provide a sufficient fine granularity and accuracy, exhibiting a vast uncertainty about the periods when the user is not active, i.e., not communicating. This represents a challenge for applications and analysis assuming ubiquitous and continuous user-tracking capability.

Some modeling techniques have been proposed in the literature to predict user movement between two places.

Authors in [39] and [40] infer the top-k routes traversing a given location sequence within a specified travel time from uncertain trajectories; they use check-in datasets from mobile social applications.[2] The proposed methods permit to identify the most popular travel routes in a city, but they do not allow constructing time-sensitive routes, i.e., the proper visiting time of places as well as the transit time between two places cannot be determined.

Authors in [41] propose a space-time prism approach, where the prism represents reachable positions as a space-time cube, given user's origin and destination points – i.e., the assumption of knowing the location of a user at one time and then again at another time fits well mobile phone data in which we only know users' position during their communication events – as well as time budget and maximum speed. Spatial prisms so allow evaluating binary statements, such as the encounter potential between two moving users. However, the maximum speed that cannot be set for all users in general, is considered as a major limitation of the model applicability.

Similarly, the authors in [42] propose a probabilistic extension of the space-time approach, applying a non-uniform probability distribution within the space-time prism. A strong assumption made therein is that users move linearly over time. This hypothesis is in a high contrast with the results obtained in [43] that show the tendency of users to stay in the vicinity of their call places.

Authors in [43] propose a probabilistic inter-call mobility model, using a finite Gaussian mixture model to determine users' position between their consecutive communication events (call or SMS) using CDRs. The model evaluates the density estimation of the spatio-temporal probability distribution of users position between calls, but it does not give an approximation of the fine-grained trajectory between calls.

User displacements using GPS traces have been analyzed in [44]; the authors find that the displacement behavior shows Levy walk properties (i.e., random walk with pause and flight lengths following truncated power laws). While very interesting in order to model inter-contact time distributions and general massive mobility, such random-based approaches cannot give precise approximations between given points on a per-user basis.

In this chapter, we assess the pertinence of different conceivable trajectory estimation approaches in terms of deviation from real available trajectories, via the analysis of real mobile network data for the state of Massachusetts in USA.

[2] In recent years, mobile social applications have become so popular that they generate huge volume of social media data, such as check-in records or geo-tagged photos. In a check-in service, users note their locations via a mobile phone to share photos, activities etc.

2.2 Airsage Dataset Description

We use a dataset consisting of anonymous cellular phone signaling data collected by AirSage [45], which converts cellular network signaling data into anonymous locations over time for cellular devices. The dataset consists of location estimations - latitude and longitude - for about one million devices from July to October 2009 in the state of Massachusetts. These data are generated each time the device connects to the cellular network including:

- when a call is placed or received (both at the beginning and end of a call);
- when a short message is sent or received;
- when the user connects to the Internet (e.g., to browse the web, or through email synchronization programs).

The location estimations[3] not only consist of identifiers of the mobile phone towers that the mobile phones are connected to, but also an estimation of their positions generated through triangulation, using the AirSage's Wireless Signal Extraction technology [45]. This technology aggregates and analyzes wireless signaling data[4] from mobile phones to securely and privately monitor the location and movement of populations in real-time, while guaranteeing acceptable user anonymity and privacy. We note that the observation period is limited to one day because the anonymized user identifiers change from one day to another to ensure user privacy.[5]

2.3 Mobility Metrics

People do not behave similarly; each person has different mobility habits and behaviors. Many studies have been conducted to find mobility patterns from network data sampling, from very complex and complete ones able to determine precise motifs (e.g., [46]), to more aggregated and synthetic ones extracting a single parameter to characterize user mobility. A sufficiently precise, synthetic and easy to compute parameter is the radius of gyration. In [30], authors show that the radius of gyration is defined as the deviation of user positions from the corresponding centroid position.

Given a user $u \in U$ (i.e., U represents the set of all users) who has been located at $n^u(t)$ locations until time t, its radius of gyration can be thus computed as:

$$r_g^u(t) = \sqrt{\frac{1}{n^u(t)} \sum_{i=1}^{n^u(t)} |\vec{r}_i^u(t) - \vec{r}_{cm}^u(t)|^2} \tag{2.1}$$

where $\vec{r}_i^u(t)$ represents the i^{th} position recorded for the user u and $\vec{r}_{cm}^u(t)$ represents its centroid defined as the center of mass of the user's recorded displacements.

[3]Each location measurement is characterized by a position expressed in latitude and longitude and a timestamp.

[4]The location measurements are generated based on signaling events, i.e., when a cellphone communicates with the cellular network's elements through control channel messages.

[5]The results of this study have been validated for many days.

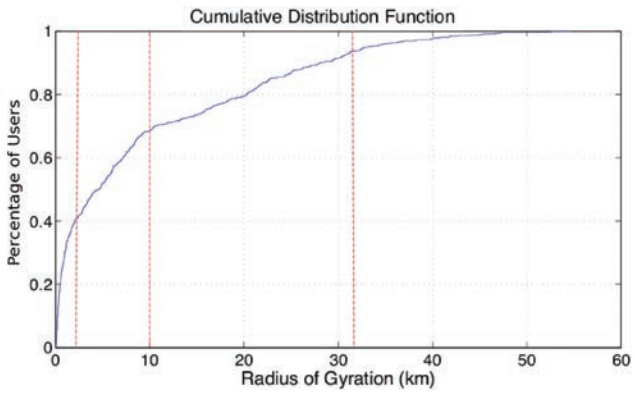


Figure 2.1: Cumulative Distributive Function of the radius of gyration.

It is obtained as:

$$\vec{r}^{\,u}_{cm}(t) = \frac{1}{n^u(t)} \sum_{i=1}^{n^u(t)} \vec{r}^{\,u}_i(t) \tag{2.2}$$

To explore the statistical properties of the population's mobility patterns, the cumulative distribution function (CDF) of the radius of gyration for the smartphone users is represented in Figure 2.1 using the dataset described in Section 2.2. It is easy to distinguish four main user categories[6] based on steep changes in the CDF slope.

- Users with $r_g \leq 3km$, who can be identified as the most sedentary people.
- Users with $3km \leq r_g \leq 10km$. They might be identified as urban mobile people as the diameter of the Boston metropolitan area is very approximately around 10 km.
- Users with $10km \leq r_g \leq 32km$. They might be identified as peri-urban mobile people as the diameter of the Boston peri-urban area is very approximately around 32 km.
- Users with $r_g \geq 32km$, who can be identified as commuters spanning the whole Massachusetts state area.

This ranking seems appropriate as the total traveled length increases with the radius of gyration, as displayed in Figure 2.2[7]. Moreover, this correlation may be interpreted by the fact that the radius of gyration can be viewed as a proper "territory" of each user, and thus increasing the territory area means that the person is able to move over longer distances.

[6]This classification depends on city size, economic degree and other parameters.

[7]The absolute length is of course overestimated with respect to the real one. After looking into details, we discover that this is due to handover flipping among close antennas. The important aspect here remains the relative (and not the absolute) increasing trend.

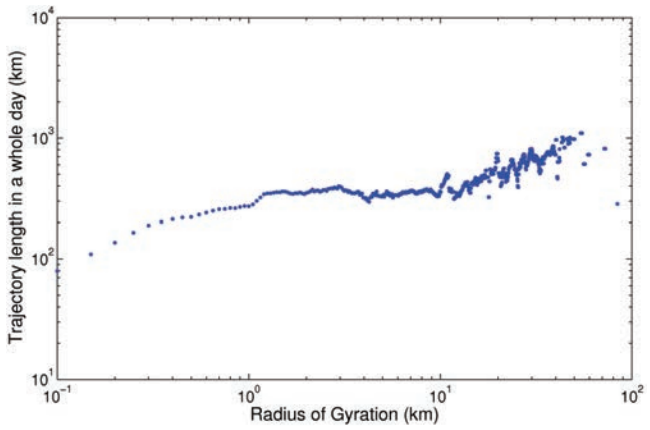


Figure 2.2: Correlation between trajectory length and radius of gyration.

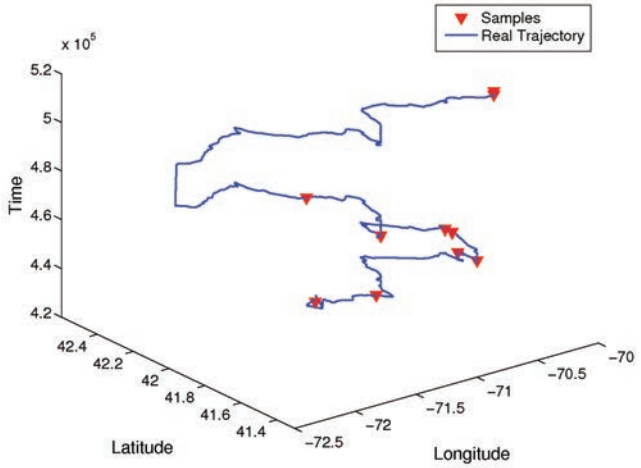
2.4 Trajectory Interpolation Methods

Different interpolation methods have been proposed in the literature to describe moving object trajectories. We present in the following a selection of classical ones, showing how they approximate the real trajectory; referring to the example in Figure 2.3.

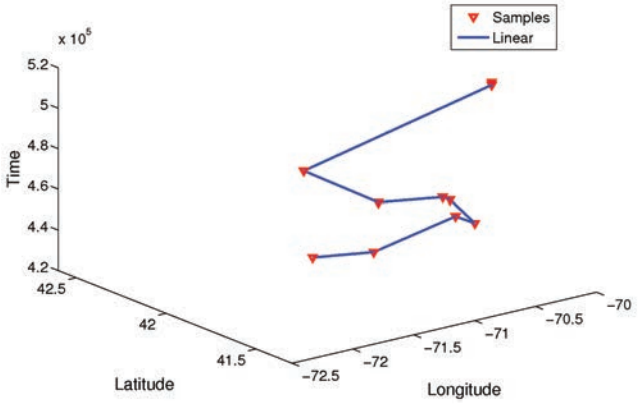
- the *Linear Interpolation*, is a popular interpolation used for moving objects [47]. It is presented in Figure 2.3(b). It is obtained by joining straight interpolating lines between each pair of consecutive samples. Users are supposed to move at a constant speed along the straight lines. One limitation of the linear interpolation is that it can fail in some situations where the interval of time between interpolated points is high. For example, suppose there are two points A and B in the road network with a curved path connecting them: with the linear interpolation we always assume the user drives along a straight line.

- the *Nearest-neighbor Interpolation*, is an interpolation often used in mapping programs [48], also known as proximal interpolation. It consists of taking, for each position, the value of the nearest sampling position in time (not plotted in Figure 2.3 because of the simplistic decision). Therefore, if we detect the same user in two different instants, at point A and point B respectively, the nearest interpolation attaches the user to position A for the first half period of time, and to position B for the second half.

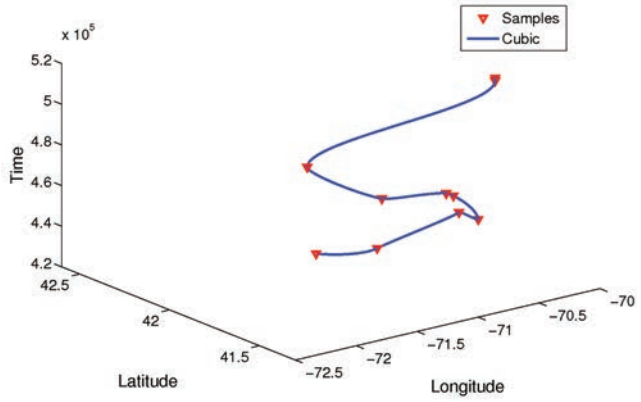
- the *Piecewise Cubic Hermite Interpolation* is often used in image processing studies (see [49]). It is depicted in Figure 2.3(c). It is a third-degree spline that interpolates the function by a cubic polynomial using values of the function and its derivatives at the ends of each subinterval. This method interpolates the samples in such a way that the first derivative is continuous, but the second derivative is not necessary continuous. The slopes are chosen in a way that the function is "shape preserving" and respects monotonicity. Suppose a



(a) Real Trajectory



(b) Linear Interpolation



(c) Cubic Interpolation

Figure 2.3: Real and estimated trajectories with different interpolation techniques.

subinterval $[x_1, x_2]$, with the function values: $y_1 = f(x_1)$, $y_2 = f(x_2)$ and the derivative values $d_1 = f'(x_1)$ and $d_2 = f'(x_2)$ are given. The cubic polynomial function in this subinterval is given by:

$$C(x) = a + b(x - x_1) + c(x - x_1)^2 + d(x - x_1)^2(x - x_2) \tag{2.3}$$

satisfying $C(x_1) = y_1$, $C(x_2) = y_2$, $C'(x_1) = d_1$ and $C'(x_2) = d_2$ This interpolation determines the coefficients a, b, c and d noting that:

$$C'(x) = b + 2c(x - x_1) + d[(x - x_1)^2 + 2(x - x_1)(x - x_2)] \tag{2.4}$$

is also continuous. The solution to this system is given by: $a = y_1$; $b = d_1$; $c = \frac{y_1' - d_1}{x_2 - x_1}$ and $d = \frac{d_1 + d_2 - 2y_1'}{(x_2 - x_1)^2}$, where $y_1' = \frac{y_2 - y_1}{x_2 - x_1}$.

2.5 Trajectory Modeling

In order to qualify the precision of different interpolation methods, we have to determine the deviation of the estimated trajectory from the real one. To determine real user trajectories, we fine-select data of those smartphone holders with a lot of samplings, typically those data-plan users with persistent Internet connectivity due to applications such as e-mail synch. By selecting users with more than 1000 connections (position samplings) during a given day, we can filter out 707 smartphone users from the whole dataset.

In order to reproduce "normal-phone user" sampling, we subsample[8] real trajectories (i.e. smartphone user trajectories) according to the experimental inter-event time distribution (i.e., the inter-event time is defined as the time between two consecutive connections, done by the same user, to the mobile network), extracted from the same dataset and given in Figure 2.4. We determine it by analyzing *real* normal-phone user samplings (for which the real trajectory is unknown), available in the Airsage original dataset. Therefore, we extract, from the real trajectory, a first random position $P_i(longitude_i, latitude_i, time_i)$, then the corresponding next positions are extracted according to the inter-event time distribution. Hence, given a real trajectory with a high number of positions, and its subsampling that reproduces normal user's activity, we apply an interpolation method (as of Section 2.4) to estimate the trajectory between the subsampled points. Given a real user's position $P_i(longitude_i, latitude_i, time_i)$, we estimate its corresponding position P_i' $(longitude_i', latitude_i', time_i)$. Then we determine the deviation between the two points P_i and P_i' as the distance separating the exact position P_i to the estimated position P_i' in the interpolating curve joining the samples.

2.6 Results

In this section, we present the main results obtained by applying the interpolation methods introduced in Section 2.4. We first define the *Trajectory Error* of a user

[8]The ratio between the number of the sampled positions to the total number of known positions (data-plan smartphone user) is defined by the subsampling ratio. The subsampling process is independent and identically distributed.

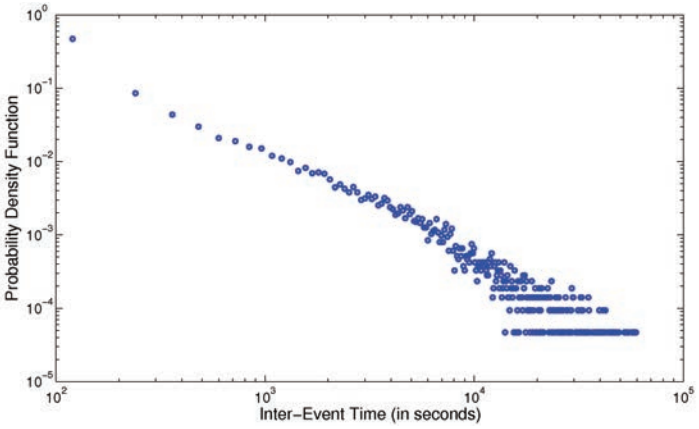


Figure 2.4: Probability density of the inter-event time.

at time t_i, $TE(t_i)$, as the ratio of the deviation between its real position from its estimated position at time t_i, $\Delta(t_i)$, (computed as described in Section 2.5) over its radius of gyration, i.e., for a given user u:

$$TE(t_i) = \frac{\Delta(t_i)}{r_g^u} \tag{2.5}$$

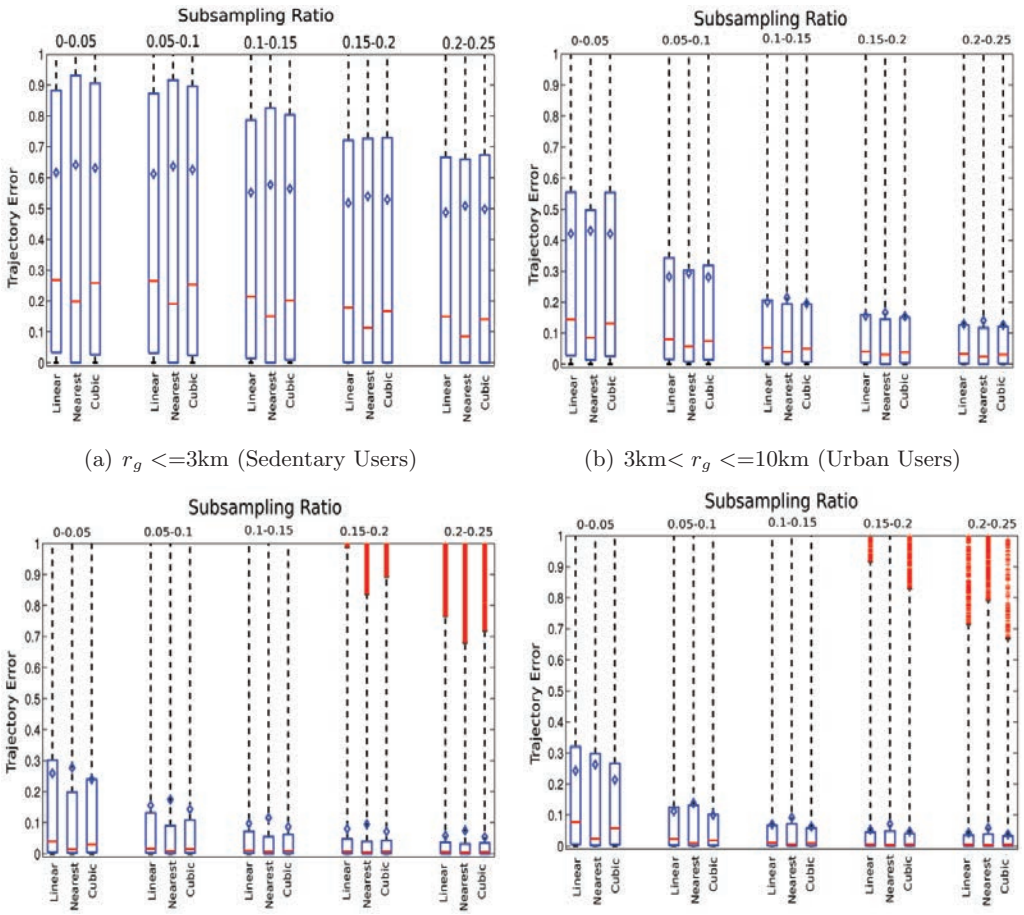
Then we evaluate the TE distributions for different interpolation methods and for different radii of gyration. We further investigate the statistical distribution of TEs with respect to mobility parameters in order to understand which method performs better for each particular category of users.

2.6.1 Interpolation Error

Figure 2.5 reports the boxplots[9] as well as the average (the blue star) statistics of the TEs for different interpolation methods. We note that the red lines in the boxplots represent the median values. At a first view, looking at the TE averages, we can assess that:

- The TE decreases with the increase of the subsampling ratio, for whatever interpolation, which is reasonable as one can get more accurate computations with more samples.
- The gap between the three interpolation methods decreases with the increase of the radius of gyration, especially for those users with a radius of gyration higher than 10 km, i.e., those who could be considered as peri-urban users and commuters (see Section 2.3).
- The lowest mean TE among different interpolation methods depends on the category to which the user belongs. Indeed, for those users having a radius of gyration less than 3 km, i.e., sedentary users, the linear interpolation method

[9] i.e., first quartile, median, third quartile, maximum, minimum and outliers. It is worth noting that some maximum and outliers are cut in the figure for the sake of readability.



(a) r_g <=3km (Sedentary Users)

(b) 3km< r_g <=10km (Urban Users)

(c) 10km< r_g <=32km (Peri-urban Users)

(d) r_g >32km (Commuters)

Figure 2.5: Boxplots of the trajectory errors with different interpolation methods.

presents the smallest mean TE when compared to other methods. Instead, for those users having a higher radius of gyration, especially for commuters (i.e., those with a radius of gyration of more than 32 km), the cubic interpolation presents the smallest mean TE. Finally, for urban users with a radius of gyration between 3 and 10 km, the linear and cubic interpolations show close performance.

Further looking into the whole statistics of the TEs, including median and quartile lines, we can determine that:

- The median is always lower than the average, which indicates that the population contains an important part of users with much higher TEs than the rest of the population.

- Overall, the nearest interpolation shows better median statistics than all the other interpolations for all user categories with different radii of gyration.

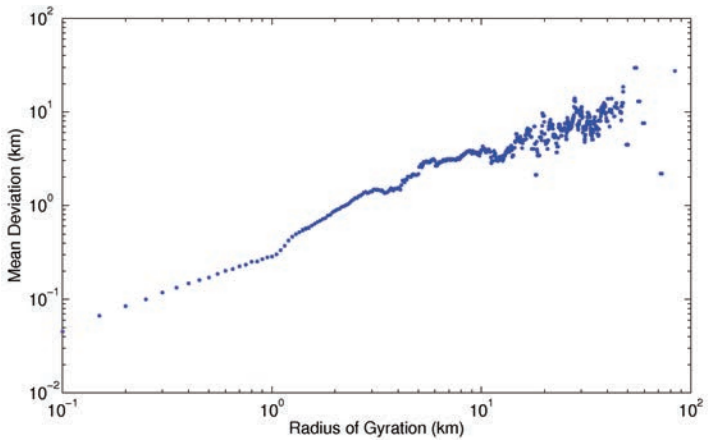


Figure 2.6: Correlation between the mean deviation and the radius of gyration.

- The median TE becomes very low for subsampling ratio of more than 0.1 for peri-urban and commuter users.

Based on these results, we see that the best interpolation method depends on the mobility category, i.e., the user radius of gyration. In order to determine the correlation function between the deviation and the radius of gyration, Figure 2.6 shows for each user (one point), the correlation between the radius of gyration and the mean trajectory deviation (just for the linear interpolation case, knowing that other interpolation methods give a very similar trend). The trend being generally increasing, we have thus a positive correlation between the mean deviation and the radius of gyration. Indeed, with the increase of the radius of gyration, users are able to move over longer distances, the distance between two samples increases, hence finding a good interpolation method that accurately approximates the real trajectory traversed by the user gets more challenging.

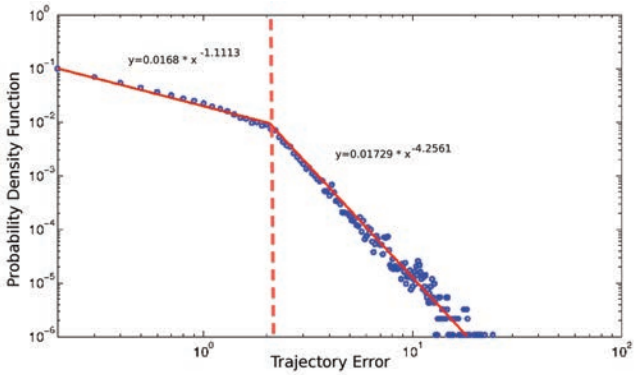
2.6.2 Interpolations' Probability Density Function

How to explain the huge gap between averages and medians, and the performance inversion indicating that nearest interpolation is on median the best interpolation, whatever the user category and the subsampling ratio are, is a matter of discussion.

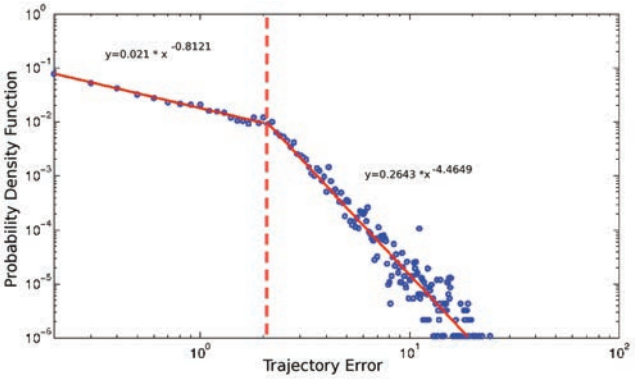
We interpret it with the fact that the median does not weight, as the average does, the TEs of those users' moves for which a trajectory interpolation, whatever the type is, is not appropriate; that is, those extraordinary moves that deviate too much from conventional paths. For example, the moves of users having a backward path behavior (e.g., tourists moving back to already visited places, etc.) can hardly be modeled by intuitive interpolations. The majority of ordinary moves, with long stops at visited places, are instead captured by the median. For ordinary moves, the nearest interpolation (introducing long stops at each sample and instantaneous displacement) appears as the best approximation.

The presence of a subset of the population which behaves very differently than the rest is confirmed by the fact that the average is often close and sometimes higher than the third quartiles in Figure 2.5, and by the presence of many outliers especially

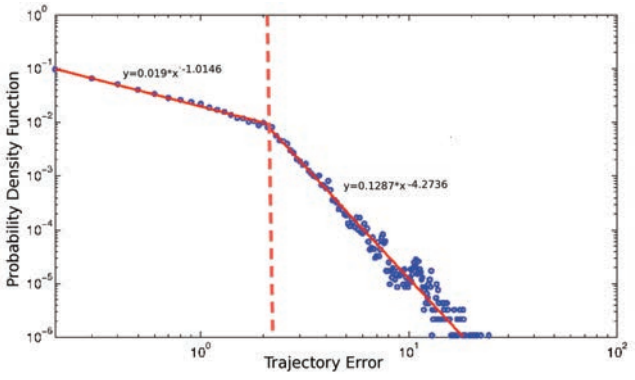
for high subsampling ratios. The ordinary moves represent more than 75% of the whole moves, while the extraordinary ones having so high TEs are around 25% of the whole moves.



(a) linear



(b) nearest



(c) cubic

Figure 2.7: Probability Density Function of the trajectory error - (subsampling ratio: 0-0.05).

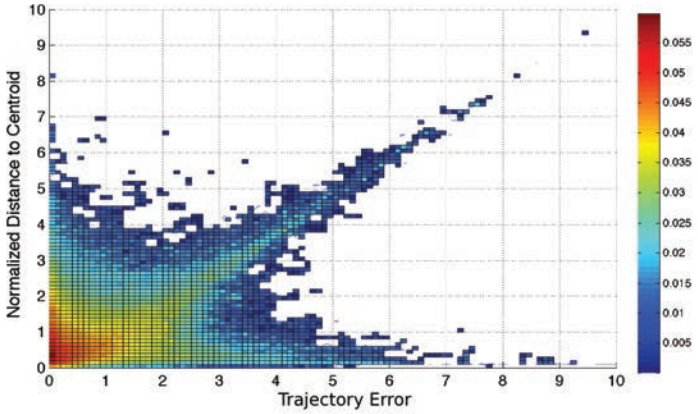


Figure 2.8: Joint probability between the trajectory error and the distance to centroid.

In order to further explore the statistical properties of the TE distribution, Figure 2.7 shows the probability density function (PDF) of the TE for the linear, cubic and nearest interpolations respectively. It is easy to notice that there are two regimes. The distribution of TEs over all users' positions is well approximated by a combination of two power law distributions joined by a breakpoint. It is surprising to notice that the breakpoint is the same (approximately equal to 2.2) for the different interpolation methods.

In practice, what does this power law breakpoint really mean? We interpret it as the point after which the trajectory error properties change abruptly. As a matter of fact, Figure 2.7 shows that, using our interpolation methods, it is highly probable to obtain TEs of less than 2.2 while the probability of having high TEs of more than 2.2 decreases rapidly (i.e., the slope of the curve after the breakpoint is higher in absolute value than that before the breakpoint). An important research question shows up here: is there any dependency between user's position (i.e., around the centroid) and its TE? Does the user located inside its "territory" (the circle of radius equal to the radius of gyration) have low TE?

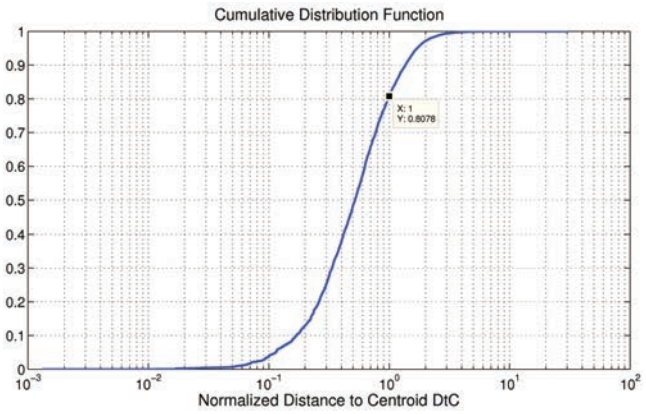
In order to evaluate this dependency, we first define the *Distance to Centroid (DtC)* parameter of a user u located at position P_i at time T_i as the normalized user's position with respect to its radius of gyration given as follows:

$$DtC = Distance(P_i, \vec{r}_{cm}^{\,u})/{r_g}^u \tag{2.6}$$

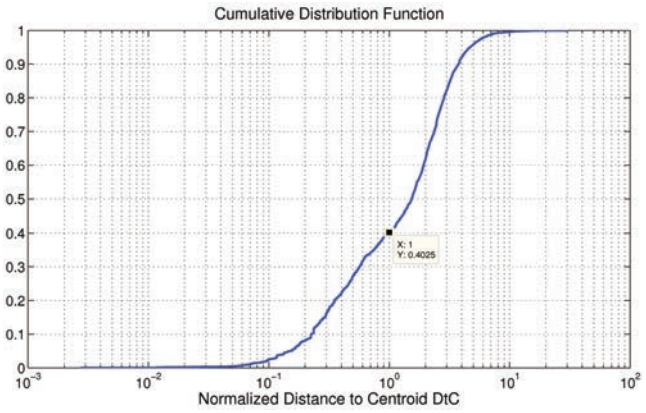
We note that if DtC is lower than 1, it means that the user is inside its territory while for DtC above 1, the user is outside its territory. We then plot in Figure 2.8 the joint probability distribution between TE and DtC.

The figure shows that for small TEs, it is highly probable that the user is within its territory (i.e., DtC lower than 1); while for high TEs, it is highly probable that the user is outside its territory.

These values can alternatively be analyzed by the conditional cumulative density distribution of the two variables, TE and DtC, as presented in Figure 2.9. We can determine therein that:



(a) TE $\leq$ 2.2



(b) TE $\geq$ 2.2

Figure 2.9: Conditional Cumulative Density Function of the distance to centroid.

- when small TEs occur, we have a high probability (81%) that the user is inside its territory, and a low probability (19%) that the user is outside it.
- When big TEs occur, we have a probability of 40% that the user is inside its radius of gyration and a probability of 60% that the user is outside its radius.

Therefore, we have an empirical evidence that the trajectory error increases and its characteristics change when the user moves beyond the territory area roughly approximated by the radius of gyration.

2.7 Summary

Motivated by recent research on human mobility characterization based on cellular network log and probe data, we study in this chapter the appropriateness of using such data for estimating the trajectory of people across urban areas. The applications are manifold, ranging from content delivery network design to urban planning, yet this study is application independent and is of a fundamental nature.

Using data for millions of users from the state of Massachusetts in USA, we select data-plane smartphone users to get very precise localization data for a few hundreds of users. Then, we subsample these paths following the experimental normal user inter-event distribution, and apply to the subsampled position different interpolation methods. Finally, we analyze their errors to better understand the appropriateness of the different methods in detail, and of interpolation methods in general, for different mobility classes.

The major findings of our work can be summarized as follows:

- The radius of gyration is an appropriate, compact and easy to compute parameter to qualify user mobility in a urban area network scope.
- The linear interpolation is the best approximation for sedentary users, linear and cubic interpolations work well for urban users, and the cubic interpolation is the best for peri-urban users and commuters.
- Separating ordinary moves following conventional paths from the minority of user moves with unpredictable displacements, the nearest interpolation appears as the best approach whatever the mobility class is.
- Interpolation methods clearly work better and have lower errors when applied within the territory of the user defined by its radius of gyration.

As already mentioned, we believe the applications are manifold. We are in particular interested in determining how content and Cloud delivery points in an urban and peri-urban environments can be identified and adapted online by inferring basic user mobility properties from big data log coming from cellular networks. These experimental findings constitute the pillars of the next chapter.

Chapter 3

Estimation of Mobile Crowded Spots

An important question in the evolution of mobile cellular networks consists of determining whether, how and where to deploy adaptive content and cloud distribution solutions at the access and backhauling network level. Intuitively, an adaptive placement of content and computing resources in the most crowded regions can grant important traffic offloading, improve network efficiency and decrease content access latency. As a matter of fact, one of the main challenges of the 5G is to reduce delivery costs and latency and to reach an end-to-end latency of less than 1 ms [50]. In this chapter[1] we further investigate this issue and propose two crowded spot estimators computed using the predefined mobility metrics in Section 2.3. Firstly, we characterize the content consumptions habits as a function of time, space and applications and review the state of the art on cell load estimation techniques.

3.1 Introduction

The exponential growth of mobile Internet usage is a relevant indicator of the overall bandwidth provisioning needed at the access networks, which has to be geographically and temporally distributed. Nevertheless, the dynamics of content consumption are very little known today and there is an insufficient coordination between traffic engineering procedures and user mobility detection in nowadays telecommunication networks.

It is quite well known that the networks of many cell-phone operators have recently collapsed after the release of mobile equipment software updates [22]. Moreover, cell-phone networks generally suffer during special events aggregating large masses of individuals sharing similar interests (e.g., sport events, conferences, city-wide cultural events, mass manifestations), hence accessing similar content in the same time and in the same places.

For these reasons, the detection of crowded spots in access networks is considered to be nowadays a necessary step towards 5G as it can be of great benefit for many use-cases. From the one hand, a dynamic placement of contents (in the "Cloud") close to the dense access points could allow a seamless service provisioning to the users, without performance degradation across the network. Moreover the

[1]The contents of this chapter are presented in [1], [4], [6], [7], and [12].

service resiliency could be also further enhanced with lower connection latency and data loss. On the other hand, a dynamic resource allocation among access points taking into account their loads can enhance users' quality of service and network performances. In this chapter, we analyze user content consumption habits from the Orange cellular network in the Paris metropolitan area. Moreover, we propose two crowded spot estimators: a "trajectory-based" and a "territory-based" estimation techniques. We evaluate these techniques using the Airsage data [45] (already presented in Section 2.5, and Orange cellular network data in France (i.e., explained in Section 3.2).

We show that the proposed estimators appear as excellent crowd detection solutions for cellular backhauling network management; their errors decrease with the cell load, and they become almost negligible for the most crowded spots.

3.2 Orange Dataset Description

The Orange dataset used in our study comes from network management tickets, generated each time a mobile device uses wireless mobile network for Internet data exchange (i.e., what is commonly referred to as "mobile Internet" service). The network probe data exploited in the study provide information on the protocol used for the communication, so it is possible to categorize the traffic by application (Web, VoIP, P2P, streaming etc.). Data are individual, so all user identifiers were irreversibly anonymized before we could access the data to protect user's privacy. The probe collects 6-minute spaced data session information, assigning the session to the cell identifier of the last used antenna. Data are recorded on a per-user basis and cover all mobile phone users of the Orange network in France.

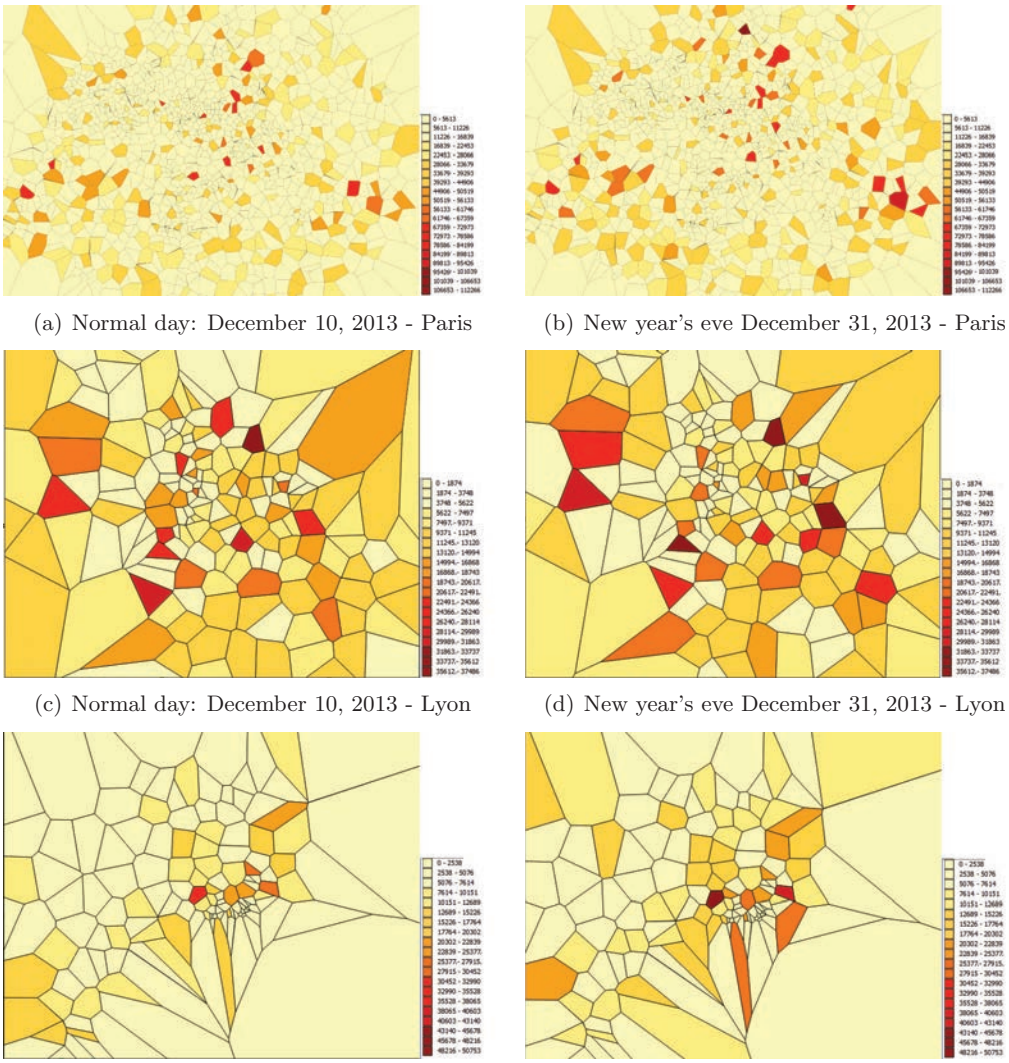
3.3 Content Consumption Cartography

In this section, we synthetically characterize content consumption habits from the Orange cellular network using the dataset described in Section 3.2, as a function of time, place and applications in different city topographies: Paris, situated within a large metropolitan area, Lyon a large business city in the central part of the country with a lower population density than Paris, and Nice, the fifth most populous city in France, located on the Mediterranean coast, with a horizontal topography rather than circular one as in Paris and Lyon and a population density close to Paris.

We consider datasets of two days, one "normal day" with standard content consumption activity (Tuesday, December 10^{th}, 2013), and one "special day" where a particular content consumption is expected (New year's eve on Tuesday, December 31^{st}, 2013).

The data cover about 1 million mobile phone users from Paris, Lyon and Nice regions (we study the mobility of around 650,000 users in Paris; 250,000 in Lyon and 100,000 in Nice). We decompose the considered regions into Voronoi cells based on base station positions, each base station being composed of few antennas able to host up to roughly 1200 users.[2]

[2]Since we do not have the real cell structures, we use the Voronoi cells to determine the coverage zone of the base stations.



(a) Normal day: December 10, 2013 - Paris

(b) New year's eve December 31, 2013 - Paris

(c) Normal day: December 10, 2013 - Lyon

(d) New year's eve December 31, 2013 - Lyon

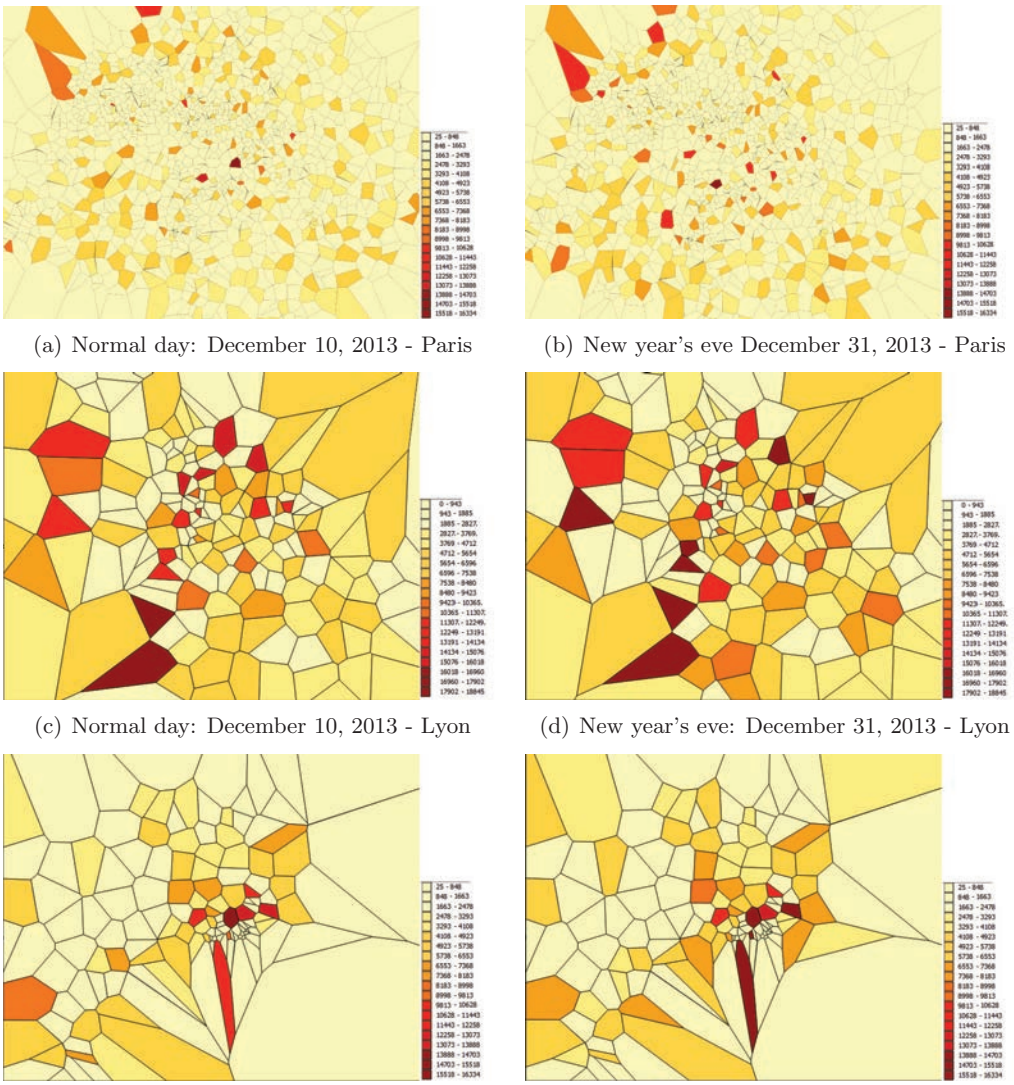
(e) Normal day: December 10, 2013 - Nice

(f) New year's eve December 31, 2013 - Nice

Figure 3.1: Traffic volume distribution in different days and different cities.

3.3.1 Content Consumption Spatial Distributions

Looking for differences in users' behavior in normal and special days is a first naturally arising research question. Figure 3.1 shows the traffic volume (i.e., in bytes) of each cell in the three regions, while Figure 3.2 shows user volume defined as the number of users attached to each cell in the considered regions. We can notice that the number of cells is not the same in all regions (i.e., Paris has the highest number of cells since it has the highest population density. We remark also that in the same region and the same day, a few cells are clearly more crowded than others presenting a large number of users and a large traffic volume; the reason is that they cover clearly identifiable content consumption spots. They are likely the public spaces where people use to gather together. The crowded spots are clearly not the same in



(a) Normal day: December 10, 2013 - Paris

(b) New year's eve December 31, 2013 - Paris

(c) Normal day: December 10, 2013 - Lyon

(d) New year's eve: December 31, 2013 - Lyon

(e) Normal day: December 10, 2013 - Nice

(f) New year's eve: December 31, 2013 - Nice

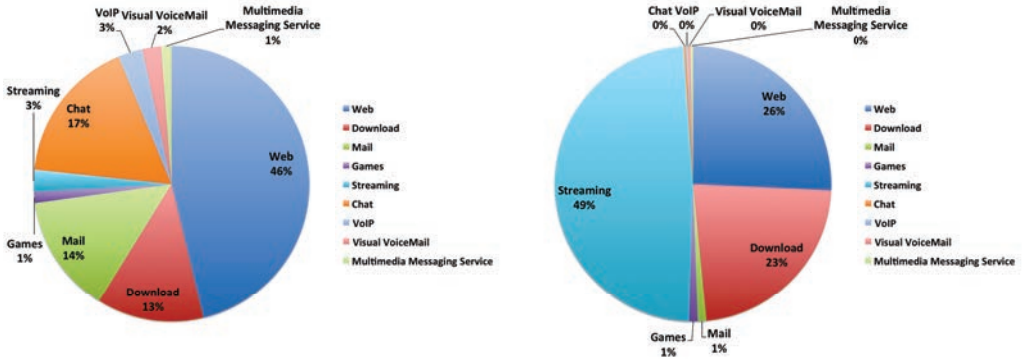
Figure 3.2: User volume distribution for different days and different cities.

the two considered days; while in the normal working day, the loaded cells seem to be the places where companies and enterprises are located, the touristic places are the most loaded ones in terms of traffic and user volume during the new year's eve.

3.3.2 Application Usages Distributions

Upon the detection of the content consumption cells, it is worth evaluating the applications that are mostly used during the special day (i.e., the new year's eve) so as to provide a basic traffic model for these situations and draw some observations.

Figure 3.3(a) reports the usage proportion of applications (i.e., the number of users connected to each application to the total number of users), while Figure 3.3(b)



(a) Application usage ratios

(b) Application traffic ratios

Figure 3.3: Application usage and traffic distributions.

represents the proportion of traffic generated from each application (i.e., the number of bytes generated from each application to the total number of bytes).[3] The application classification is based on network-level and transport-level protocol header information. We can notice that:

- Web applications are the most used ones in terms of number of clients, attracting around 46% of the usages. This happens as people increasingly use social networks (e.g., Twitter, Facebook), using HTTP. However the traffic generated from these applications is not the majority (i.e., 26% of the total traffic) because they do not require much bandwidth.

- The trend toward social communications is also confirmed by the large usage of chat and messaging applications covering together around 17% of the usages. However, as seen in Figure 3.3(b), this sort of applications has a negligible contribution to the total traffic since their required bandwidth is very limited.

- A third class of applications is the one of bulk transfers, file and mail download applications (e.g., relying on the POP, SMTP, FTP protocols), which cover around 26% of the usages and contribute in 24% of the total traffic.

- Real-time applications, video streaming, gaming and VoIP, represent only 7% of the usages, they are the most bandwidth consuming applications (i.e., contributing in 50% of the total traffic) as shown in Figure 3.3(b).

We can clearly affirm that there are important differences in the geographical location and application distribution of content consumption spots. As evidenced, their location can temporarily change. Moreover, a large majority of the traffic volume is related to bulk transfers and web services whose content could be partially cached or whose Cloud server could be located closer to the crowded spots. Users could indeed be better served if delivery facility could be located closely to the dense cells. The backhauling network performance could also benefit from traffic

[3]It is worth mentioning that a user connected to two different applications at the same time, is counted in both applications.

offloading and traffic engineering techniques aware of user mobility and content consumption point deflections.

3.4 Crowded Spot Estimators

In this section, after reviewing the state of the art on cell load estimation, we present the two possible crowded spot estimation techniques.

3.4.1 Related Work on Cell Load Estimation

A limited amount of work exists in the literature on the estimation of crowded spots and rendez-vous points in access networks. E.g., in [51] vehicular data is exploited to determine accident-risk points. Authors in [52] propose a framework that discovers regions of different functions in a city using both human mobility among regions and points of interests located in a region. Many other works, such as [38], [53], [54] and [55], while assuming the availability of mobility information, focus on user-profile aware QoS provisioning, load balancing and network signaling improvement techniques.

Traffic load forecasting has also been investigated from an analytical and mathematical modeling perspective. For example, authors in [56] show how under certain conditions periodic sinusoidal functions can be used as cellular traffic profile. Unfortunately, the simplicity and the too theoretical properties of these approaches fail from precisely matching with the actual real traffic load, which is a strict requirement of our investigation. In the same context of modeling the spatial cell density of a mobile network, authors in [57] propose a stochastic model to compute the probability of staying in a given location for a given period of time as well as the probability of moving from a given location to another one, using a random waypoint-based mobility pattern. Also, relevant works targeting mobility pattern detection from real cellular network data have been studied e.g., [19] [30] [31].

3.4.2 Trajectory-based Crowded Spot Estimator

Motivated by the usage of Airsage signaling mobile phone data (please refer to Section 2.2 for more details about data structure) that give real trajectories of smartphone users, we extract the real positions of crowded spots and compare them with the estimated positions that one can get by applying the interpolation methods defined in Section 2.4. Decomposing the state of Massachusetts into census blocks[4] [58], we compute the real load of each block in the region (i.e., expressed as the users' number of visits to each block) as shown in Figure 3.4.

The small map in the upper right corner is a zoom in of the Boston urban area, the state's largest city where small blocks exist. The figure clearly shows the load difference among the blocks and the existence of crowded blocks that define the most visited places where large masses of people usually visit.

We estimate the load of each of these blocks by choosing for each user category the best interpolation method obtained from results of Section 2.6.1 (i.e. for seden-

[4] A census block is the smallest geographic unit used by the United States Census Bureau. Blocks are typically bounded by streets, roads or creeks.

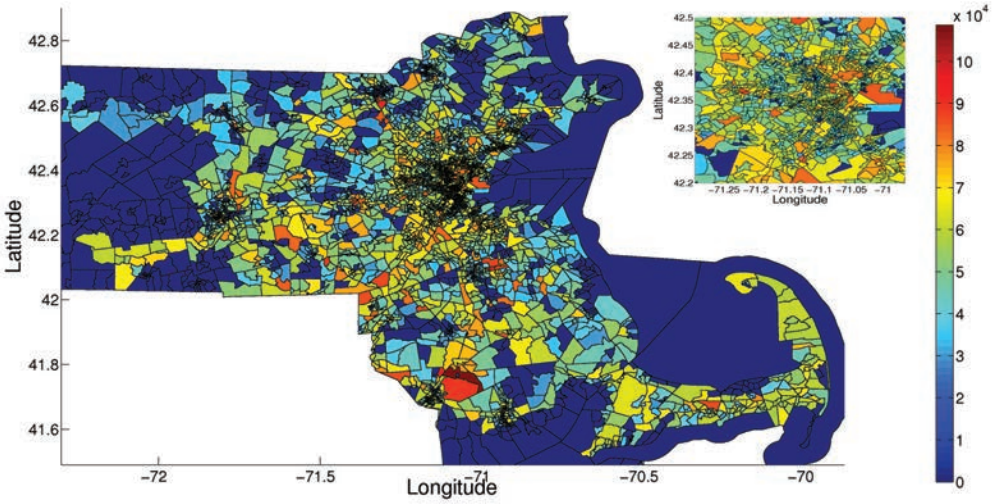


Figure 3.4: Real block load.

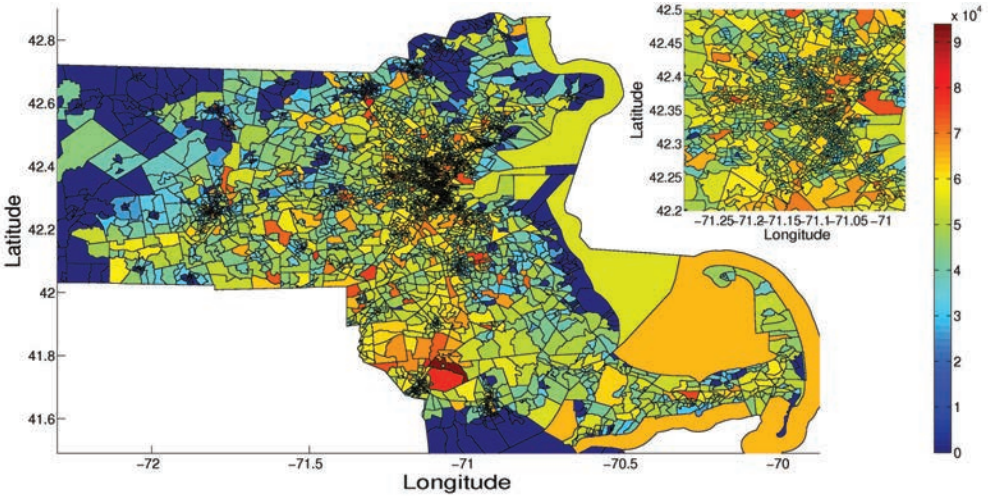


Figure 3.5: Estimated block load.

tary and urban mobile users, we use the linear interpolation method to join the samples, while for peri-urban mobile users and the commuters we follow the cubic interpolation).

Evaluation

The results of the trajectory-based estimator are presented in Figure 3.5. We notice that the load is overestimated especially for the less crowded blocks. What about the crowded blocks? How does the estimation error vary in the most crowded places?

Figure 3.6 represents the variation of the load estimation error with respect to the real load. We can state that:

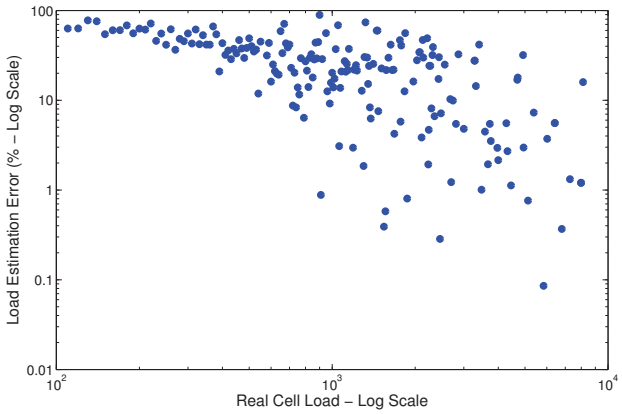


Figure 3.6: Load estimation error.

- The estimation error is very high for the less visited blocks in the region.
- The estimation error rapidly decreases with the increase of the real load.
- For the most crowded blocks, we notice that the estimation error is significantly smaller.

By choosing different thresholds beyond which we identify the crowded blocks (i.e., if a block has a load, expressed by the total number of users' visits during the day, that exceeds the chosen *Crowd Selection Threshold, CST*, it is considered as a crowded block), we note that a CST of x^{th} percentile refers to the selection of blocks having a load higher than the load of $x\%$ of the blocks. We plot for each case the cumulative distribution function of the block load estimation error. The results are shown in Figure 3.7. We can state that:

- The median estimation error decreases with the increase of CST.
- The median estimation error is equal to 37% if we select all the blocks; while for those having a load higher than the load of 50^{th} of blocks, the median estimation error reaches 25%. With stricter definition of CST to 70^{th}, the median reaches 10% and finally by selecting the blocks having a load higher than 90^{th} of blocks we get a median estimation error of 5.5%.

As a conclusion we can clearly confirm that using the trajectory interpolation methods as a crowded spot estimation method, we can obtain very high accuracy.

Implementation Complexity

We should note here that the "trajectory-based crowded spot estimator" is scalable in the sense that taking a sample of users instead of the whole population enables us to find the crowded spot positions in a relatively accurate way. For each new user position, Equations (2.1) and (2.2) are updated handling only a limited arbitrary number N of last positions during a time interval T. Hence the user's category

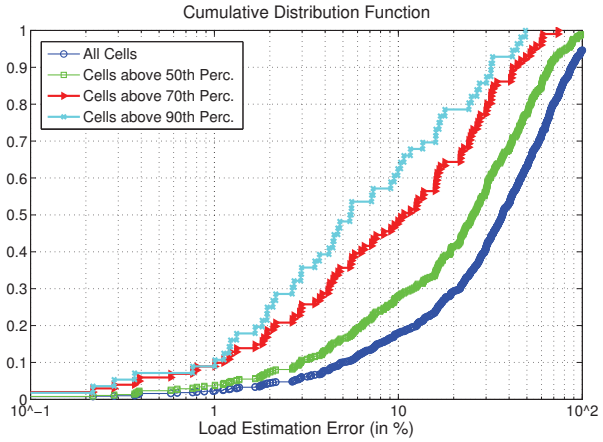


Figure 3.7: Cumulative Distribution Functions of the load estimation error for different CSTs.

as well as the best interpolation method suitable for that user change. All these operations have a $O(N * |U_m(T)|)$ time complexity, $U_m(T) \subset U$ being the subset of users moving in the time interval T.

3.4.3 Territory-based Crowded Spot Estimator

Alternatively to the trajectory-based crowded spot estimator, we investigated other options not dependent on the trajectory but rather on the characterization of users' territory. In the following, we use the centroid and the radius of gyration to account for user's spatial coverage, parameters already defined in Section 2.3. Indeed, the 'territory' of a user can be defined as the area covered by the circle centered at the user's centroid having a radius equal to the user's radius of gyration.

This second estimator we propose, takes into account the non-negligible intersections of different users' territories to estimate the cellular user density. For example, in Figure 3.8, based on the centroid and radius of gyration of the two users, it is reasonable to account for the possibility that the two users pass by site A. As already argued, the efficiency of the estimator should be evaluated toward its capability of estimating crowded cells rather than lightly loaded cells.

Let $m[r_g^u(t), \vec{r}_{cm}^{\,u}, c, u, t]$ be a spatial mapping counter equal to 1 if the circle of user u, with radius $r_g^u(t)$ around the centroid $\vec{r}_{cm}^{\,u}$ at time t, covers at least 10% of cell c, 0 otherwise (i.e., the 10% is actually 10% of the area of each cell); other thresholds than 10% could certainly be considered, depending also on the way the environment is architected, its aim being to avoid the small overlaps since not all the cells covered by one user's radius will reasonably be visited.

Such a spatial mapping counter is the core metric of our estimator. Simply counting the number of intersections significantly covering a given cell would certainly lead to an over counting that needs to be appropriately scaled. A simple scaling could be, e.g., to divide it by the average number of users per cell during the past measurements, yet this does not prevent from high deviations. We propose to scale it by the scale factor that would generate null estimation error in an arbi-

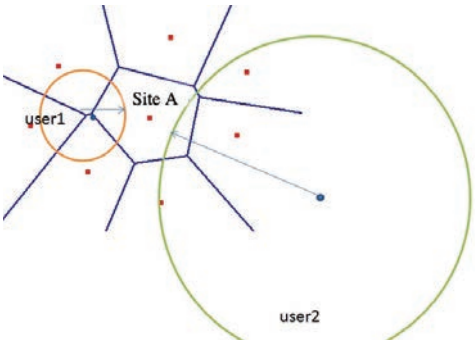


Figure 3.8: Illustration of the territory-based estimator.

trary instant $t - RTW$ in the recent past, with a regression time window (RTW) adequately set.

Then, as an estimator of the number of users per cell c, we propose:

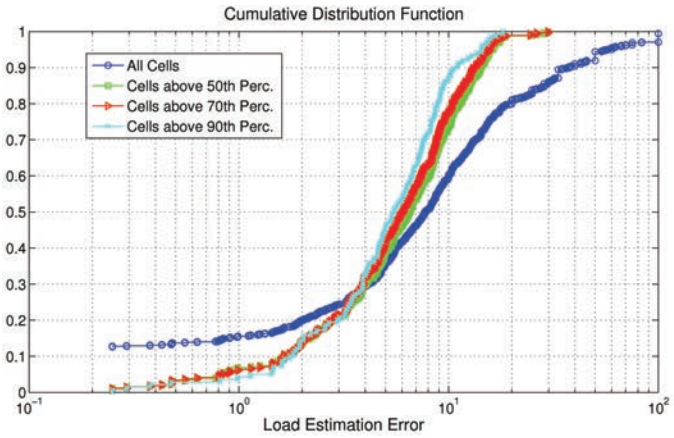
$$\hat{e}(c,t) = \frac{\sum_{u \in U} m[r_g^u(t), \vec{r}_{cm}^u, c, u, t]}{\sum_{u \in U} m[r_g^u(t - RTW), \vec{r}_{cm}^u, c, u, t - RTW]} * n(c, t - RTW) \tag{3.1}$$

It uses therefore a linear regression over past measurements to adequately weight the spatial mapping counter. Intuitively, the smaller RTW is, the closer $\hat{e}(c,t)$ is to the unknown $n(c,t)$, i.e., the more accurate the estimation is. However, RTW should be big enough to allow network management system to retrieve the real number of users in cell c at time $t - RTW$, i.e., $n(c, t - RTW)$. Depending on network management tickets, session duration, network latency and network size, this parameter RTW could range from a few dozens of minutes to a few hours.

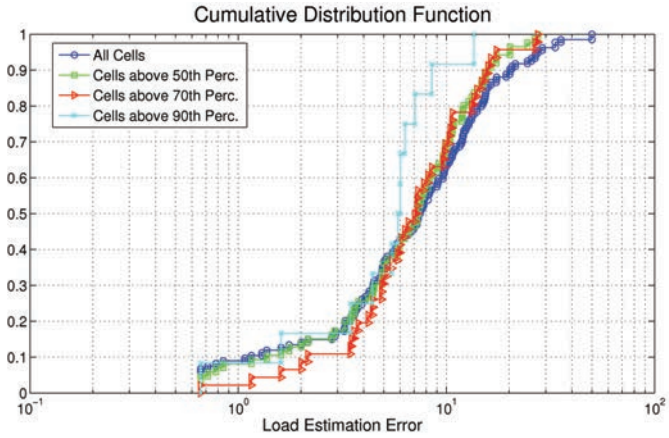
Evaluation

We evaluate our estimator using the available network management dataset of the Orange France cellular network (i.e., please refer to Section 3.2 for more details on data structure) for the new year's eve, by computing centroid and radius of gyration of all users passing by the three regions: Paris, Lyon and Nice. We set a RTW to 1 hour, which is a quite pessimistic value (in practice, in carrier grade networks, it could even be set to a few minutes, hence allowing a higher accuracy; later we show the influence of varying the RTW on the estimator's performance). We consider the user position samples from the whole day in the computation of users' centroids and radius of gyration.

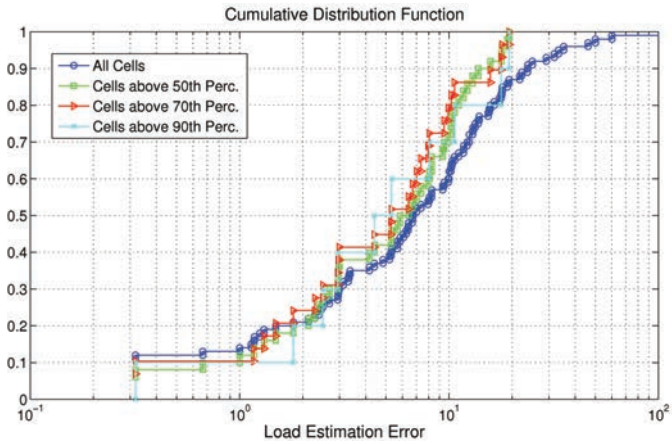
To qualify the accuracy of the proposed crowded spot detection method, we analyze the estimation error (i.e., the difference between the real and the estimated number of users, to the real number of users) for all the cells in the Paris, Lyon and Nice metropolitan area network. Figure 3.9 depicts the CDFs of the estimation error in the three considered cities for different possible load thresholds CSTs. The CDFs show that including all cells, and for all the considered regions, the median estimation error is lower than 8%. Including only cells having a CST of 70% (i.e., having a load higher than the load of 70% of the cells), the median estimation error is always inferior to 7%. With stricter definitions of CST (90% percentile),



(a) Paris

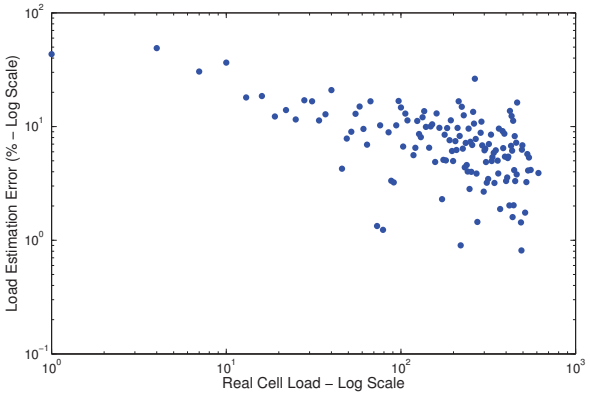


(b) Lyon

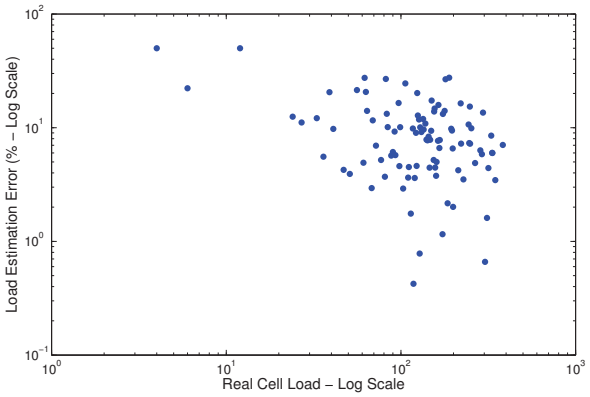


(c) Nice

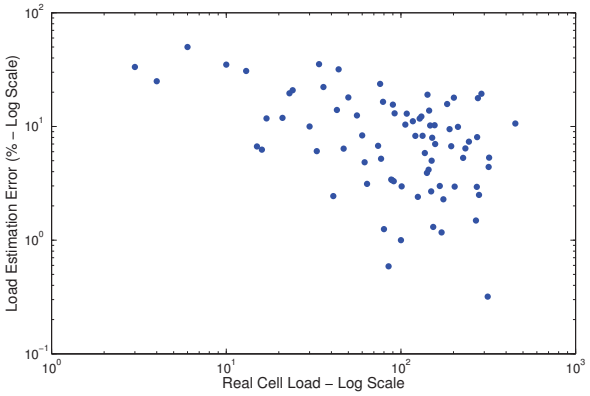
Figure 3.9: Estimation error distribution in the three regions.



(a) Paris



(b) Lyon



(c) Nice

Figure 3.10: Load estimation error in different regions.

the median estimation error becomes inferior to 6%. It is worth stressing that the estimator's performances seem to be city-independent because it shows similar

results in different regions.

Figure 3.10 represents the log-log scatter representations of the estimation error as a function of the real number of users per cell for each region. We can observe that, for all regions, the estimation error is high for low-loaded cells and it decreases more than linearly as the real cell load increases, which is, as already argued, a very desirable behavior.

Implementation complexity

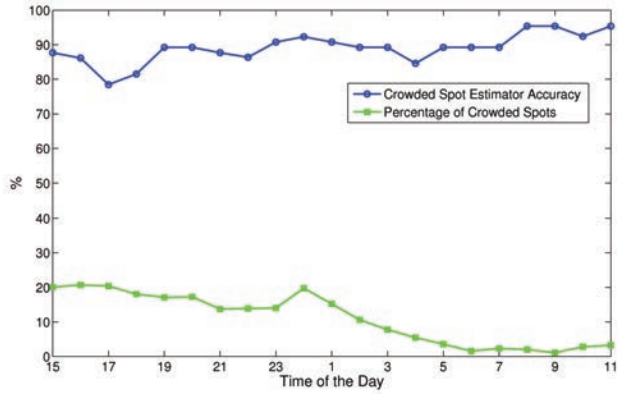
With our estimator, for each new user position, Equations (2.1) and (2.2) could be updated handling only a limited arbitrary number N of last positions, then Equation (3.1) can be updated. Hence, all these operations have a $O(N * |U_m(t)|)$ time complexity, $U_m(t) \subset U$ being the subset of users moving and sampled in the sampling interval $(t - RTW, t)$. Thus from an implementation complexity, both trajectory-based and territory-based crowded spot estimators have the same complexity.

As already mentioned, the regression time window RTW should be chosen so that it is sufficiently higher than the computation time of the estimator. However from a more practical perspective, in very high mobility environments, the above complexity could become quadratic, which may raise scalability concerns. Scalability concerns could also rise from the volume of the data to mine in order to extract estimator's metrics. The larger the sample temporal window is, the larger the data volume is and the higher the computation complexity is. Hence the last questions we want to answer are: does the crowded spot detection accuracy decrease as a function of time? What happens at different times of the day? In order to answer these questions, we test our estimator in the same period interval (between 3 pm of December 31, 2013 till 11 am of the January 1, 2014) in the three selected regions. We note that we do not have access to data of January 1, after 11 am.

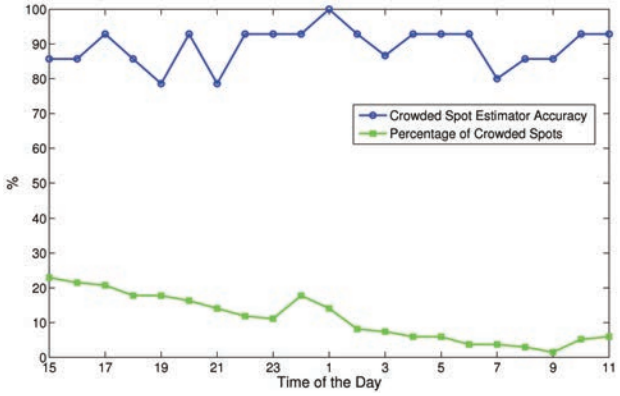
Figure 3.11 shows our evaluation results with a RTW of 1 hour and a step of 1 hour starting by 3 pm for the three considered regions. For example, when performing the crowded spot estimation at 9 pm, we exploit the data collected from 3 pm to 9 pm. These results highlight two important aspects: firstly, the accuracy of our estimator is time-independent and thus mobility-independent because it shows a constant behavior at different times of the day. Another important aspect is that the percentage of crowded spots decreases when the overall user mobility decreases, and vice versa. The percentage of crowded spots is thus mobility-dependent; the maximum percentage is obtained at midnight (19% for Paris, 17% for Lyon and 15% for Nice) when likely people gather together to celebrate the new year's eve, then this ratio decreases when people start to split up and return back to their home locations.

As a last analysis, we look at the influence of the regression time window on the accuracy of the estimator. For simplicity, and in order to minimize the computation time, we select 5% of the most active users in the regions and we compute the estimation error by varying the RTW as shown in Figure 3.12.

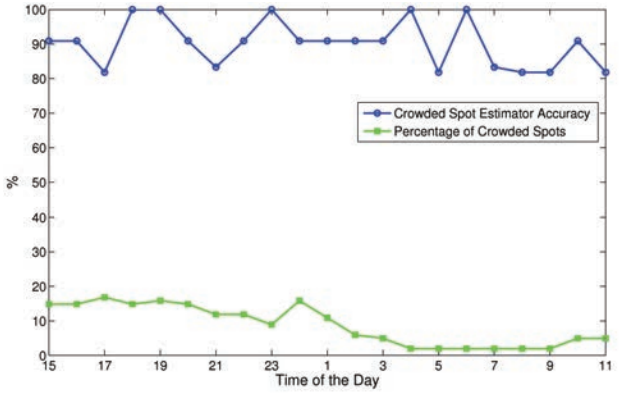
We clearly notice that the estimator's accuracy increases with smaller RTW (i.e., in Paris for example, the median error is 18% when RTW=1 hour, 12% for RTW=48 min and less than 5% with a RTW=12 min; we have similar results also for the other two regions).



(a) Paris



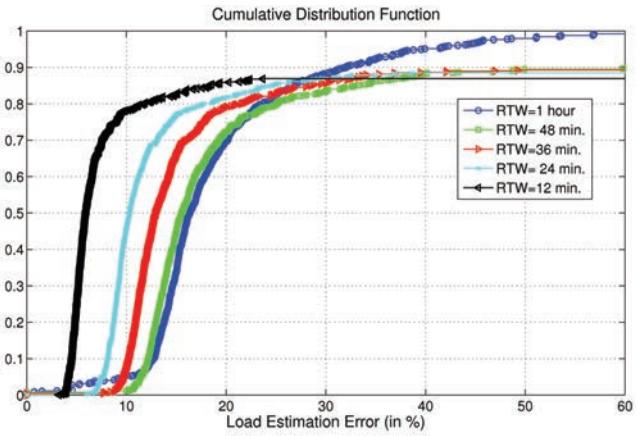
(b) Lyon



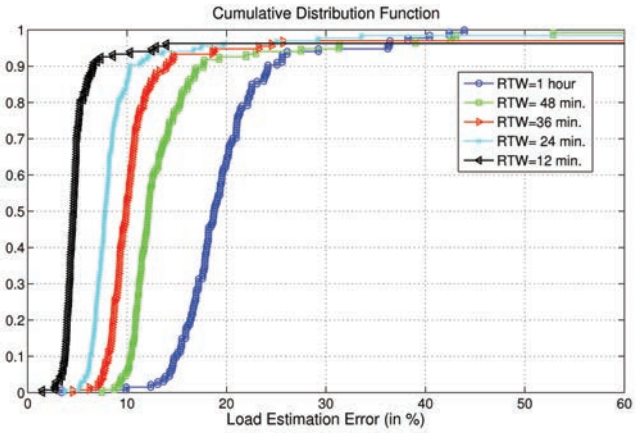
(c) Nice

Figure 3.11: Crowded spot estimator accuracy and percentage of crowded spots as a function of time in different regions.

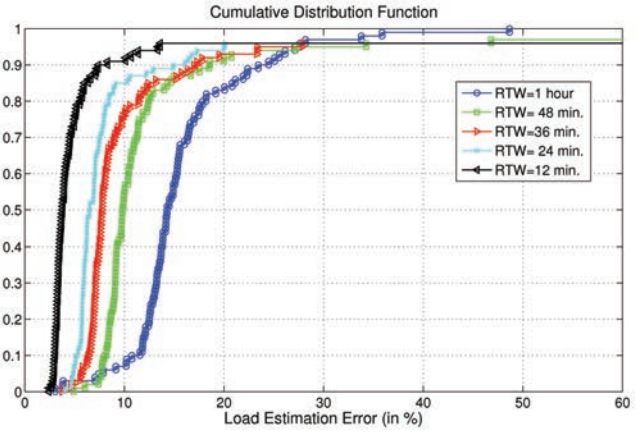
As a summary, we can affirm that the territory-based crowded spot estimator, evaluated on real data for the three cities, appears as an excellent crowd cell detection solution of cellular and backhauling network management. Its error strictly



(a) Paris



(b) Lyon



(c) Nice

Figure 3.12: CDF of the load estimation error for different RTWs.

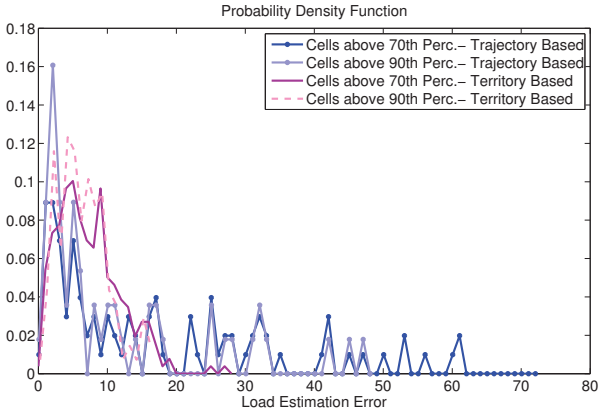


Figure 3.13: Probability Density Function of the load estimation error for both estimators and for different CSTs.

decreases with the cell load, and it becomes very small for highly loaded cells. It is also quite scalable against mobility data volume and against time variations.

3.4.4 Comparison between Estimators

The above-evaluated estimators have shown quite close median load estimation errors especially for high CST values; however the advantages of one estimator over the other are still not straightforward. In this section, we conduct a comparative study of the behavior of both estimators: Trajectory-based and Territory-based crowded spot estimators. Figure 3.13 shows the probability density function of the load estimation error for both estimators and for different CST values (i.e., 70% and 90%). It is easy to notice that the trajectory-based estimator has the highest load estimation errors, for both values of CST, with respect to the territory-based one (i.e., for the trajectory-based estimator, some cells have an estimation error higher than 30% while for the territory-based estimator, all cells have an error lower than 30%). So as a conclusion, we can affirm that the territory-based estimator seems more suitable and more accurate in determining the load of crowded spots than the trajectory-based estimator.

3.5 Summary

A present matter in mobile access network management is to find scalable and effective ways to offload the backhauling and cellular networks as a function of user mobility and consumption patterns. The motivation is the arising weight of mobile Internet traffic taken over legacy wireline access traffic in todays Internet Service Provider networks. Different offloading techniques are currently studied, to cope with this evolution, some of which are presented in the next chapters. In this scope, the contribution of this chapter is twofold.

Firstly, working on real network probe data from the Orange cellular network, we show how content consumption spots spatially move as a function of the occurrence of special events. We provide useful user, traffic volume and traffic type experimental distributions to the research community.

Secondly, motivated by the experimental findings, we propose two crowded spot estimation techniques based on two compact and easy-to-compute user mobility metrics, i.e. the user's centroid and radius of gyration. Results on real data show that the error of our estimators decreases with the cell load and that, for the crowded spots, the estimation errors are quite small.

Upon the determination of highly loaded cells and in order to satisfy users demand in those cells, while ensuring a good quality of service, our interest moves now to find suitable solutions to cater the high traffic volume in those crowded cells. This is in fact the main goal of the next part of this dissertation.

Part II

Resource Contention and Offloading Solutions

Chapter 4

Horizontal Traffic Offloading with Small-Cell Networks: Cooperative Resource Allocation Approaches

The increasing need of offloading solutions is caused by the explosion of Internet data traffic, especially the growing portion of traffic going through mobile networks. In order to meet mobile Internet demand while addressing the lack of available mobile spectrum and the expense of new infrastructure, service providers are severely challenged. They need to master the needed capacity expansion in their access network; otherwise the data traffic will sooner or later clog their networks. Next-generation network deployments promise to deliver higher bandwidth and speed, but they often imply high capital and operational expenditures.

An alternative economically and technically viable way is represented by mobile offloading solutions. Such solutions aim to optimize the resource utilization, to reduce the traffic on operator's licensed spectrum, and to lower the traffic load on base stations. Wi-Fi and small-cell technologies are considered the primary offloading technologies used today by the industry stakeholders. In this chapter, we tackle mobile traffic offloading over small-cell networks, while the next chapter is dedicated for the study of mobile data traffic offloading over Wi-Fi access points. In particular, in this chapter, we propose a strategic self-organization solution among small-cell networks (i.e, femtocell networks in our study) to cooperatively manage interference and resource allocation in urban crowded environments.[1]

4.1 Introduction

The femtocell technology [59] [60], a promising small-cells technology, aims to offer better indoor voice and data services for cellular networks via the deployment of tiny cellular repeaters, differently backhauled and synchronized. Instead of redimensioning macro-cells at the base station level, the modular installation of short-range and little mobile access points can grant multiple benefits [61] as explained in Ta-

[1]The contents of this chapter are presented in [2], [8], and [9].

Table 4.1: Major benefits and advantages of femtocell deployment.

Technical Aspects	**Material Aspects**
improve throughput	extremely simple installation
improve indoor coverage	very low equipment cost
offload outdoor cells	hardware already in the market
handset compatibility	energy savings

ble 4.1. Strategically, femtocells are attractive solutions because they are easy to install, inexpensive, with no real-estate cost, and with already present handset devices and hardware components. Moreover, they can grant energy saving, especially to the user's terminal, thanks to a lower required power. Technically, femtocells can drastically increase the download capacity, with a much higher throughput, while offloading the macro-cells.

Femtocells work on the same licensed spectrum as the macro-cells of cellular networks and thus do not require special hardware support on mobile phones, thus simplifying data offloading procedures [62]. However, despite the benefits of femtocells networks in offloading data traffic via horizontal handovers from macro to femto cells and vice versa, one should not forget the inherent constraints of such networks due to cross-tier[2] and co-tier[3] interferences that should be taken into account when installing femtocells [64], especially in suburban and urban environments. Under certain assumptions, cross-layer interference with the macro-cell is manageable, while co-layer interference among femtocells requires forms of explicit coordination and signaling among neighboring femtocells.

In this context, we can refer to these networks as collaborative femtocell networks since coordination or cooperation mechanisms are needed between independent and opportunistic femtocells to manage reciprocal interferences and resource allocation. The independence of femtocells resides in the fact that the installation of a femtocell for residential or enterprise usages is expected to be subject to separate billing, while the opportunistic behaviour can be motivated by the attempt of each femtocell to satisfy its users, by acquiring the maximum number of resources. Therefore, inter-femto resource allocation needs to be managed via collaborative approaches that have as motivation the performance improvement for all the participating femtocells.

In this chapter, we envision a strategic resource allocation among independent subscribers in collaborative femtocell networks. This is especially needed in dense urban environments, with a high density of femtocells, and where femtocells have different levels of interference and resource demands, and the overall demand exceeds the available bandwidth. The motivation behind our approach is that femtocell's interference level and demands volume, femtocell subscriber independence from other subscribers, as well as opportunistic behaviour of those femtocells, should all be factors taken into account when allocating resources to users. We model such situa-

[2]The cross-tier interference [63] is defined as the decrease in signal quality of macro-cell users due to the presence of femto users sharing the same spectrum and vice versa.

[3]The co-tier interference occurs when all femtocells share the same spectrum.

tions using cooperative game theory, whose principles are resumed in Appendix A, which guarantees that interference management and resource allocation solutions are strategically and rationally justified. We show that our approach grants important improvements in throughput and fairness.

4.2 Related Work on Resource Allocation in Femtocell Networks

Resource allocation in OFDMA femtocell networks has been considered in recent research works. The general objective of these works is the computation of efficient allocation of time-frequency resource slots, called 'tiles', while accounting for cross-layer interference (interference between macro-cell and femtocell users) and co-layer interference (interference between femtocells' users) [65]. In the following, we discuss a selection of relevant approaches proposed to solve these types of interferences.

Authors in [65] present an overview of possible approaches to manage cross-layer and co-layer interferences. Two main directions are outlined: shared spectrum and split-spectrum schemes. In the first, co-channel assignment mitigates the capacity problems of both femtocells and macro-cells thanks to the use of larger spectrum, but cross-layer interference needs to be managed. A proposal in this direction is [66], where the authors propose a distributed utility-based non cooperative SINR adaptation approach for femtocells to alleviate the cross-tier interference. Their solution relies on a channel-dependent SINR equilibrium at each femtocell that strongly discourages interferer femtocells to use large transmit powers. In the second scheme, an orthogonal channel assignment eliminates cross-layer interference by dividing the spectrum into two independent fragments, one used by the macro-cells and the other by the femtocells. For both cases of shared spectrum and split-spectrum, the authors in [65] outline the requirements of centralized approaches, called C-DFP (Centralized−Dynamic Frequency Planning), and distributed approaches, called D-DFP (Distributed−Dynamic Frequency Planning). For C-DFP, a subchannel broker receives demands and interference information from the femtocells and/or the macro-cells, so as to compute the best resource allocation considering a tradeoff between optimality and computational complexity. For D-DFP, each femtocell, or cluster of femtocells, independently creates subchannel priority lists starting by the least interference subchannel and periodically allocates the resources in a distributed fashion. Some of the relevant works in this direction are discussed below.

4.2.1 Split-Spectrum Distributed Schemes

Split-spectrum schemes appear particularly interesting for urban environments with a dense deployment of femtocells since they offer a good tradeoff between spectral efficiency and network management simplicity. In this chapter, we propose a split-spectrum approach. A number of related proposals have been made in the literature.

In [67], the authors show that with a dynamic spectrum splitting among femtocell and macro-cell, the area spectral efficiency can be optimized. In their framework, within the femtocell layer, an adapted version of ALOHA (for the time-frequency domain, called Frequency ALOHA, F-ALOHA is proposed to schedule the femtocell access to the co-tier shared spectrum. However, the pseudo-random characteristic

of this method can generate irrational resource allocations, because they are not strategically computed.

In [68], each femtocell determines its set of interferer neighbors and uses a hashing scheme in order to determine its allocation in a fully distributed manner. Then, for collided tiles, rehashing is performed again in order to solve contention in the subsequent frames. A fractional frequency reuse technique that adjusts the frequency reuse factor to alleviate inter-femto interference is presented in [69]. In this case, femtocells are grouped depending on the amount of reciprocal interference by a femtocell gateway that determines the minimum number of orthogonal subchannels for each group, and adjusts the transmit power of each femtocell based on the received signal strength.

A similar approach is proposed in [70], where to exploit cooperation among neighboring femtocells and to improve resource allocation and throughput satisfaction, a hybrid centralized/distributed approach is proposed. First, femtocells are grouped in a distributed fashion into disjoint clusters with respect to interference maps. Then, within each cluster, resource allocation is centralized at a cluster-head that periodically optimizes the throughput satisfaction. Finally, possible resource contention among neighboring femtocells is locally solved.

4.2.2 Game-Theoretic Approaches

Extending the femtocell cooperation assumption made in [69] and [70], it appears appealing to formally model rational strategies among independent femtocells or groups of femtocells. As already mentioned, when accounting for strategic interactions and network agents independency, it is appropriate to adopt game-theoretic approaches, as practical solutions or as benchmarks. Recently, there has been significant interest in applying game theory to the analysis of collaborative communication networks, with the aim to identify rational strategic solutions for multiple decision-maker situations. Indeed, as opposed to mono-decision maker approaches, game-theoretic approaches adopt a multi-agent perspective to account for different objective functions and/or counter objections to rationally non justified solutions [71].

When the collaboration among network agents does not imply binding agreements and need just coordination, non-cooperative game theory can identify strategic solutions as a function of various types of game equilibrium [72]. For example, the already-mentioned work in [66] relies on game equilibrium. Similarly, the authors in [73] propose a solution for resource allocation in K-user fading multiple access channels (MAC) based on a Bayesian game modeling, with incomplete information about the fading channel gains and unilateral utility functions.

When binding agreements are required to motivate cooperation, coalitional (cooperative) game theory allows solutions with the desirable properties of efficiency and rationality (among others) [74]. Specifically, the authors in [75] model the femtocell spectrum sharing problem as a coalitional game in partition function form [76] [77], using signal-to-noise ratio utility functions. In this approach, femtocell networks are topologically partitioned into disjoint clusters so that the overall performance is increased, outperforming selfish solutions where femtocells access a shared spectrum in an uncoordinated manner. Then, within a network partition, a

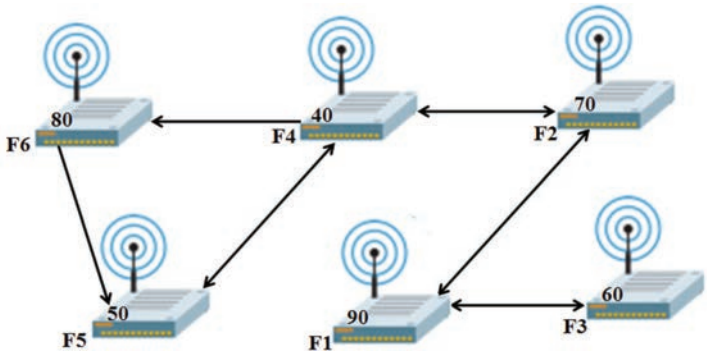


Figure 4.1: Example of a 6-femtocell network.

joint online scheduling is adopted among cooperating femtocells [78].

In this chapter, while taking similar hypothesis on the independency of femtocells, we focus on related yet different issues. Differently than in [75], our objective is to define cyclic spectrum allocation rules. Rather than partitioning the femtocell network topology in disjoint clusters as in [69] [70] [75], we allow femtocells negotiate resources in multiple femtocell groups, where groups are locally detected as function of interferer femtocell neighbors. As in dense urban environments, joint online scheduling among groups of femtocells as [78] may be counterproductive in terms of signaling overhead, we target a solution in which the resource allocation is periodically pre-computed, based on changing femtocell resource demands and interference maps. In particular, we consider dense environment situations in which the overall demand is quite often higher than the available bandwidth on the shared media, which mathematically corresponds to a bankruptcy game situation [79] (please refer to Appendix A for more details on bankruptcy game and coalitional game modeling), representable in canonical form [77]. We investigate two solution concepts: the well-known Shapley value [80] (already adopted in a variety of situations in networking such as inter-domain routing [81] and network security [82]), and the less-known Nucleolus [83] (used, for instance, in strategic transmission computation [84] [77]), which shows additional interesting properties in bankruptcy situations.

4.3 Context and Problem Formulation

We consider a network composed of a macro-cell area with several femtocell access points (FAPs) that represent residential or enterprise networks. As in [65] and [67], we assume an orthogonal channel assignment that eliminates the cross-layer interference. The femtocells and the macro-cell are assumed to operate using the OFDMA technology (e.g., WiMAX or LTE) whose frame structure can be viewed as time-frequency slots, also called tiles. A certain number of users attach to each femtocell; user demands represent the required bandwidth, expressed in number of required tiles.

As already mentioned, in urban dense environment, we expect that the sum of demands of the femtocells often exceeds the available resources. Therefore, our objective is to find, for such congestion situations, a strategic resource allocation that satisfies throughput expectations while controlling the interference between

femto-femto users.

In the following, we first present the corresponding optimization problem, then, we highlight possible alternative solutions, and finally describe the properties of bankruptcy games along with possible solutions.

4.3.1 Notations

Let $\mathcal{F}$ be the set of femtocells in the network, d_i the demand of $F_i \in \mathcal{F}$, and x_i the number of allocated resources to F_i. Also, let $\mathcal{I}_i$ be the interference set of F_i, which corresponds to the set of femtocells composed of F_i and the femtocells causing interference to F_i. This interference set is computed based on the SINR, the path loss model[4] given in the A1 scenario for indoor small office and residential of WINNER for the frequency range 2−6 GHz [85]. It is worth noting that interference is not symmetric since it depends on user positions, and therefore $\mathcal{I}_i \in \mathcal{P}(\mathcal{F})$, where $\mathcal{P}(\mathcal{F})$ is the set of different subsets of femtocells in the network. We note here that for all $i \neq j$, I_i and I_j do not form necessarily two disjoint subsets.

For example, consider the situation depicted in Figure 4.1, with six femtocells where arrow's direction indicates the femtocell whose users suffer from a neighboring femtocell interference. The interference relationships are reported in Table 4.2.

Table 4.2: Interference relationships.

Femtocell	**Interferers**
F_1	$\{F_2, F_3\}$
F_2	$\{F_1, F_4\}$
F_3	$\{F_1\}$
F_4	$\{F_2, F_5\}$
F_5	$\{F_4, F_6\}$
F_6	$\{F_4\}$

4.3.2 Related Centralized Optimization Problem

For the sake of clarity, we model here the resource allocation problem as a centralized mono decision-maker optimization problem, i.e., as the C-DFP approaches mentioned in Section 4.2. The problem can be formulated as:

$$\begin{aligned} \text{objective} \quad & f(d_i, x_i) \\ \text{subject to} \quad & 0 \leq x_i \leq d_i, \ \forall F_i \in \mathcal{F} \\ & \sum_{j|F_j \in \mathcal{I}_i} x_j \leq E, \ \forall \mathcal{I}_i \\ & x_i \in \mathcal{Z}^+, \ \forall F_i \in \mathcal{F} \end{aligned}$$

[4]The path loss model depends on the number of floors, traversed walls and users' positions.

where E is the number of tiles in an OFDMA frame (also referred to in the following as 'estate'). The objective typically depends on the demand and the allocated resources; in our case it is the minimization of the maximum gap between demand and allocation, $\min \max_i(\frac{d_i - x_i}{d_i})$. The constraints are integrity constraints, on the allocated tiles to individual femtocells and to femtocells belonging to same interference sets. Later, we compare our approaches to this C-DFP solution highlighting the interest in strategic approaches and stressing the tradeoffs between them.

4.3.3 Possible Distributed Approaches

For each interference set, we have therefore a situation in which a group of femtocells can either: (i) randomly access the spectrum hoping that collision will not occur (e.g., as in F-ALOHA [67] which is an application of ALOHA in time-frequency domain); or (ii) self-organize to define an online joint scheduling (as in [78]); or (iii) divide the available spectrum proportionally, (iv) rationally adapt the allocation to each femtocell claim and interference situation.

Clearly, (i) excludes any form of coordination and would favor opportunistic wealth-aversive behaviors (e.g., setting a minimum waiting time upon collision in F-ALOHA) that other femtocells cannot control. Approaches like (ii) risk to generate enormous signaling for large interference sets (likely in urban dense environments). Under (iii), inefficiency can arise whether many demands are less than the proportional share, and a weighted proportional share would favor cheating demands (higher claims than what is really needed).

The path forward is therefore towards cooperative approaches that dissuade malicious behaviors in setting demands, under an adequate binding agreement fixing common rules on shared information and allocation scheme. Before detailing our algorithmic approach, let us introduce the bankruptcy game that can model interactions among femtocells belonging to the same interference set.

4.3.4 Bankruptcy Game Modeling

In urban environments, a dense deployment of femtocells is expected, so that situations in which the overall resource tile claim (i.e., sum of the demands) surpasses the number of available tiles (E) in the shared spectrum are likely. Assuming that femtocells belonging to the same interference set, share information about respective demands, the interaction can be modeled as a cooperative coalitional game. The game-theoretical model we propose is composed of two main phases:

1. Interference Set Detection phase.
2. Bankruptcy Game Iteration phase.

Formally, our proposed game represents a binding agreement between cooperating femtocell subscribers.

Interference Set Detection

Upon each significant change in demands or in network topology (femtocells can be freely moved inside a residential space or office), each femtocell determines the set of

interferer femtocells (i.e., the femtocells that cause interference to its users) based on the minimum required Signal to Interference plus Noise Ratio (SINR).[5] Femtocells are able to share their interference set with other femtocells in the network.[6]

Next, the list of interference sets are sorted, firstly with respect to their cardinality, and secondly with respect to the overall demands, both in a decreasing fashion (i.e., first the largest sets with highest overall demands).

Bankruptcy Game Iteration

In the second phase, resources are eventually allocated, proceeding by solving a cooperative coalitional game (see Appendix A) for each interference set, following the order in the sorted list from the first phase. Note that, since a femtocell can belong to many interference sets, if it has already participated to a game in a previous game iteration, it is excluded from the next game iteration in which it appears. Each game iteration therefore includes only the femtocells for which an allocation has not been computed yet. This corresponds in iterating a game differing in that:

- $\mathcal{N}$ includes only the unallocated femtocells in the set;
- the estate E is decreased by the amount already allocated to the set's femtocells.

In order to solve this game we need to define its characteristic function, that represents the profit attributed to each coalition of players. In the most pragmatic case, each coalition of players is able to share what the other femtocells have left after getting what they claimed. That is, $E - \sum_{i \in \mathcal{N} \setminus S} d_i$, where $\mathcal{N} \equiv \mathcal{I}_i$. Such a characteristic function corresponds, in fact, to what is known as 'bankruptcy game' precisely defined hereafter.

Definition 4.3.1. A bankruptcy situation is defined by a pair (E, d) where $E \geq 0$ is an estate that has to be divided among the members of $\mathcal{N}$ (the claimants) and $d \in R_+^{|\mathcal{N}|}$ is the claim vector such that:

$$E < \sum_{i \in N} d_i \tag{4.1}$$

Definition 4.3.2. A bankruptcy game [79] is defined as $G(\mathcal{N}, v)$ where $\mathcal{N}$ represents the claimants of the bankruptcy situation and v is the characteristic function that associates to each coalition its worth, defined as the part of the estate not claimed by its complement:

$$v(S) = max(0, E - \sum_{i \in \mathcal{N} \setminus S} d_i) \, , \forall S \subseteq \mathcal{N} \setminus \{\emptyset\} \tag{4.2}$$

[5] In LTE networks, user feedback reports can include interferer femtocell identifiers (Physical Cell Identity) [86].

[6] In LTE networks, this can be aggregated at, or relayed by, Home-enhanced Node B (i.e., femtocell) gateways (i.e., HeNB-GW) [87].

Equation (4.2) has been proven to be superadditive [71]. Moreover, it satisfies the supermodularity property [80] [88], stronger than the superadditivity, which means that the marginal utility of increasing a player's strategy rises with the increase in other player strategies:

$$v(S_1 \cup S_2) + v(S_1 \cap S_2) \geq v(S_1) + v(S_2),\ \forall S_1, S_2 \subset \mathcal{N}. \tag{4.3}$$

Supermodular games are also called convex games. For more details on bankruptcy games and their characteristics, please refer to Appendix A.

4.3.5 Possible Imputation Schemes

Solutions to cooperative games are essentially qualified with respect to the satisfaction of rationality constraints, desirable properties and existence conditions. Namely, the Core of a game is the set of imputations that satisfies individual and collective rationality (one or a coalition gets at least what it would get without cooperating), and efficiency (all the estate is allocated). Another interesting set of imputations is the Kernel, which contains those imputations that nullify the bargaining power of the players.

A commonly adopted solution for cooperative games in networking is the Shapley value, because it shows desirable properties in terms of null player, symmetry, individual fairness, and additivity [80]. It is defined as:

$$\Phi_i(v) = \sum_{S \subset \mathcal{N} \setminus \{i\}} \frac{|S|!(N - |S| - 1)!}{N!} [v(S \cup \{i\} - v(S)] \tag{4.4}$$

i.e., computed by averaging the marginal contributions of each femtocell in the network in each strategic situation i.e., (players' permutation). In convex games as the bankruptcy games, the Shapley is the Core center. Nevertheless, the Shapley value is not consistent [79], in the following sense.

Definition 4.3.3. An allocation $x = (x_1, x_2, ..., x_N)$ is consistent if $\forall i \neq j$ the division of $x_i + x_j$, prescribed for claims d_i and d_j, is $(x_i; x_j)$.

This means that no player or group of players can gain more by unilaterally deviating from a consistent solution since it will always obtain the same profit. For cooperative femtocell networks, this discourages clustering-like solutions inside an interference set. Another appealing solution concept, the Nucleolus, is not only consistent, but also the unique consistent solution in bankruptcy games, and is in the Core [83]. The Nucleolus is the imputation that minimizes the worst inequity. It is computed by minimizing the largest excess $e(x, S)$, expressed as:

$$e(x, S) = v(S) - \sum_{j \in S} x_j \ \ , \forall S \subset \mathcal{N} \tag{4.5}$$

The excess $e(x, S)$ measures the amount by which the coalition S falls short of its potential $v(S)$ in the allocation x; the Nucleolus corresponds to the lexicographic minimum imputation of all possible excess vectors. For more details on Shapley value and Nucleolus computation, please refer to Appendix A.

4.3.6 An Illustrative Example

We consider a femtocell network composed of six FAPs as shown in Figure 4.1, where the interference relationships are represented by the arrows in the graph and reported in Table 4.2. The corresponding interference sets are reported in Table 4.3.

Table 4.3: Interference sets.

Steps	**Femtocell Sets**
1	$\{F_1, F_2, F_3\}$
2	$\{F_1, F_2, F_4\}$
3	$\{F_4, F_5, F_6\}$
4	$\{F_2, F_4, F_5\}$
5	$\{F_1, F_3\}$
6	$\{F_4, F_6\}$

To each femtocell, we associate a value representing the demand of attached users (expressed in number of tiles). The OFDMA frame is composed of $E = 100$ frequency/time slots (tiles). We recall that the interference is asymmetric between femtocells because it depends on the positions of attached users.

The interference set list is presented in Table 4.3; the first step includes the players of a bankruptcy game $G(\mathcal{N}, v)$ where $\mathcal{N} = \{F_1, F_2, F_3\}$, and the coalitional payoffs are given in Table 4.4; $v(\mathcal{N}) = E = 100$ since no femtocell has participated to any previous game.

Table 4.5 reports the Shapley values (rounded) as well as the details on each femtocell's marginal contributions (columns).

For the Nucleolus, one starts at an arbitrary point such that $x_1 + x_2 + x_3 = 100$, e.g., $(50, 30, 20)$, as in Table A.3. Then, one minimizes the largest excess, corresponding to coalition F_3 in our case; but, this coalition can claim that every other coalition is doing better than it is. So, one tries to improve this coalition by making x_3 larger or, equivalently, $x_1 + x_2$ smaller since $x_3 = 100 - x_1 - x_2$ (feasibility property); but, decreasing the excess of F_3, the excess of $F_1 \cup F_2$ increases at the same rate and these excesses then meet at -30, when $x_3 = 30$. Clearly, no allocation x can make the excess smaller than -30 since at least one of the coalitions F_3 or $F_1 \cup F_2$ can have at least an excess of -30. Hence, $x_3 = 30$ is the first component of the Nucleolus. Proceeding in the same manner, one finally obtains the Nucleolus allocation $(35, 35, 30)$.

We move now to the second step. In this situation we have three femtocells, F_1, F_2 and F_4, and among them F_1 and F_2 have already taken their required resources; the remaining resources are assigned to F_4.

Then, at the third step, the total estate to distribute among femtocells is not 100 tiles any longer since F_4 has already participated to a game and obtained its resources; thus the new game is formed of two players, F_5 and F_6, and the total payoff $v(\mathcal{N})$ is then equal to $E - x_4 = 100 - 26 = 74$ tiles, as reported in Table 4.7. The Shapley value computation for this second game is illustrated in Table 4.8.

Table 4.4: Coalitional payoffs.

Coalition	$v(S)$
$\emptyset$	0
F_1	0
F_2	0
F_3	0
$F_1 \cup F_2$	40
$F_1 \cup F_3$	30
$F_2 \cup F_3$	10
$F_1 \cup F_2 \cup F_3$	100

Table 4.5: Shapley Value computation.

Permutation	F_1	F_2	F_3
F_1,F_2,F_3	0	40	60
F_1,F_3,F_2	0	70	30
F_2,F_1,F_3	40	0	60
F_2,F_3,F_1	90	0	10
F_3,F_1,F_2	30	70	0
F_3,F_2,F_1	90	10	0
Average	42	32	26

Table 4.6: Nucleolus computation.

Coalition	$e(x, S)$	$(50, 30, 20)$	$(38, 32, 30)$	$(35, 35, 30)$
F_1	$-x_1$	-50	-38	-35
F_2	$-x_2$	-30	-32	-35
F_3	$-x_3$	-20	-30	-30
$F_1 \cup F_2$	40-x_1-x_2	-40	-30	-30
$F_1 \cup F_3$	30-x_1-x_3	-40	-38	-35
$F_2 \cup F_3$	10-x_2-x_3	-40	-52	-55

Moreover, for the Nucleolus, we obtain the Table 4.9. The algorithm stops at this point since all femtocells have received their resources. As it can be noticed, the Nucleolus smoothes the maximum and the minimum allocation, preventing from extremely low and extremely high allocations for femtocells that interfere a lot and interfere a little, respectively.

Table 4.7: Coalitional payoffs.

Coalition	Payoff
$\emptyset$	0
F_5	0
F_6	24
$F_5 \cup F_6$	74

Table 4.8: Shapley Value computation.

Permutations	F_5	F_6
F_5, F_6	0	74
F_6, F_5	50	24
Average	25	49

Table 4.9: Nucleolus computation.

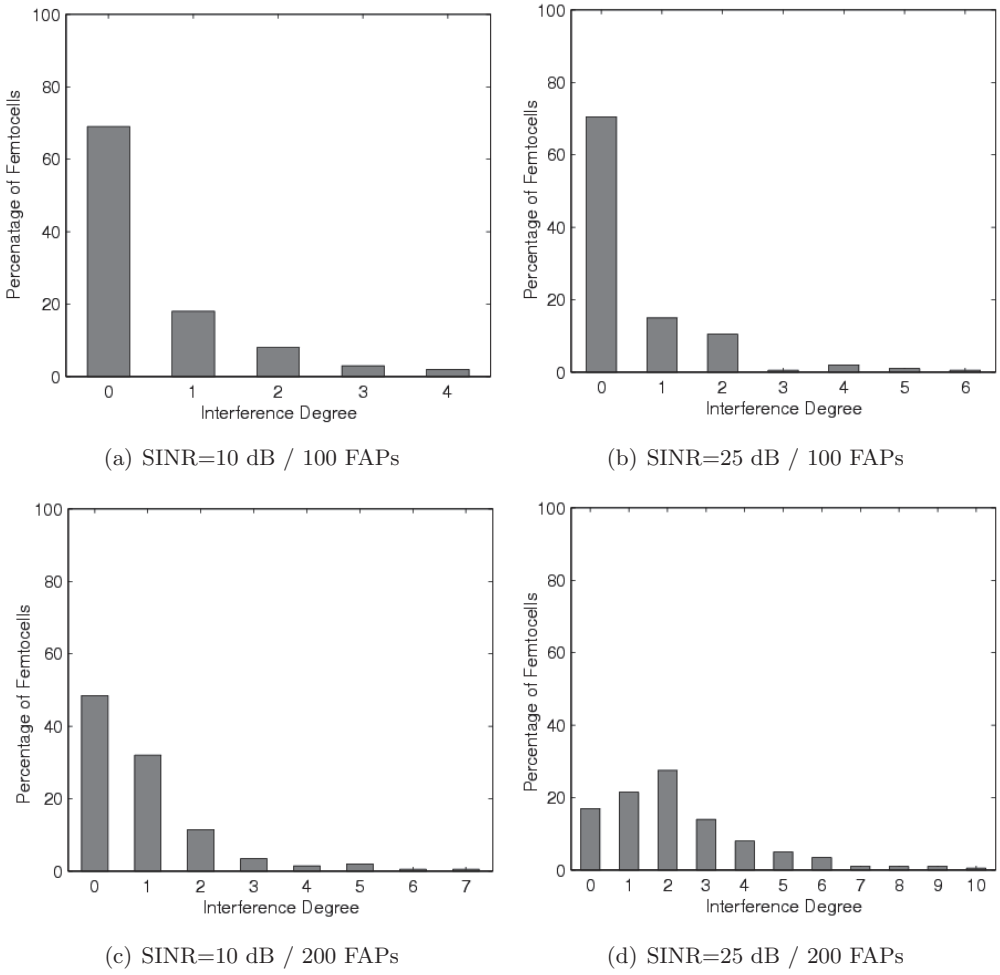
Coalition	$e(x, S)$	$(40, 34)$	$(25, 49)$
F_5	-x_5	-40	-25
F_6	24-x_6	-10	-25

4.4 Performance Evaluation

In this section, we evaluate the performance of the proposed game-theoretic approaches (i.e., Shapley value and Nucleolus) on large instances. C-DFP and F-ALOHA schemes, presented in Section 4.2, are used as benchmarks.

We simulate realistic scenarios with two different network sizes, with 100 and 200 FAPs, where for each simulation FAPs are randomly distributed in a 400m×400m area. We consider two interference level scenarios, a low-level one and a high-level one, based on two SINR thresholds, 10 and 25 dB, to show the impact of the interference degree on the performance metrics. Based on the SINR, the path loss model given in the A1 scenario for indoor small office and residential of WINNER for the frequency range 2−6 GHz [85] where the path loss depends upon the number of floors and traversed walls, and with static user positions; each femtocell determines the set of its interferer femtocells depending on the received signal strength. Users are uniformly distributed within the femtocells with a maximum number of four users per femtocell. Each user uniformly generates its traffic demand that can be directly translated to a certain number of tiles, with a maximum value of 25 tiles per user. As in [67], the analysis is achieved using a typical OFDMA frame (downlink LTE frame) consisting of $E = 100$ tiles.

Before delving into the exploration of the results, Figure 4.2 gives an idea about the topologies obtained for the four datasets, with the femtocell interference degree



(a) SINR=10 dB / 100 FAPs

(b) SINR=25 dB / 100 FAPs

(c) SINR=10 dB / 200 FAPs

(d) SINR=25 dB / 200 FAPs

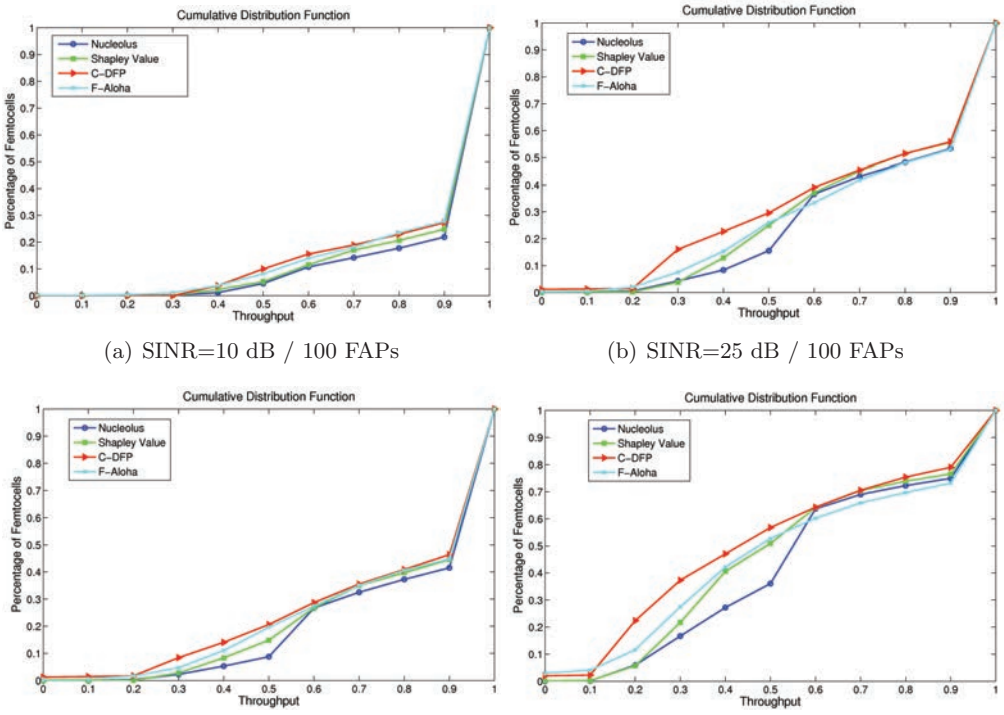
Figure 4.2: Interference degree distribution for the different datasets.

distribution (corresponding to the number of neighboring femtocells causing interference). As it can be noticed, 100-FAP topologies present a majority of isolated femtocells that do not suffer from interference, while this is no longer the case for 200-FAP topologies; both present a slight increase in the high degrees with the 25 dB SINR threshold.

Let us now focus on the comparison among the different strategies based on the offered throughput, the allocation fairness and the computation time. The results are obtained over many simulation instances for each dataset, with a margin error less than 3%; we do not plot corresponding confidence intervals for the sake of presentation.

4.4.1 Throughput Analysis

Figure 4.3 reports the mean normalized throughput (i.e., mean ratio of the number of allocated tiles to the total demand; in the following referred to as throughput)



(a) SINR=10 dB / 100 FAPs

(b) SINR=25 dB / 100 FAPs

(c) SNIR=10 dB / 200 FAPs

(d) SNIR=25 dB / 200 FAPs

Figure 4.3: Throughput Cumulative Distribution Function (CDF) for the four datasets.

for the four considered datasets.

For 100-FAP networks, Figs. 4.3(a) and 4.3(b), the game-theoretic approaches outperform the other schemes, for both interference levels. We can here appreciate how much the strategic constraints, and in particular the individual and collective rationality, contribute in avoiding low throughputs. This is further evidenced for larger instances as shown in Figs. 4.3(c) and 4.3(d). In particular, we can assess that:

- At low throughputs, F-ALOHA and C-DFP offer very low performance, especially in dense environments; e.g., in the 200-FAP case with high interference, around 3% of the femtocells obtain null throughput, and about 30% obtain a throughput less than 30%, while these numbers (percentage of femtocells) are roughly halved with game-theoretic approaches.

- The median throughput is always higher for the Nucleolus; e.g., in the 200-FAP case with high interference, 55% for the Nucleolus, 47% for the Shapley value, 45% for F-ALOHA and 41% for C-DFP, and the gap between the Nucleolus and the other methods decreases at lower interference and femtocell density levels.

- At high throughputs, F-ALOHA shows a small benefit over the Nucleolus, but in all cases the median throughput of the Nucleolus is still the highest among

Table 4.10: Mean fairness indexes.

FAPs	SINR	Nucleolus	Shapley Value	C-DFP	F-ALOHA
100	10 dB	0.96744	0.95742	0.95499	0.94248
	25 dB	0.90070	0.89288	0.85736	0.88003
200	10 dB	0.92465	0.91511	0.88741	0.90291
	25 dB	0.81668	0.78391	0.71604	0.72324

all approaches.

- Among the game-theoretic approaches, the Nucleolus persistently outperforms the Shapley value, with relevant differences at medium-low throughputs.

All in all, the Nucleolus seems the most appropriate approach with respect to the offered throughput, especially in high femtocell density and high interference environments as, e.g., in urban environments with a dense deployment of femtocells. Moreover, the C-DFP approach appears as the most inadequate one, and the F-ALOHA offers low throughputs to a significant portion of femtocells.

4.4.2 Fairness Analysis

We evaluate the fairness of the solutions using three aspects.

(i) with respect to the Jain's fairness index [89], defined as:

$$J_I = \left(\sum_{i=1}^{N}(x_i/d_i)\right)^2 / \left(N\sum_{i=1}^{N}(x_i/d_i)^2\right) \tag{4.6}$$

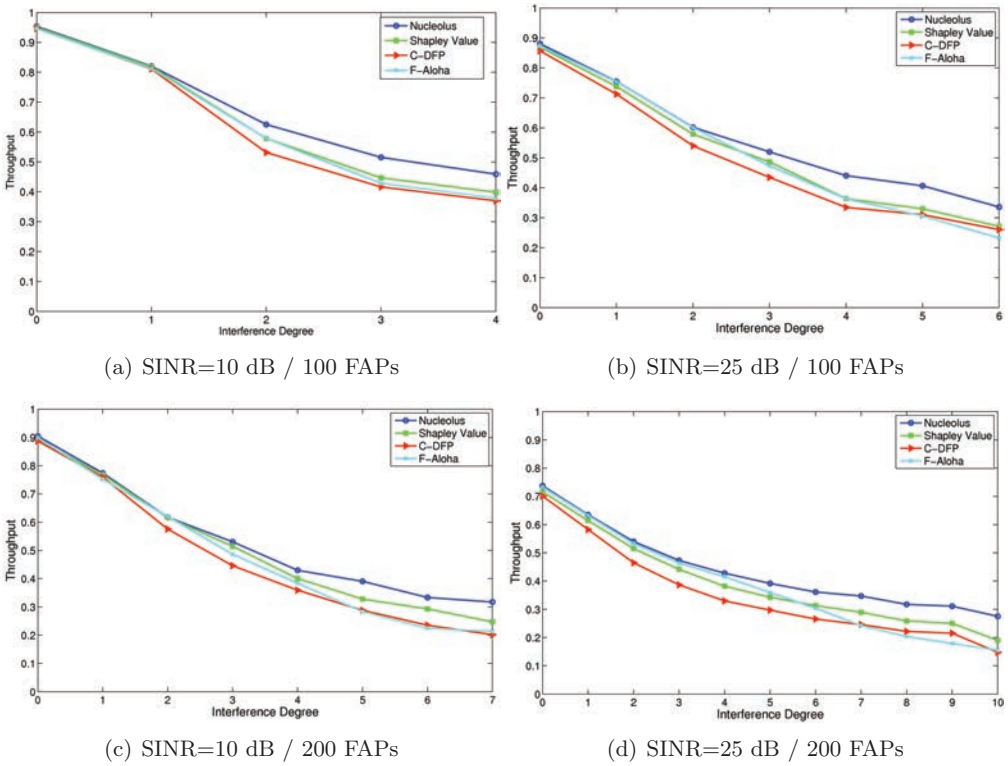
and reported in Table 4.10. It is easy to notice that game-theoretic approaches give the highest fairness, thanks to the strategic constraints that avoid penalizing femtocells presenting low interference degree and those with lower demands. Again, game-theoretic approaches outperform the others, with important differences with the 25 db / 200-FAP dataset.

(ii) Figure 4.4 further investigates how femtocell interference degree is taken into account, illustrating the mean normalized throughput as a function of the interference degree (recall that the interference degree of each femtocell corresponds to the cardinality of its interference set, minus one). We can assess that:

- The Nucleolus always outperforms the other methods, especially for femtocells with high interference degrees.

- The Shapley value behaves similarly to F-ALOHA and C-DFP, especially for small networks, while in large networks it shows a roughly 5% better throughput than F-ALOHA and C-DFP.

- Globally, C-DFP appears as the less efficient solution.



(a) SINR=10 dB / 100 FAPs

(b) SINR=25 dB / 100 FAPs

(c) SINR=10 dB / 200 FAPs

(d) SINR=25 dB / 200 FAPs

Figure 4.4: Throughput distribution as a function of the interference degree.

It seems appropriate to conclude that the interference degree is taken into account in a significantly different way with the Nucleolus, showing an interesting fairness performance certainly, especially desirable for urban dense environments.

(iii) in order to assess how the allocated resource is affected by the demand volume, Figure 4.5 plots the throughput as a function of the femtocell demand (just for the 200-nodes and 25 dB SINR threshold case as the other cases show little variations). Globally, the ALOHA and C-DFP approaches show a roughly constant behavior, and on the other hand, game-theoretic approaches decrease with growing demands. The results show that the Nucleolus favors low demands with respect to high demands significantly more than the Shapley value. This may be interpreted as unfair for high demands. However, under a network management standpoint, it might be seen a positive behavior as the Nucleolus can discourage too greedy demands at the benefit of lower 'normal' demands.

4.4.3 Time Complexity Analysis

Last but not least, it is important to assess if the overall good performance of game-theoretic approaches comes at the expense of a higher time complexity.

Figure 4.6 reports boxplots (i.e., quartile boxes plus maximum, minimum and outliers) of the computation time for the three pre-computation approaches.

It is easy to notice that C-DFP has quite high computation times, on the order

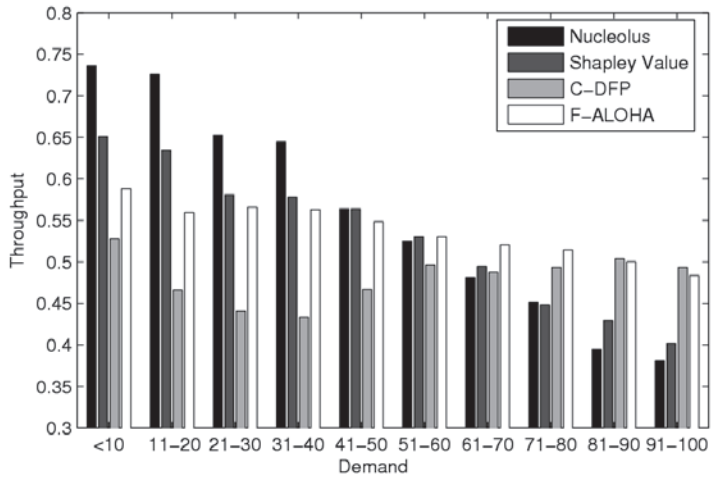


Figure 4.5: Throughput as a function of the demand (SINR=25 dB / 200 FAPs).

of seconds for 100-FAP networks and dozens of seconds for 200-FAP networks. A stronger dependence on the interference set size (higher for high interference levels) appears for the Shapley value, which is not surprising since the number of marginal contributions equals the factorial of the interference set size. In turn, the Nucleolus does not show any important dependence neither on the network size nor on the interference level, with a median computation time of roughly 2s for very dense high-interference environments.

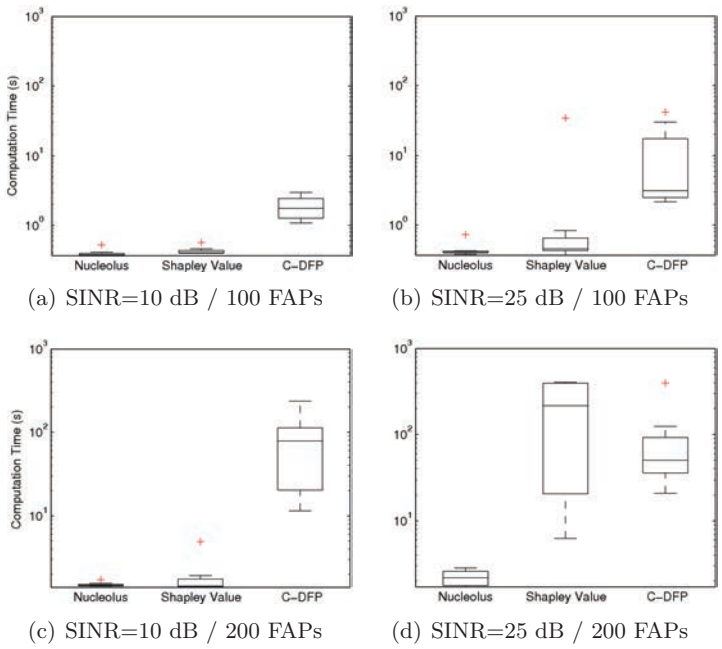
4.5 Wireless Mesh Networks Use-Case

Another small-cells context for which the proposed approach is appealing is the one of residential Wireless Mesh Networks (WMNs). WMNs are emerging as a key solution to provide broadband and mobile wireless connectivity in a flexible and cost effective way.[7] In this section, we investigate user cooperation path to implement strategic resource allocation in WMNs, under the assumption that users want to control their interconnections. The mesh networks rely on OFDMA communications so the resources can be expressed in the time-frequency domain and are organized in subchannels. For the sake of conciseness, we just report the results as the mathematical model is similar to the one presented in Section 4.3.4. We note that the players of this game are the Mesh Routers (MRs), and the estate that needs to be distributed among the players is the number of subchannels in an OFDMA frame.

4.5.1 Performance Evaluation

We compare the proposed approach to two state-of-the-art OFDMA allocation schemes: the centralized C-DFP and the distributed Frequency-ALOHA. We simulate realistic scenarios with three different network sizes (25, 50 and 100 nodes) representing respectively low, medium and large densities. MRs are randomly distributed in a 5km×5km area. Each MR determines the set of its interferers, inside

[7] Please refer to [2] and [9] for more details.



(a) SINR=10 dB / 100 FAPs (b) SINR=25 dB / 100 FAPs

(c) SINR=10 dB / 200 FAPs (d) SINR=25 dB / 200 FAPs

Figure 4.6: Computation time comparison for the four datasets.

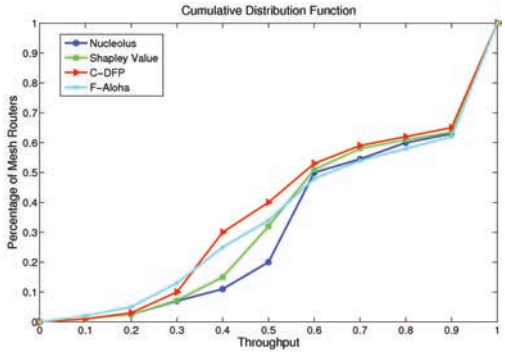
its coverage area. Mesh clients are uniformly distributed within a MR radius of 275m, and each one of them uniformly generates its traffic demand that can be directly translated to a certain number of subchannels. We consider a typical downlink OFDMA frame consisting of E = 60 subchannels. Figure 4.7 reports the mean normalized throughput (i.e., mean ratio of the number of allocated subchannels to the total demand) for the three considered datasets. We can here appreciate, as before, how much the strategic constraints in game theory approach, and in particular the individual and collective rationality, contribute in avoiding low throughputs. In particular, we can assess that:

- At low throughputs, F-ALOHA and C-DFP offer very low performance, especially in dense environments; e.g., the 100-node case, in F-ALOHA around 6% of the MRs obtain null throughput, and about 23% in C-DFP obtain a throughput less than 30%, while these numbers (percentage of nodes) are roughly halved with game-theoretical approaches.

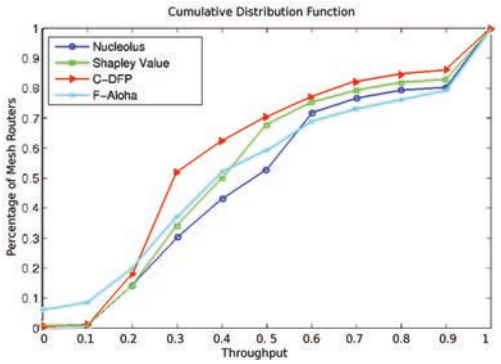
- The median throughput is always higher for the Nucleolus; e.g., in the 100-node case, 47% for the Nucleolus, 39% for the Shapley value, 37% for F-ALOHA and 29% for C-DFP.

- At high throughputs, F-ALOHA shows a small benefit over the Nucleolus, but in all cases the median throughput of the Nucleolus is still the highest among all approaches.

- Among the game-theoretic approaches, the Nucleolus persistently outperforms the Shapley value, with relevant differences at medium-low throughputs.



(a) 50 nodes



(b) 100 nodes

Figure 4.7: Throughput Cumulative Distribution Function (CDF) for the three cases.

All in all, the Nucleolus seems the most appropriate approach with respect to the offered throughput, especially in high-density environments. Moreover, the C-DFP approach appears as the most inadequate one, and the F-ALOHA offers low throughputs to a significant portion of the MRs.

4.6 Dealing with Cheating Behaviors

We note that in the proposed approaches, we may face the problem of cheating behaviors by some small-cell nodes (MRs or femtocells) as an effect of the demand-allocation approach, i.e. the small-cell nodes could end up with higher allocations if they claim higher demand. While in non-cooperative game theory cheating behaviors can be undetectable due to the uncoordinated nature of the decision-making process, we can manage this problem in cooperative games by a binding agreement that fixes the rules of the cooperation, i.e., our algorithm to compute the allocation, and possibly also the implementation of node blacklisting mechanisms. Such a mechanism should be operated upon explicit automated signaling by spectrum sensing nodes, detecting that allocated slots to other neighboring nodes are finally not used enough (since the channel is shared, this sort of operation is easily implementable by nodes' antennas during idle periods). A result of the blacklisting is the

isolation of the cheating node in the collaborative resource allocation and/or the systematic dropping of its traffic to be relayed in the small-cell network.

4.7 Summary

On the way toward fixed-mobile convergence, modern telecommunication networks are introducing novel technologies that better meet user requirements in terms of experienced quality. Small-cell offload is a promising direction to enable broadband access via legacy mobile handsets. For example, femtocells are expected to be commercialized as an added-value service subject to separate billing; strategic self-organization among femtocells is needed to cooperatively manage interference and resource allocation. In this chapter, we have investigated novel resource allocation approaches based on the theory of cooperative games, motivated by the fact that such approaches allow accounting for strategic interactions among independent femtocells, and by the intuition that they shall offer better performance in small-scale urban and dense environments.

In particular, we proposed a game-theoretic approach for strategic resource allocation in cooperative small-cell OFDMA networks. Upon distributed detection of interference maps, our approach iterates bankruptcy games from the largest interference set with highest demand to the lower sets. We motivated the adoption of solutions from coalitional game theory, the Nucleolus and the Shapley value, highlighting how their properties can help meeting performance goals. Through extensive simulations using realistic datasets, we compared our game-theoretic approaches to state-of-the-art proposals. With respect to throughput and fairness, our approaches outperform the others. In particular, the Nucleolus solution is strictly superior to all the others, achieving higher throughputs also for low-demand and high-interference femtocells. Moreover, computationally, the Nucleolus is far more competitive than the other approaches. The Nucleolus approach represents therefore a promising approach for resource allocation in future femtocell network deployments.

Chapter 5

Vertical Traffic Offloading over Passpoint Access Points

In the previous chapter, we have investigated the traffic offloading solutions over small-cell networks. Despite the promising results therein in terms of achievable performance, the explicit coordination and signaling among femtocells add a certain level of complexity and therefore significant investments need to be undertaken in order to implement this type of traffic offloading. In this chapter, a much simpler, inexpensive and lightweight solution consisting of using Wi-Fi hotspots for data traffic offloading is discussed and evaluated.[1]

5.1 Introduction

Wi-Fi technology has always been an attractive solution for catering the increasing data demand in mobile networks because of the ubiquity of Wi-Fi networks, the high bit rates they provide, the simplicity in deployment and maintenance, and the lower CAPEX [90] [91]. However, the legacy WiFi technology lacks of seamless interworking between Wi-Fi and mobile cellular networks on the one hand, and between Wi-Fi hotspots on the other hand.

Nowadays, the recently released Wi-Fi Certified Passpoint Program (also known as 'Hotspot 2.0' and referred to in the following shortly as 'Passpoint') [92] provides the necessary control-plane for these operations. It aims to make the WiFi network a "true extension of service provider networks", letting users roam from one hotspot to another with no manual effort, just like cell phone network that already switches seamlessly from one cell tower to another. The Passpoint technology provides all control-plane functionalities for automated and seamless connectivity to Wi-Fi hotspots.

Service providers can henceforth look to such Wi-Fi systems as a viable way to seamlessly offload mobile traffic and deliver high-bandwidth services. At the same time, subscribers no longer have to face the frustration and service degradation typically experienced when connecting to legacy Wi-Fi hotspots.

[1] The contents of this chapter are presented in [3] and [13].

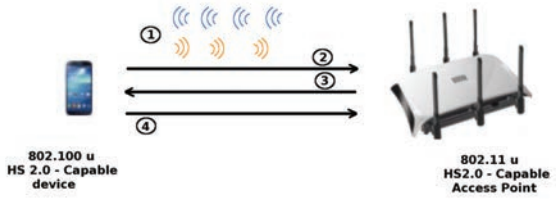


Figure 5.1: Passpoint hotspot association.

In this chapter, we evaluate the capacity as well as the energy saving gains that one can get by offloading cellular data traffic over Wi-Fi Passpoint hotspots as a function of different hotspot placement schemes and of access point selection policies.

In the beginning of the chapter, we give an insight on the hotspot-device signaling information exchanged with Passpoint and moreover we review the state of the art on Wi-Fi data traffic offloading.

5.2 Passpoint Hotspot-Device Signaling

Passpoint can work in any network and overcomes the limitations of proprietary, non-interoperable solutions offered by some providers today. Devices certified in the Passpoint program are able to manage network association, authentication, sign-up, and security in the background, in a way that is completely transparent to the subscriber and that consistently works in any Passpoint network [93] [94]. When a user with a "Hotspot 2.0" (HS2.0) capable mobile device (i.e., based on IEEE 802.11u) comes within the range of a HS2.0 capable hotspot, it automatically opens up a dialog with that hotspot to determine its capabilities before proceeding to authentication. It is worth noting that Passpoint logic is already implemented in many mobile devices, such as Android-based ones. Moreover, since Passpoint discovery is based on pre-authentication, there are considerable savings of time and battery life compared to existing methods [92].

Figure 5.1 illustrates the four different required steps for Passpoint hotspot association. The Access Network Query Protocol (ANQP) is used for device-hotspot signaling [92].

1. The 802.11u-capable access point broadcasts its HS2.0 support, so that HS2.0-enabled devices can recognize such support.

2. The 802.11u-capable device is able to process ANQP messages, containing useful information such as the 'reachable' authenticators, and various hotspot capabilities. The 802.11u-capable device requests full authenticators list.

3. The hotspot responds to the ANQP query with the requested information.

4. Device compares provisioned network-selection policy with HS2.0 data from hotspots and associates itself to the best hotspot suitable for its needs.

Table 5.1: Beacon and probe response information elements in Passpoint.

Access Network Type	identifies whether hotspot is for public, private or guest access.
Internet Bit	indicates if the hotspot can be used for Internet access.
Advertisement Protocol	indicates if the hotspot supports GAS/ANQP
Roaming Consortium element	provide a list of up to 3 names of reachable service providers.
Venue Information	describes the venue where the hotspot is situated.
Load Element	provides information on channel utilization and the current number of associated devices.

Table 5.1 shows some of the information elements provided by the hotspot to the mobile devices. In the specifications, those six elements are mentioned to date. Most elements provide simple configuration and network reachability and locality information. The most interesting element for efficient Passpoint selection is the Load Element, which allows a mobile device to be informed about hotspot channel utilization and the current number of associated devices to a Passpoint hotspot.

5.3 Related Work on WiFi Traffic Offloading

Because of its recent specification, the scientific papers from the literature discussed in this section do not consider the Passpoint technology explicitly along with its hotspot selection capabilities. We present thereafter a selection of Wi-Fi offloading strategies available in the literature.

Authors in [95] measure the offloading potential of the public WiFi based on city wide vehicular traces. Compared to the vehicle based high mobility scenario in [95], the authors in [91] study the performance of 3G mobile data offloading through Wi-Fi networks in a more general mobile scenario with empirical pedestrian traces. They distinguish two different types of Wi-Fi offloading: *on-the-spot* and *delayed* offloading. The first type consists of spontaneous connectivity to Wi-Fi and transfer data on the spot; when users move outside the Wi-Fi coverage area, the offloading is stopped and all the unfinished transfers are transmitted back to cellular networks. In the delayed offloading, each data transfer is associated with a deadline and as users come in and out of Wi-Fi coverage areas, their data transfer is repeatedly resumed until the transfer is complete or the deadline is reached. Based on a study done over some smartphone users and on the statistical distributions of their Wi-Fi connectivity, the authors evaluate the Wi-Fi offloading efficiency for various amount of Wi-Fi deployment, different deployment strategies, different traffic intensity and delay deadlines, showing that Wi-Fi in such configurations can offload up to 65% of the total mobile data traffic. Similarly, the authors in [96] explore the benefits in terms of energy savings that can be achieved by offloading traffic loads to Wi-Fi networks. Using different traffic types, they show that a saving of up to 70% is reached by opportunistically powering down cellular radio network equipment to

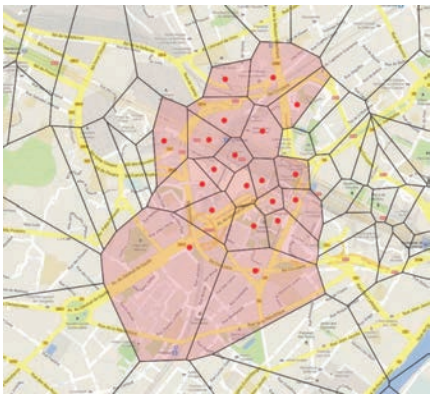


Figure 5.2: The considered region: La Défense.

offload users traffic to Wi-Fi hotspots.

In [97], a WiFi offloading scheme is proposed from a transport layer perspective. A multipath protocol called oSCTP is proposed to offload the 3G traffic via WiFi networks and maximize the user's benefit. The philosophy of oSCTP is to use WiFi and 3G interfaces simultaneously if necessary, and schedule packets transmitted in each interface every schedule interval. By modeling user utility and cost both as a function of the 3G and WiFi network usage, the user's benefit, i.e., the difference between the utility and the cost, is maximized through an optimization problem. Following the same direction, the authors in [98] propose a framework for 3G traffic offloading based on the idea of motivating mobile users with high delay tolerance to offload their traffic to Wi-Fi networks. A feasible approach consists of delaying all delay tolerant applications until their maximum delay tolerance, and then resorting to the cellular networks if the applications can not finish. However, this approach does not appear much effective, considering that the user has to wait even when there is actually no available Wi-Fi connection. To solve this problem, the authors in [99] propose an adaptive approach that computes an offload handing-back time, after which the user stops waiting for offloading through Wi-Fi connections, hence resorting to the cellular network service. This allows achieving a better trade-off between offloaded volume and user satisfaction.

5.4 Cellular Network Dataset

The dataset used in our study consists of Orange probe's data explained in details in Section 3.2. We limit the study in the chapter to the "La Défense" region, a major business district in the northwest of Paris. The region is decomposed as shown in Figure 5.2 at base station level, where red dots represent the base stations and the surrounding polygons represent the Voronoi cells. The size of a Voronoi cell depends on two basic factors: the geographical position and the coverage area (determined according to power level) of the corresponding base station. We analyze the data in a normal working day from 8 am to 10 am when people make their regular home-to-work travel. We choose this period to capture users mobility in the region. Upon this selection, we extract data consumption of about 20000 users.

5.5 Data Traffic Offloading over Passpoint Hotspots: Methodology

Given a sample geographical distribution of Passpoint hotspots, we extract user displacement information from the data. Along its trajectory, a mobile device encountering a Passpoint hotspot, or a number of Passpoint hotspots at one location, can learn about the service providers available via each of them, as well as other characteristics of the hotspot, via the ANQP protocol. Thanks to this signaling, the mobile device can discover a comprehensive profile of the hotspot before association, so it can quickly identify, prioritize hotspots suitable for its needs and select the best match while still in the user's pocket. We draw the whole offloading procedure in the flow chart presented in Figure 5.3.

The hotspot selection policy is therefore of paramount importance for both the user, able to associate to the best access point, and the network, which should avoid hotspot and backhauling link congestion. We compare three different hotspot selection policies, each taking into consideration one different parameter, as described in the following:

1. *Number of Associated Devices*: the user is attached to the hotspot with the least number of associated devices (this information is provided by the hotspot in its response to the ANQP query as presented in Table 5.1).

2. *Channel Utilization*: the user is attached to the hotspot with the least Channel Utilization defined as the percentage of time the hotspot senses the medium busy (this information is also provided by the hotspot in its response to the ANQP query).

3. *Signal Quality*: the user is attached to the hotspot with the best received signal power.

While the first two are retrievable information via the ANQP Passpoint signaling, the latter instead does not strictly depend on Passpoint and can be considered as a policy that could easily be implemented with a relatively limited programming of mobile device's drivers ignoring hotspot capabilities.

After selecting the suitable hotspot, the mobile device is automatically authenticated. In Passpoint, this is done using Extensible Authentication Protocols (EAP) based on a Subscriber Identity Module (SIM) authentication, an authentication that is widely used in cellular networks today [93]. This procedure is specified in such a way that the process is completely transparent to the subscriber and that consistently works in any Passpoint network.

Then, the offloading process starts; only delay-tolerant traffic is offloaded to Passpoint hotspots, while retaining delay-sensitive traffic in mobile cellular networks. The delay-tolerant traffic is nothing else than TCP traffic. If the user moves out of the coverage of the Passpoint hotspot and finds no other hotspots in the environment, it returns back to the cellular network transparently.

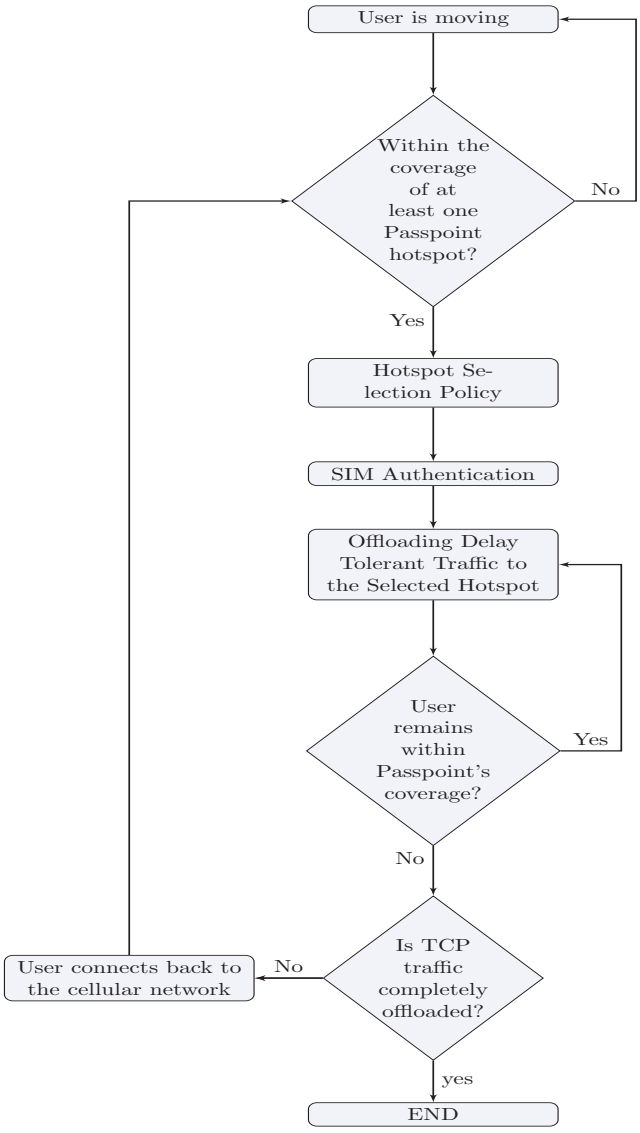


Figure 5.3: Offloading algorithm.

5.6 Data Traffic Offloading over Passpoint Hotspots: Simulation Results

In this section we describe the simulation framework we used to evaluate different offloading policies. For each simulation, the Passpoint hotspots are distributed in the selected region presented in Figure 5.2 of approximately 1 km^2. The results are obtained over many simulation instances, with a margin error lower than 3%; we do not plot corresponding confidence intervals for the sake of presentation. In the following, we first present the radio model then we compare different offloading

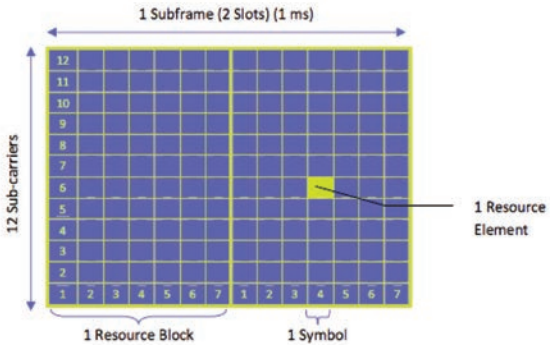


Figure 5.4: OFDMA frame structure.

policies and hotspot placement strategies.

5.6.1 Radio Model

The macro-cells are assumed to operate using the OFDMA technology (e.g., in LTE) whose frame structure is based on time-frequency slots (see Figure 5.4), also called tiles or resource blocks (RBs). A set of parameters for typical transmission bandwidths for LTE in the downlink is shown in Table 5.2, where the subcarrier spacing is $\Delta f = 15$ kHz. We select 20 MHz as the transmission bandwidth, therefore the number of resource blocks per frame is equal to 100 RBs, e.g., allowing a max throughput of 100.8 Mb/s for the 64 QAM modulation.

Table 5.2: Typical parameters for downlink transmission.

Transmission bandwidth [MHz]	20
Number of resource blocks	100
OFDMA symbols per 1 ms	14
Modulation symbol rate (Mb/s)	16.8
QPSK Bit Rate (Mb/s)	33.6
16QAM Bit Rate (Mb/s)	67.2
64QAM Bit Rate (Mb/s)	100.8

These parameters are used to compute user demands in terms of RBs knowing only the volume in bytes. We note here that the modulation used by each user depends on its Signal to Noise plus interference (SINR) level and the path loss. We use the COST-231 Hata path loss model [100], devised as an extension to the Okumura-Hata model, which is the most widely used radio frequency propagation model for predicting the behavior of cellular transmissions in urban areas [101].

For the Passpoint hotspots, we employ a SINR interference model [102]. Each hotspot is assigned one channel from the 13 available channels in France on the 2.4 GHz frequency range. If the hotspot j transmits signals to user i, the SINR

Table 5.3: Channel overlapping degree.

Channel Distance	0	1	2	3	4	5	6	≥7
Overlapping Degree	1	0.7272	0.2714	0.0375	0.0054	0.0008	0.0002	0

computed by user i is expressed as follows:

$$SINR_i = \frac{Pd(i,j)^{-\alpha}}{N + \sum_{k \in A, k \neq j} P\lambda(i,k)d(i,k)^{-\alpha}} \tag{5.1}$$

where:
P is the transmission power of the hotspot (i.e., for simplicity we assume all hotspots use the same transmission power P of 20 dBm);
$d(i,j)$ is the distance between user i and the hotspot j;
α is the path loss index (a value typically between 2 and 4);
N is the background noise (i.e., we set this value to -96 dBm);
A is the group of the hotspots existing in the network;
$\lambda(i,k)$ is the channel overlapping degree between the channels used by i and k; it decreases when the channel distance between i and j increases. The channel overlapping degree is computed by [103] and shown in Table 5.3. We note that when the channel distance is 5 or above, the overlapping degree becomes negligible. The maximum achievable capacity of access points is set to 54 Mbps.

In the following, we compare various scenarios with respect to the capacity gain (CG) that we can get by offloading users traffic to Passpoint hotspots. The CG is defined as:

$$CG = RB_{freed}/RB_{total} \tag{5.2}$$

where RB_{freed} is the total number of RBs freed from the cellular mobile after offloading data traffic over Passpoint hotspots, and RB_{total} is the total number of RBs required by users before offloading data traffic over Passpoint hotspots.

5.6.2 Achievable Gain with different Hotspot Selection Policies

Figure 5.5 illustrates the capacity gain (in percentage) that we get for the three different selection policies with a random distribution of hotspots in the selected region. We can clearly notice that:

- the capacity gain increases with the Passpoint density, as the probability of encountering a Passpoint while moving increases;
- the capacity gain with the Passpoint-agnostic Signal Quality policy gives results similar to those at the state of the art only for very high hotspot density, over 120 hotspots per square km;
- the *Channel Utilization* offloading policy outperforms the other ones and offers the highest capacity gain. A reasonable justification of this behavior is that this policy equally distributes the users to hotspots taking into account traffic volume and hence allowing hotspot resources to be efficiently utilized;

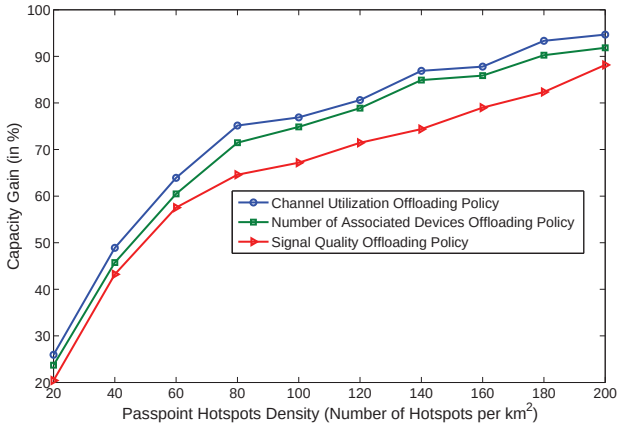


Figure 5.5: Capacity gain for different Passpoint hotspot selection policies.

- with the *Signal Quality* offloading policy, all users in a close location are assigned to the same hotspot because they will all receive AP signals with the same power. As a result, there will be a larger number of users competing for limited resources in the unilaterally best hotspot whereas the resources in the other hotspots remain free and hence wasted;
- The *Number of Associated Devices* offloading policy does not take into account the traffic volume required by each user and thus inefficiently distributes the users to hotspots.

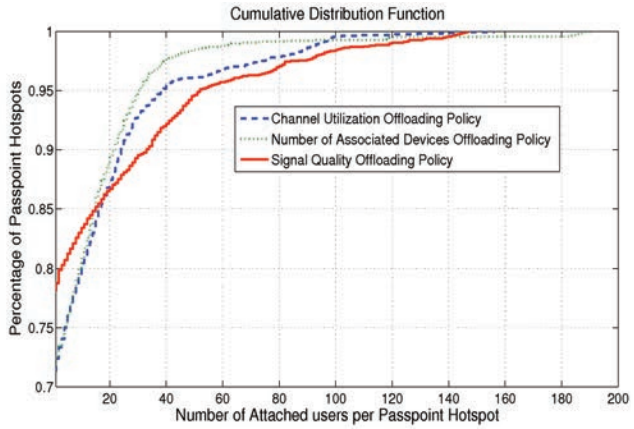


Figure 5.6: CDF of the number of attached users per Passpoint hotspot (density of 80 hotspots/km^2).

Furthermore, Figure 5.6 and 5.7 show, respectively, the cumulative distribution function of the number of users attached as well as the traffic volume per Passpoint hotspot using the three offloading policies (for a hotspot density of 80 hotspots/km^2). We notice that the percentage of low-loaded hotspots is higher in

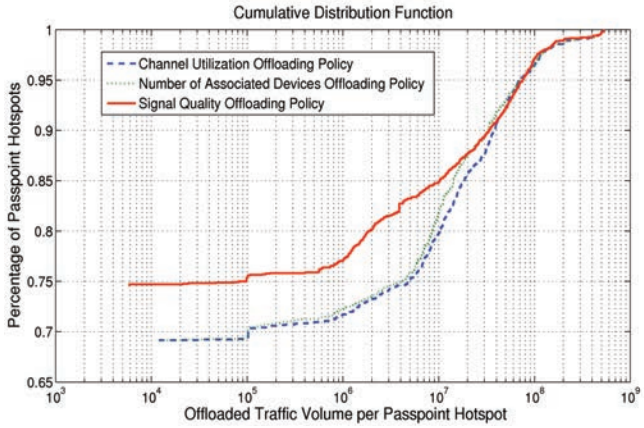


Figure 5.7: CDF of offloaded traffic volume per Passpoint hotspot (density of 80 hotspots/km^2).

the *Signal Quality* offloading policy than the other two policies. More precisely, with this *Signal Quality* offloading policy, approximately 80% of Passpoint hotspots have less than four attached users while 73% of hotspots have this value in the other two policies. Also 77% of hotspots offloading each less than 1 MB of traffic in *Signal Quality* while 71% and 72.5% in *Channel Utilization* and *Number of Associated Devices* respectively. Moreover, the percentage of highly-loaded hotspots is bigger in *Signal Quality* offloading policy than the other two: 5% of hotspots with more than 52 users in *Signal Quality* while 3.5% in *Channel Utilization* and 1.5% for *Number of associated devices*; 11.5% of hotspots offloading each more than 30 MB in *Signal Quality* and *Number of Associated Devices* policies to 13% in *Channel Utilization* offloading policy. These results confirm the previous findings and emphasize the more efficient usage of resources and distribution of traffic among different hotspots in the *Channel Utilization* offloading policy.

All in all, starting from a discrete Passpoint hotspot density, the gain of using the best among Passpoint offloading policies (i.e., the Channel Utilization one) and the offloading policy implementable without Passpoint (the Signal Quality one) is of roughly 15%. These results are obtained for a random distribution of Passpoint hotspots, so the next question to answer is what is the most appropriate hotspot placement scheme.

5.6.3 Passpoint Placement Schemes

We compare different Passpoint placement schemes in order to assess the impact of Passpoint positions on the offloading system performance.

Given the base station antenna-centric nature of cellular access, and more generally of wireless access, we consider different placement schemes depending on a parameter expressing the Distance To Borders (DTB) defined as:

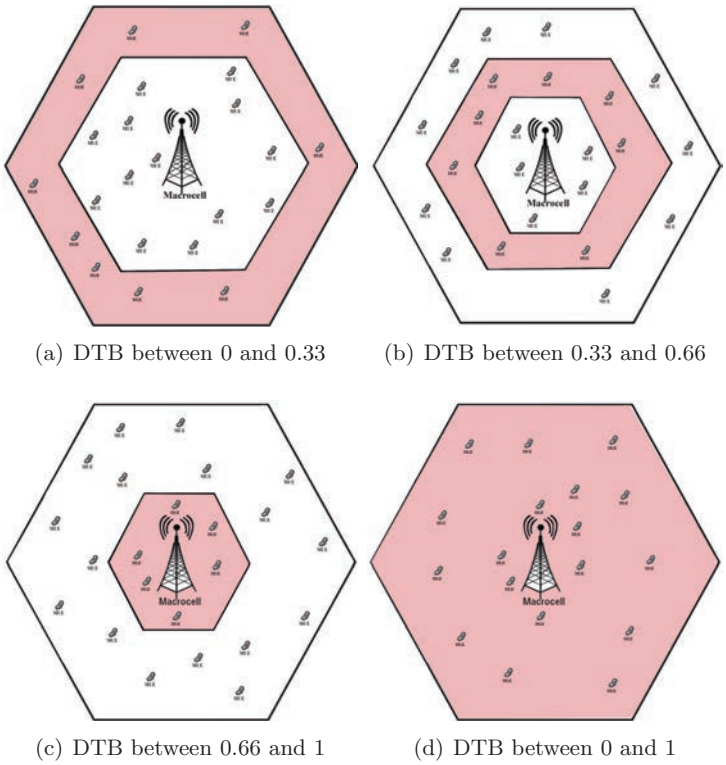
$$DTB_{i,j} = \frac{distance(P_i, \diamond_j)}{distance(M_j, \diamond_j)} \tag{5.3}$$

where:
P_i is the i^{th} Passpoint and M_j is the j^{th} macro-cell in the region.
$\diamond_j$ is the polygon that surrounds the coverage area of Macro-cell j.
$distance(P_i, \diamond_j)$ is the minimal distance from the Passpoint P_i to all ribs of $\diamond_j$.
Based on the DTB parameter, we select four different placement schemes, presented in Figure 5.8 where the colored zone represents the region of installing the Passpoint hotspots.

We consider the placement of Passpoint hotspots in the:

- outer annulus (i.e., zone close to the edge) of the macro-cell coverage, with a DTB between 0 and 0.33, as in Figure 5.8(a);
- middle annulus (i.e., central zone) of the macro-cell coverage, with a DTB between 0.33 and 0.66, as in Figure 5.8(b);
- inner annulus (i.e., zone closest to the base station) of the macro-cell coverage, with a DTB between 0.66 and 1, as Figure 5.8(c);
- whole macro-cell zone, randomly distributed, with a DTB between 0 and 1, as in Figure 5.8(d).

Figure 5.9 illustrates the results obtained by varying the hotspot placement schemes. We consider here the *Channel Utilization* policy which appears as the best Passpoint offloading policy. We can clearly notice that:



(a) DTB between 0 and 0.33 (b) DTB between 0.33 and 0.66

(c) DTB between 0.66 and 1 (d) DTB between 0 and 1

Figure 5.8: Illustration of different hotspot placement schemes.

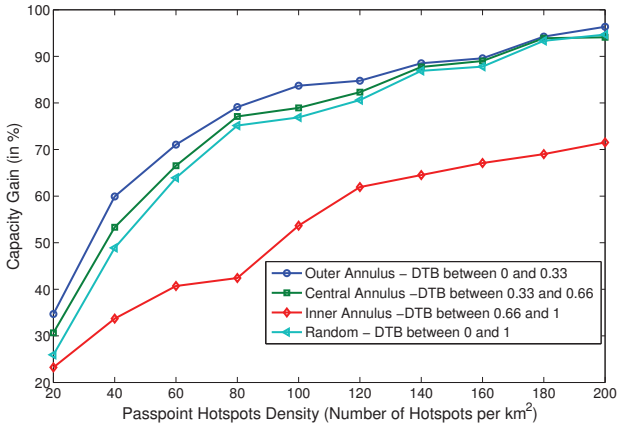


Figure 5.9: Capacity gain for different hotspot placement schemes under the best hotspot selection policy.

- the hotspot placement with DTB between 0 and 0.33 (i.e., installing Passpoint hotspots in the outer annulus of the macro-cell coverage) is the best placement scheme, which guarantees the highest capacity gain. The interpretation is straightforward as users located at the edge of the macro-cell base station suffer from a low SINR; therefore, the modulation chosen for those users is the one that requires the least number of bits per symbol (i.e., QPSK modulation in our case) to reduce the symbol error rate. Those users have low bit rates and thus require more time and more RBs to transmit their traffic. By offloading their traffic to Passpoint hotspots, we free a big number of RBs from the cellular networks.

- The topology corresponding to DTB between 0.66 and 1 (i.e., inner annulus) is the worst among others. Differently than for the outer annulus case, users close to the macro-cell base station use the modulation that requires the highest number of bits per symbol: those users have a high bit rate and require less time and RBs. So offloading their traffic is not very beneficial for cellular networks.

- The topology corresponding to DTB between 0 and 0.33 overcomes the random one (DTB between 0 and 1) with a mean capacity gain of roughly 5%, and that with DTB between 0.33 to 0.66 (i.e., central annulus) with a mean gain of roughly 3%.

Finally, we evaluate different offloading policies under the best hotspot placement scheme, i.e., the case where Passpoint hotspots are placed only in the outer annulus. Figure 5.10 illustrates the obtained results, where the dotted lines refer to the random hotspot placement replicated from Figure 5.5. The figure shows that the gap between Passpoint policies and the signal quality policy is further increased when placing the hotspot in the outer annulus only. We notice a mean difference between the outer and random placement schemes of around 11% for low hotspots

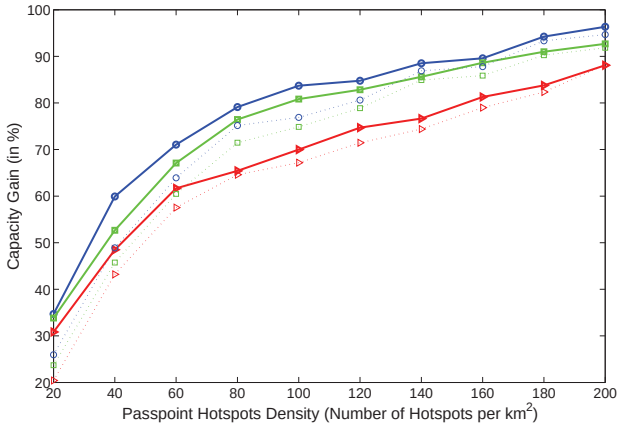


Figure 5.10: Capacity gain for different Passpoint hotspot selection policies under the best placement scheme.

density and this difference decreases for high hotspots density with a mean difference of around 3%. Overall, with hotspot placement in the outer annulus, the gain increases when using the Passpoint-enabled offloading policies rather than the signal quality one and this gain is around 15%.

5.6.4 Energy Saving Gain

Yet we have studied the capacity gain that an operator can get by offloading data traffic over Passpoint hotspots but what about the gain from users' point of view? Does this offloading solution increase the battery lifetime of mobile phones?

We therefore study whether offloading mobile data traffic over Passpoint hotspots is worthwhile, in terms of energy. As we know, marcocells cover a relatively large area compared to Passpoint hotspots. Therefore the long range radio communication between the mobile phone and the macrocell have high energy per MB requirement compared to shorter-range wireless technologies (i.e., Passpoint hotspots in our case). To estimate the energy consumption in both technologies, we use the parameters of Table 5.4 reported from [104] and [105]. The first column refers to the energy cost for transferring 1 MB of data for both macrocells and WiFi technologies; it is easy to notice that the transfer energy of 1 MB of data over macrocells is 20 times higher than that of WiFi. The second column refers to the additional energy cost of maintaining a connection alive independently of the traffic volume; it is equal to zero for macrocells since we assume that the phone is always connected to the cellular network. The third column refers to the energy cost for connection establishment.We define the Energy Saving Gain (ESG) as follow:

$$ESG = 1 - \frac{Energy\ Consumption\ after\ Offloading}{Energy\ Consumption\ before\ Offloading} \tag{5.4}$$

Figure 5.11 shows the average energy saving in % (i.e., average energy saving of all users in the region) that one can get by offloading mobile data traffic over

Table 5.4: Energy consumption of smartphone networking interfaces.

	Transfer (Joule/MB)	Idle (Watt)	Scan (Watt)
Cellular	100	0	0
WiFi	5	0.77	1.29

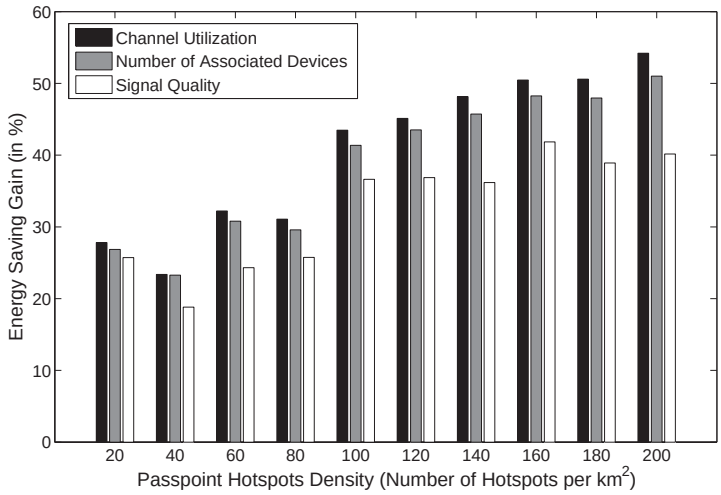


Figure 5.11: Energy saving of offloading mobile data traffic over Passpoint hotspots for different offloading policies.

Passpoint hotspots for different offloading policies. The same dataset of previous simulations is used. As before, the hotspots are randomly distributed. We can clearly notice that:

- The energy is better saved when the number of Passpoint hotspots increases, as the user's probability to encounter a Passpoint and thus offload its data traffic over Passpoint increases (WiFi communications are 20 times more energy-efficient than macrocells, indeed).

- The *Signal Quality* offloading policy is less energy-efficient than the other two and tends to be stable after a density of 100 hotspots per km^2, this can be explained by the higher percentage of highly-loaded hotspots in *Signal Quality* offloading policy compared to the other two policies as reported in Figure 5.6 and Figure 5.7. Thus with the *Signal Quality* offloading policy, users compete more with each others to get access to the selected hotspot and end up sometimes without being able to transfer their traffic over that hotspot, which results in a waste of energy.

- The *Channel Utilization* offloading policy outperforms the other policies in terms of energy consumption and saves from 28% of energy in low hotspots

density to 54% in high hotspots density. It saves up to 4% and 13% of energy comparing to the *Number of Associated Devices* and *Signal Quality* policies.

All in all we can emphasize the strength of offloading mobile data traffic over Wifi-certified Passpoint hotspots in terms of both user's device and spectrum capacity gain from the cellular network operator.

5.7 Summary

Traffic growth is outstripping the capacity of cellular mobile networks, especially in urban and densely populated zones. Moreover, operators are under pressure to find solutions to keep up with their customers' insatiable demand for data intensive applications. Data traffic offloading to Wi-Fi hotspots has always been an attractive solution for catering the increasing data demand in mobile networks, despite the existence of some drawbacks that limit their usage. Nowadays, with the advent of the Passpoint Certified Program [92], offloading data traffic to Passpoint hotspots is back to the forefront. The Passpoint Certified Program was created to address critical business needs for mobile data, streamline access and to help ease operator data traffic offload to Wi-Fi networks in a completely transparent way for the user.

In this chapter, we compare different conceivable mobile data traffic offloading approaches over Passpoint hotspots, using real mobile consumption data gathered from the Orange mobile network in Paris. First, we provide a brief analysis of mobile data consumption and characteristics. Then, we compute the capacity gain as well as the energy saving gain that one can get by offloading users traffic while taking into account different offloading policies and hotspot placement schemes. In particular, we show that offloading using Passpoint control-plane information can grant up to 15% capacity gain and 13% energy saving gain with respect to Passpoint-agnostic ones based on signal quality information.

Chapter 6

Content Offloading in Information Centric Networking

In the previous two chapters, we have investigated solutions for cellular radio overloading situations using small-cell networks and Wi-Fi certified Passpoint hotspots. Traffic overloading can also manifest in cellular backhauling wire-line links. In this context, a new solution appears to the forefront to manage traffic overload: Information Centric Networking (ICN), that allows offloading content distribution from content service providers by means of in-network caching. In this chapter, we focus on high cache contention involving multiple Content Providers (CPs) and one ICN provider having to give them access to its caches. We propose a resource allocation and pricing framework to support the network provider in the cache allocation to multiple CPs, for situations where CPs have non-overlapping sets of files and untruthful demands need to be avoided.[1]

6.1 Introduction

With the advent of broadband and social networks, the Internet became a worldwide content delivery platform [106] [107], with high bandwidth and low latency requirements. To meet the always increasing demand, contents are pushed as close as possible to their consumers and Content Providers (CPs) install dedicated storage servers directly in the core of Internet Service Provider (ISP) networks [108].

Conscious of the mismatch between the network usage and its conception, the research community recently proposed the concept of *Information Centric Networking* (ICN) [108] [109]. With ICN, content objects can be accessed and delivered natively by the network according to their name rather than relying on IP addresses [107]. By naming information directly at the network layer, ICN favors the deployment of in-network caching, and therefore each content can be found potentially anywhere in the network, moved or replicated at different locations [110] [111].

ISP networks then become native distributed storage systems, i.e., ICN providers can directly sell caching capabilities to content providers instead of hosting their

[1] The contents of this chapter are presented in [14].

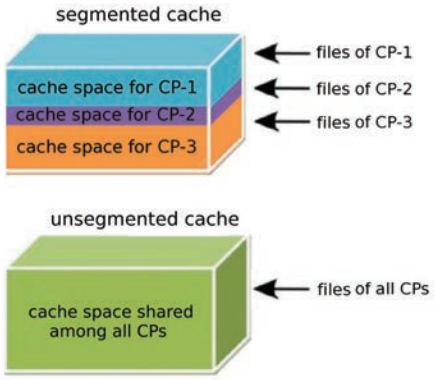


Figure 6.1: Representation of segmented and unsegmented caches with many content providers (CPs).

servers. However, it is most probable that the storage demand exceeds the total ISP storage offer, at least for those content caching locations the closest to the users. So far, the contention is solved by considering each storage as one autonomic and self managed cache (e.g., using LRU, least-recently-used, mechanism), as depicted in the bottom of Figure 6.1.

In this chapter, we propose to address this contention situation by segmenting the storage on a per-content provider basis, as depicted in the top of Figure 6.1. Each content provider receives a portion of the storage space depending on its storage demand. For this, based on *mechanism design theory* [112] [113], we propose a 2-step game-theoretic design that computes a fair and rational sharing of resources between CPs. The first step relies on a ICN cache allocation algorithm where, as a function of content cache demands coming from CPs, the ICN provider decides the *imputation* of cache spaces to CPs. The second step uses a predefined *payment rule* by auctions to decide the selling price of the storage unit in the network; its purpose is to prevent content providers from lying about their true demand.

As cache imputations to CPs need to be fair and robust against over-claiming, we evaluate in this chapter common proportional and max-min fairness (PF, MMF) allocation rules, as well as coalitional game rules, the Nucleolus and the Shapley value. Results show that that the naive least-recently-used-based ICN approach provides proportional fairness. Moreover, the game-theoretic rules outperform in terms of content access latency the naive ICN approach as well as PF and MMF approaches, while sitting in between PF and MMF in terms of fairness.

6.2 Related Work on Cache Allocation in ICN

Several researches have recently proposed various cache allocation solutions for ICN. In this section, we review a selection of them.

Authors in [114] compare the in-network caching performance in homogeneous (i.e., where the routers have the same overall cache size) and heterogeneous cache deployments (i.e., where the routers have not the same cache size). In the latter

case, they propose to allocate cache capacity proportionally to the router's centrality metric measured according to different criteria: degree, stress, betweenness, closeness, graph and eccentricity centrality. They show that allocating cache capacity across the network in a heterogeneous manner slightly improves network performance compared to the homogeneous manner; however, the benefits of heterogeneous deployments become apparent with larger networks (e.g., more than 100 nodes). Moreover, authors in [115] study the influence of content popularity distribution on network performance, showing that (i) for uniformly distributed content demands (e.g., catch-up TV), pushing caches into the core yield better performance while (ii) highly skewed popularity request patterns (e.g., YouTube, mobile VoD system or Vimeo) are better served by edge caching. (ii) is also confirmed in [116].

Recently, there has been significant interest in applying game theory to the analysis of communication networks, with the aim to identify rational strategic solutions for multiple decision-maker situations. Indeed, as opposed to mono-decision maker problems, game-theoretic approaches adopt a multi-agent perspective to account for different objective functions and counter objections to rationally non justified solutions [71]. Thus far, many papers from the literature have tackled game-theoretic approaches in ICN using non-cooperative game theory. These papers consider servers or routers or networks as selfish entities seeking to maximize their own profit at the expense of globally optimum behavior. For example, authors in [117] study a non-cooperative game to characterize the problem of replication of contents by a set of selfish routers aiming to minimize their own costs. In the same context, authors in [118] characterize the caching problem among selfish servers without any central coordination using a non-cooperative game. For each content in the network, selfish servers have two possible actions: either caching the content if all its replicas are located too far away or not caching it if one of its replicas is located at a nearby node. As in [117], they show the existence of pure strategy Nash equilibrium of the caching game. Motivated by the intuition that forms of collaboration between different ICN nodes could yield an enhancement in network performance, authors in [119] propose a game whereby the routers behave as rational selfish agents that seek to minimize their aggregate content access cost. Going beyond routers, authors in [120] describe how content providers could shape their content access prices and discounts to favor the emergence of ICN distribution overlays across independent networks, toward the formation of incentive-prone ICN overlay equilibrium.

Under a similar rationale, yet a broader context, in this chapter we investigate how the ICN provider, modeling CPs as players of a game, can design an ICN cache allocation framework so that cache imputations to CPs are strategically fair and robust against cache space over-claiming, while outperforming legacy approaches in terms of content access latency. Up to our knowledge, there are no other works precisely addressing this problem, despite the above-cited works do share similar concerns in cache allocation and ICN component sharing. As detailed in the following, we propose various cache allocation rules, including coalitional game theory rules for bankruptcy situations [79] (please refer to Appendix A for more details on bankruptcy games) to solve the atomic cache contention problem, motivated by the fact that a similar algorithmic approach has shown high performances in strategic shared spectrum allocation problems (i.e., Wireless Mesh Networks [2] and femtocell networks [8]) as presented in Chapter 4.

6.3 ICN Cache Allocation Framework and Rules

In the context of an ICN network provider, ICN cache capacity is used to host content files in order to enhance users' quality of experience by decreasing content access latency. Assuming contents are owned by external CPs, the ICN provider would need to offer a neutral interface to access its caches, guaranteeing a fair allocation of caches with respect to cache space demands, which are in turn a function of content popularity. In this section, we formulate the problem, and then we detail the cache allocation algorithm and the corresponding ICN pricing framework.

6.3.1 Problem Formulation

Let us assume that there are n Content Providers (CPs), and each CP owns a given number of files. With the possibility to cache some files in the network between them and the users (by renting cache space from the ICN provider), the CPs can enhance their users' quality of experience by decreasing content access latency. Since different files can have different popularity, depending on how much cache space each provider is willing to pay for, the demand for a cache space by each content provider would not cover all its catalog's size. Let d_i be the cache space demand of the i^{th} CP, indicated in the following as CP-i, and $\vec{d}$ the vector of all demands.

Since the ICN provider's global cache space, denoted by E, is limited, we target the expected situation for an economically viable ICN deployment in which the ICN provider receives more demands than what it can satisfy, i.e., $\sum_{i=1}^{n} d_i \geq E$. If this was not the case, i.e., if the total demand is less than the available space, then the ICN provider would be able to allocate for every CP the exact space demanded, yet contention would likely still manifest for at least those few best nodes that are at the most attractive cross-points of users' demands (as far as these few best nodes would not be able alone to satisfy the whole demand).

In this context, there is a competition in accessing the ICN nodes' caches. From an ICN provider perspective, the risk is that CPs partially ally between each other, forming sub-coalitions when designing their respective demands. In order to master this behavior and avoid the formation of oligopolies, the ICN provider shall take into account the possible sub-coalitions in the allocation of cache sizes to CPs, designing an appropriate pricing framework. More precisely, the ICN provider has to:

1. decide on the *allocation rule*, i.e., how to assign cache space to which CP based on CPs' individual content cache size demands.

2. decide on the *payment rule*, i.e., how to fix prices for the allocated space given by step 1.

To emphasize the need of these two separate ICN provisioning rules, let us consider first the (unrealistic) case where the ICN provider announces that the space is given for free for the highest demand: every CP would then have an incentive to announce a very high demand, lying on the value of their real needs, to get free space. Suppose now another (more realistic, but naive) case with an announced fixed price per unit of cache: also in this case, because the space is limited, each CP has an incentive to announce a higher *untruthful* demand so that it can get more space. In order to avoid these situations, the ICN provider should design both steps

in advance to make sure that the outcome of the overall scheme is a desired one. For this purpose, we propose to adopt *mechanism design* theory's concepts [112]. In particular, we refer to approaches for single-dimensional environments to make sure that the allocation scheme provides strong performance guarantees (as explained hereafter, performance guarantees are based on fairness criteria), and at the same time it provides strong incentives for the CPs to be truthful in communicating their real demand.

The allocation and payment rules are interrelated in general. However, the mechanism design theory successfully deals with the two steps in a consecutive manner. First we suppose that the CPs are communicating their truthful demand. Based on these demands, we design a cache allocation scheme giving each CP its share of the limited resource E. Then, we design a payment rule for the CPs such that the dominant strategy for the CPs is to send their real demand (i.e., with no incentives to lie about it). Under this approach, the ICN provider can shape a strategic allocation making its ICN provisioning architecture rationally acceptable and attractive for additional CP customers.

6.3.2 Cache Allocation to Content Providers

An *allocation rule* is a function f having as an input the demands of the CPs (the demand vector $\vec{d} \in \mathbb{R}_+^n$) and the total available cache space $E \in \mathbb{R}_+$, and giving as output an *imputation vector* $\vec{x} \in \mathbb{R}_+^n$ containing the cache space portion to allocate to each CP (i.e., the values in $\vec{x}$ ranges between 0 and E such that $\sum_{i=1}^{n} x_i = E$), i.e., $f : (\vec{d}, E) \rightarrow \vec{x}$.

Let $\vec{d}_{-i}$ be the vector of demands of all the CPs other than i, so with a little abuse of notion, let us indicate the imputation for CP i as $x_i = f_i(d_i, \vec{d}_{-i}, E)$. For convenience, we also define $\bar{x}_i = x_i/E$ as the normalized imputation, i.e., the proportion of E allocated to CP-i. Let us give the following definition.

Definition 6.3.1. (Monotone Allocation Rule) An allocation rule is monotone if for each $(\vec{d}, E)$ and for each CP-i having $d'_i > d_i$, we have:

$$f_i(d'_i, \vec{d}_{-i}, E) \geq f_i(d_i, \vec{d}_{-i}, E), \ \forall \vec{d}_{-i} \tag{6.1}$$

In other words, fixing all the other CPs' demands $\vec{d}_{-i}$, if the demand of CP i increases from d_i to d'_i, then the imputation x_i should not decrease. Monotonicity plays an important role in designing the payment rule (we get back to this issue in Section 6.3.4).

The allocation of resources to those claiming higher demands than what is available is referred to in the literature as a *bankruptcy* problem (the term derives from the evident connection with the problem of bankruptcy where a person or other entity cannot repay the debts claimed by creditors). For this reason, we sometimes refer to the CPs as *claimants*, or the total available cache space to partition as the *estate*.

There are different possible approaches from the literature that can be used as allocation rules for a bankruptcy situation. We present thereafter the most common.

Allocation by Proportional Fairness (PF)

PF distributes the resources proportionally to the demands subject to total space constraint [121], i.e.,

$$\frac{f_i(\vec{d}, E)}{d_i} = \frac{f_j(\vec{d}, E)}{d_j} \quad \text{for any pair of CPs } (i, j).$$

It is straightforward to note that PF is monotone.

Allocation by Max-Min Fairness (MMF)

MMF maximizes the profit of the lowest claimant, then it maximizes the second lowest demand in the game, and so on [122]. Formally, if we order the CPs according to their increasing demand, i.e., $d_1 \leq d_2 \leq \cdots \leq d_n$, then MMF allocates the available space E as follows:

$$f_i(\vec{d}, E) = \min\left(d_i, \frac{E - \sum_{j=1}^{i-1} f_j(\vec{d}, E)}{n - i + 1}\right) \text{ for } i = 1, \ldots, n.$$

Intuitively, MMF gives the lowest claimant (assuming $d_i < E \ \forall i$) its total demand and evenly distributes unused resources to the other users. It is also straightforward to note that MMF is monotone.

Both MMF and PF allow computing fair imputations without considering the possibility that CPs could ally when formulating their demands. Alternatively, game theoretic allocation rules can be attractive toward the computation of a strategically fair imputation. Before presenting some game-theoretic allocation rules, let us formally define the bankruptcy game for our settings where the CPs are the players.

Allocation by Shapley Value

The Shapley Value [80], computed by averaging the marginal contributions of each player in the game in each strategic situation, is given by:

$$f_i(\vec{d}, E) = \sum_{S \subset \mathcal{N} \setminus \{i\}} \frac{|S|!(|\mathcal{N}| - |S| - 1)!}{|\mathcal{N}|!} [v(S \cup \{i\} - v(S)] \tag{6.2}$$

The Shapley value allocation rule is monotone since Eq. (A.12) can be rewritten as follows:

$$f_i(b_i, \vec{b}_{-i}, E) = \sum_{S \subset \mathcal{N} \setminus \{i\}} \alpha_S \phi_S(b_i), \tag{6.3}$$

where $\alpha_S = \frac{|S|!(|\mathcal{N}|-|S|-1)!}{|\mathcal{N}|!}$ and:

$$\phi_S(b_i) = \begin{cases} b_i & \text{if} \quad b_i \leq \max(0, E - \sum_{j \in \mathcal{N} \setminus \{S,i\}} b_j) \\ \max(0, E - \sum_{j \in \mathcal{N} \setminus \{S,i\}} b_j) & \text{otherwise.} \end{cases} \tag{6.4}$$

so by fixing $\vec{b}_{-i}$, $\phi_S(b_i)$ is non-decreasing function in b_i for any set S. Thus the Shapley value allocation is monotone.

Allocation by Nucleolus

The Nucleolus [83] that is computed by minimizing the largest excess of different coalitions of the game. The excess is expressed as:

$$e\left(f(\vec{d},E),S\right) = v(S) - \sum_{j\in S} x_j \ , \forall S \subseteq \mathcal{N} \tag{6.5}$$

This excess measures the amount by which the coalition S falls short of its potential $v(S)$ in the imputation $\vec{x}$ (Please refer to Appendix A for more details on Shapley value and Nucleolus computation). It is worth noting that Nucleolys is monotone thanks to the intrinsic consideration of individual rationality constraints.

6.3.3 Cache allocation algorithm

The total cache space in the network is formed from the collection of the routers' caches. These caches are distributed in heterogeneous locations in the network. For example, it might be more convenient for CPs to be allocated a cache space closer to the end users (thus their contents are closer to clients reducing content access latency). Therefore, it is important that the ICN network provider distributes a homogeneous cache space to CPs (every unit of cache space should have the same value from the content providers' perspective). In this respect, it should *cluster* routers that have similar properties for CPs. According to [114] and [115], three commonly accepted criteria for grouping the routers are: the proximity to the user-network edge, the router degree, and the router centrality (betweenness). More precisely, the *contention* metrics that we investigate are defined as follows:

- *Router Proximity to network edge (RP)*: the number of hops separating a router from network edge.
- *Router Degree (RD)*: the number of links incident to a router.
- *Router Betweenness (RB)*: the number of times a node is along the shortest path between two other nodes.

Upon ranking routers according to the contention metric, we propose the following allocation algorithm:

Algorithm 1 ICN Cache Allocation Algorithm

1: Form clusters of ICN routers with the same contention metric, and order them from the highest importance (in CPs' perspective) to the lowest one;
2: Take the cluster with the highest importance and apply the allocation rule to routers of the cluster;
3: Decrease the demand of each CP by the amount allocated in the cluster;
4: Take the next cluster and apply the allocation rule;
5: Stop when all clusters are treated or there is no remaining demand.

For the game-theoretic allocation rules, this corresponds in iterating a game $G(\mathcal{N}, v)$ differing in that, at each iteration:

- $\mathcal{N}$ includes all the CPs, but with different demands d_i.[2]
- The available cache size (E), varies as a function of the cluster size and the capacities of routers in the cluster. For instance, if the cache capacity of each router is given by C_r, the corresponding estate is given by: $E = \sum_{r \in \mathcal{R}_c} C_r$ where $\mathcal{R}_c$ is the set of routers in the cluster c.

It is worth noting that since the routers within the same cluster have the same contention metric, the allocated cache space to each CP in a cluster can be evenly allocated from any cache among the routers in that cluster.

Algorithmic game theory adds one more requirement to the design of the system: the complexity of obtaining the allocation should be computationally efficient. Polynomial time algorithms should be used for finding the allocation. Even though the complexity for calculating the Nucleolus or the Shapley values can be computationally hard (in particular the Shapley value), in our instance this does not cause a real problem because the number of CPs asking for the resource in a network is typically low (less than 10) and the complexity of the allocation scheme is a function of the number of CPs (and not a function of the potentially huge number of content files). So we can relax in our design the polynomial time algorithms requirement for computing the cache allocation and allow to use exponential time algorithms.

6.3.4 Pricing Framework

As already argued, a robust pricing framework needs to be designed by the ICN network provider to ensure true demands are formulated by CPs. Actually, the same unit of cache space can have different values for the different CPs (e.g., those with higher traffic consider a unit storage space to be more valuable). We assume that the value of a unit of cache space for a CP is proportional to its demand (e.g., a CP having a demand d_i evaluates the price of its unit of cache to be d_i/E).

Along with the fairness of the allocation scheme, the payment rule should be designed to give strong guarantees that the CPs are truthful in communicating their real demand. Under this perspective, it becomes natural to think of the demands as bids (as in auctions), and the cache partitioning as the allocation outcome from an auction. The demand vector is given by $\vec{d}$ where d_i is the (truth) demand by CP-i (also considered as the private value of i). The bid vector is given by $\vec{b}$ where b_i is the value communicated by CP-i to the ICN provider (could be equal to d_i if i declares the truth).

The truthful communication of demands should be a dominant strategy. This is known as the *dominant-strategy incentive-compatible* (DSIC) property [113]. The normalized allocation $\bar{x}_i$ is the proportion of the full available cache space allocated to content provider i (i.e., $\bar{x}_i$ ranges in the interval $[0, 1]$). The payment rule is given by $\vec{p}$, where p_i is the price of the allocation paid by CP-i. The utility of a content provider is given by:

$$U_i = V_i(d_i, f_i(b_i, \vec{b}_{-i}, E)) - p_i(b_i, \vec{b}_{-i}, E) \tag{6.6}$$

[2] As opposed to the algorithm presented in Chapter 4, all the players participate in every game while in Chapter 4, the number of players varies as a function of the interference level.

where $V_i(d_i, f_i(b_i, \vec{b}_{-i}, E)) = d_i \bar{x}_i$ is the value of the allocated space from the CP-i perspective, and $p_i(b_i, \vec{b}_{-i}, E)$ is the price paid. Every CP has the incentives to maximize its utility.

Definition 6.3.2 (DSIC)**.** The tuple $(\vec{x}, \vec{p})$ is DSIC if: 1) each truth-telling CP is guaranteed a non-negative utility and 2) each CP has as dominant strategy the communication of its truthful demand, i.e., for all CPs, and for any b_i,

$$V_i(d_i, f_i(d_i, \vec{b}_{-i}, E)) - p_i(d_i, \vec{b}_{-i}, E) \\ \geq V_i(d_i, f_i(b_i, \vec{b}_{-i}, E)) - p_i(b_i, \vec{b}_{-i}, E).$$

Then, the tuple $(\vec{x}, \vec{p})$ is DSIC if when $b_i = d_i$, this strategy maximizes the utility of CP-i no matter what the other CPs do. Being that the utility $U_i = d_i \bar{x}_i - p_i$, for example with the pricing rule $p_i = b_i \bar{x}_i$, no one has an incentive to communicate the true demand. Because with that pricing rule, the utility would be $U_i = 0$ for truth-tellers while it can be increased if everyone declared a slightly lower demand. This would lead to a situation where everyone declares a lower demand than their real one. On the other hand, for a fixed price per storage space (i.e., $p_i = \alpha \bar{x}_i$ for a given $\alpha \in \mathbb{R}_+$) every CP having $d_i > \alpha$ has an incentive to increase its communicated demand (b_i) to receive more space increasing its utility. The question that can be raised here: what pricing rule should we select so that the CPs have no incentives to lie (given the Shapley and the Nucleolus-based allocation rules)? It turns out that by Myerson's Lemma [123] from mechanism design theory we can design the prices to meet our objective:

Theorem 6.3.3 (Myerson's Lemma [123])**.** *If $\vec{x}$ is monotone, then there is a unique payment rule $\vec{p}$ such that the mechanism $(\vec{x}, \vec{p})$ is DSIC.*

The monotonicity is given by Definition 6.3.1, and the four presented allocation rules are monotone as already discussed.

Finding this payment rule requires a bit more analysis. Since for a given CP i, the characteristic function of the game (Eq. (4.2)) is piece-wise linear as function a demand b_i, then it follows directly for Shapley value that the allocation $f_i(b_i, \vec{b}_{-i}, E)$ (given by Eq. (A.12)) is piece-wise linear as function of b_i. Similar argument for Nucleolus shows that it is piece-wise linear as a function of b_i. This means that the allocation as function of the demand looks as in Fig. 6.2. The price can be then given by Myerson's Lemma [123] as follows:

$$p_i(b_i, \vec{b}_{-i}, E) = b_i f_i(b_i, \vec{b}_{-i}, E) - \int_0^{b_i} f_i(z, \vec{b}_{-i}, E) \mathrm{d}z \tag{6.7}$$

The price can be interpreted as an area above the curve (as given by Fig. 6.2). Notice that by considering this pricing rule, each content provider maximizes its utility U_i by communicating its true demand no matter what others do, i.e., U_i is maximized when $b_i = d_i$ for every $\vec{b}_{-i}$.

In fact, for the Shapley value allocation, we can identify precisely the points where the curve changes its slope, in order to determine the price; equation (6.3) can be reformulated using (6.4) as follows:

$$f_i(z, \vec{b}_{-i}, E) = g(\vec{b}_{-i}) + \left(\sum_{T \in \mathcal{T}} \alpha_T \right) z \tag{6.8}$$

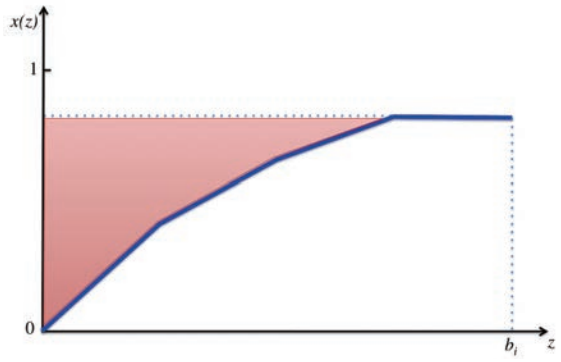


Figure 6.2: The blue curve is the piecewise-linear allocation function $\bar{x}_i$ given by $x(z)$ for CP i when varying its demand from 0 to b_i (z axis). The red area is the payment rule (price to pay by the CP).

where $g(\vec{b}_{-i})$ is a scalar function independent of z and $\mathcal{T}$ is a relevant set of the sets $T \subseteq \mathcal{N}\backslash\{i\}$. Define the vector $\Phi(\vec{b}_{-i})$ having in decreasing order the elements $\max(0, E - \sum_{j\in\mathcal{N}\backslash\{S,i\}} b_j)$ for all sets $S \subseteq \mathcal{N}\backslash\{i\}$. For a given value $z \in [0, b_i]$, let us define the index k to be the first index in $\Phi(\vec{b}_{-i})$ such that $\Phi_k(\vec{b}_{-i}) > z$ and $\Phi_{k+1}(\vec{b}_{-i}) \leq z$. Therefore, for a given z, $\mathcal{T}$ contains the sets $S_1, \ldots, S_k$ corresponding to the top k elements in Φ. The slope at any value z is then given by $\sum_{j=1}^{k} \alpha_{S_j}$, and so it is non-increasing when increasing z. The points where the curve changes its slope are the values of z for which a transition in the index k takes place.

As a result of the discussion in this section, the ICN provider can declare a pricing accordingly to (6.7) to all the CPs, so that none of the CPs has an incentive to declare a different demand than their real one, and based on these (truthful declarations) the allocation using the proposed ICN allocation algorithm is carried out. It is important to note that this pricing framework does not necessarily maximize the profit for the ICN provider, but it is the *unique* pricing rule [123] that provides strong incentives for truthful declaration of demands by the CPs. Any other pricing rule can cause the CPs to communicate false demands to maximize their utilities.

6.4 Performance Evaluation

We consider a network composed of 25 routers of same caching capacity C (i.e., homogeneous cache size). We considered two networks with a tree (where there is only one path from an end-user to a CP) and a partial mesh (where there can be multiple paths from an end-user to a CP), having both an edge-to-CP shortest path length up to 6 hops. In the tree topology, the CPs are connected to the root router of the tree, and the end-users are connected to its leaves. In the partial mesh, the CPs are all connected to one router in the network, while the end users are connected randomly to some of the other router nodes of the network.

We include 5 CPs, denoted CP-i for $i = 1, \ldots, 5$, connected all to the same router and each supplying different contents/files. We assume that each content j has a uniform size (1 MB for example) and a popularity $P_j \in [0; 1]$ reflecting the

request's frequency made by end-users for the content (i.e., the number of times end users issue 'interest' messages to retrieve the content) [107]. Contents are always delivered via a shortest path. In the simulations, the popularity is chosen uniformly at random such that the sum of all files' priorities in the network is equal to 1, i.e., $\sum_j P_j = 1$. We recall we assume that each content is offered by only one CP.

To take into account network cases with a heterogeneous set of demands, we suppose that, among the five CPs, the CP-1 has the lowest demand d_1, and that CP-2, CP-3, CP-4, and CP-5 have, respectively, three, five, seven and nine times the demand of CP-1. The overall demand of CP-1 is set to 80 files (i.e., 80 MB). The *Contention Level* in the network is then computed as: $C_L = 1 - (25C/\sum_{i=1}^{5} d_i)$.

The popularity of contents in the network is modeled using the Zipf's law [124] that determines its frequency of occurrence. Each CP runs the LRU cache replacement policy that we approximate using the Che approximation [125].

We do compare the results under different allocation rules also for the case of a network without in-network caching, i.e., in which end user requests need to go all the way up from the edge to the CP containing the needed file at the network provider CP edge. Moreover, for the in-network caching cases, we include a naive ICN case in which there is no router clustering and there is no CP-specific cache allocation [109]; instead, contents are delivered following the shortest path and cached on-the-fly by the LRU caches collocated on the traversed routers. As a reminder, we evaluate the four allocation schemes listed in Section 6.3.2: PF, MMF, Shapley Value, and Nucleolus. The following evaluation focuses on a performance analysis based on content access latency reduction and on a fairness analysis based on cache imputations.

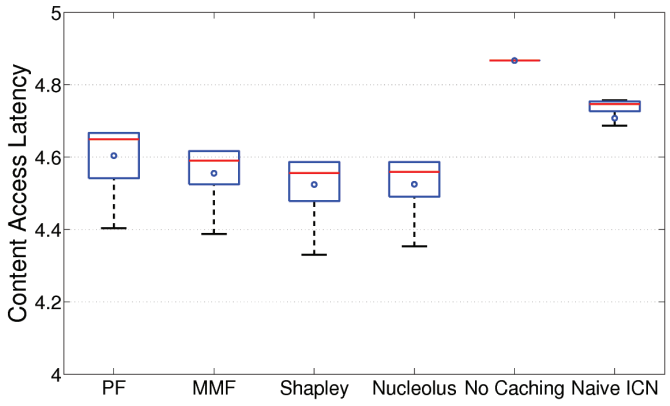
6.4.1 Content Access Latency

We evaluate the performance of different approaches with respect to the most important user's quality of experience metric for ICN technologies, i.e., the content access latency. We compute the average content access latency (expressed in number of hops) as a function of the edge-to-content path, and the average hit ratio on each router along the path as given by the Che approximation [125]. To model the case of high cache contention situation, we set C_L to 80% (i.e., the total cache space is equal to only 20% of the total CPs demands).

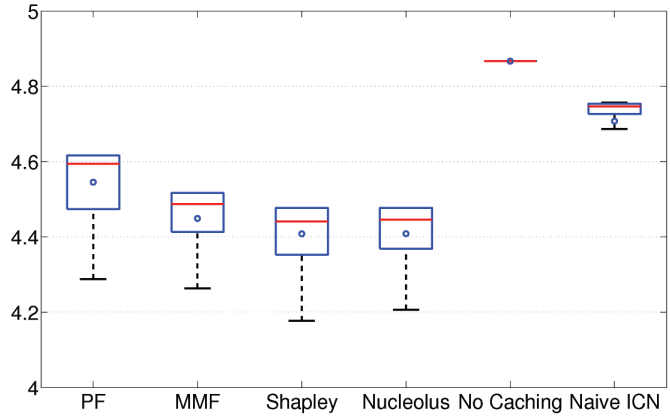
Fig. 6.3 and Fig. 6.4 show the boxplot statistics (max, min, quartiles, median as a red line, average as a star) of the content access latency for the network's contents using the above mentioned metrics for the tree and partial mesh topology, respectively. We can notice that:

- Comparing in-network caching approaches to the one without caching, the former outperform the latter one for all the cases; e.g., for the partial mesh topology and using the RD metric, the median content access latency decreases, from the approach without caching, by 9% with the game-theoretic approaches, 8% for the MMF, 4% with PF, and 2.5% for the basic ICN case.

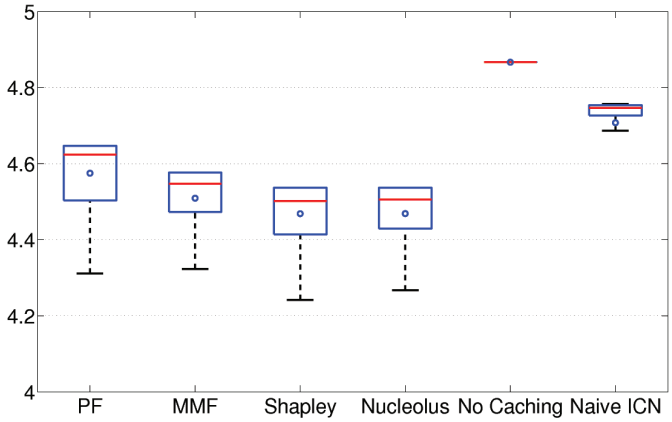
- Comparing the basic ICN approach to the router aggregation case with the four allocation rules, the content access latency decreases with the latter one for all the cases (e.g., for the partial mesh topology and RD metric, the median



(a) Router Proximity to network edge (RP)

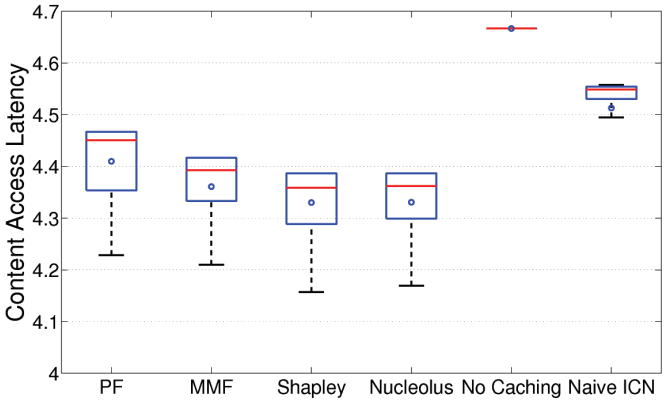


(b) Router Degree (RD)

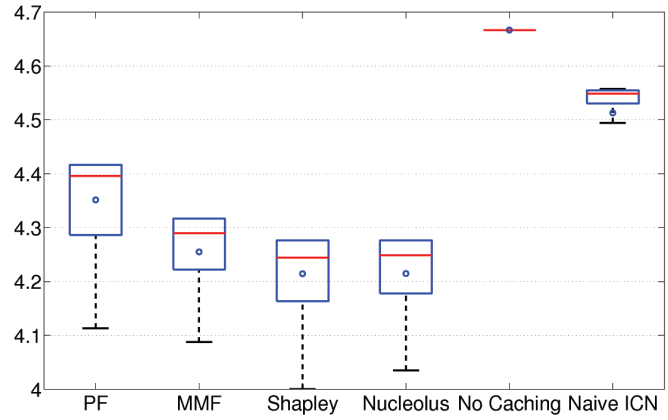


(c) Router Betweenness (RB)

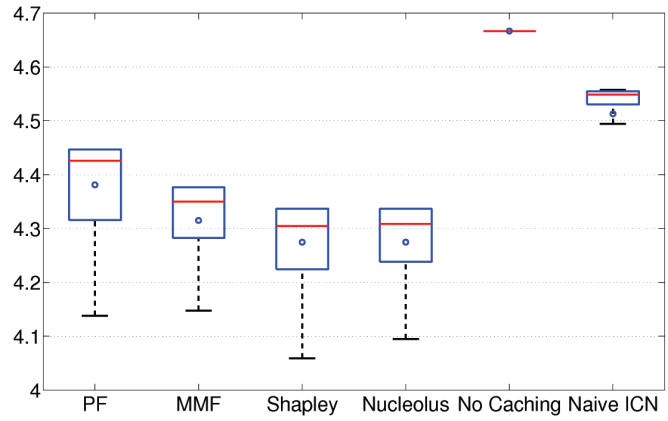
Figure 6.3: Content access latency distributions for a tree topology and with different ICN router clustering metrics.



(a) Router Proximity to network edge (RP)



(b) Router Degree (RD)



(c) Router Betweenness (RB)

Figure 6.4: Content access latency distributions for a partial mesh topology and with different ICN router clustering metrics.

access latency decreases from basic ICN by 3% with PF, 5.6% with MMF and 6.5% for game-theoretic approaches).

- The game-theoretic approaches, Nucleolus and Shapley value, give very close performances for the different cases. They outperform the PF and MMF approaches for all the cases; e.g., for the tree topology and using the RB metric, the median content access latency is lower by 2.5% with respect to PF, and by 1.6% with respect to MMF.

- The partial mesh topology outperforms the tree one, likely because it allows multiple paths between network's routers differently than the tree topology with a single path from each router to the root.

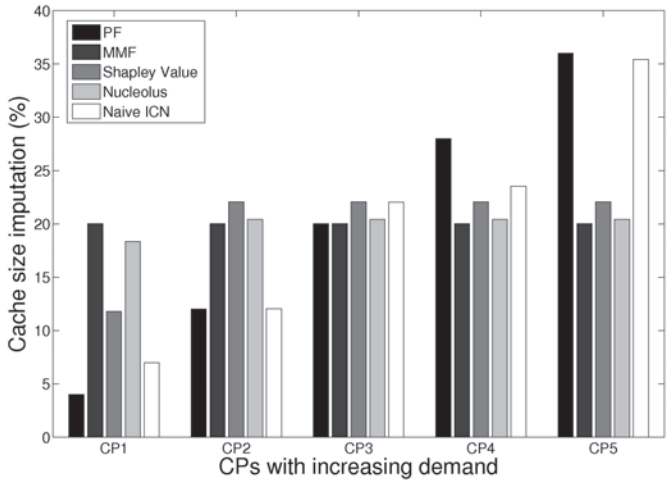
- The RD router clustering metric outperforms the other metrics for all the in-network caching cases; e.g., in the mesh topology, the content access latency for the Nucleolus decreases from the RP by 3% and 1.25% to the RD and RB metrics, respectively. This somehow confirms previous findings of [114] where RD was shown to be superior to all other metrics. As a new insight, the gain of RD with respect to RB is less important than with respect to RP.

All in all, these highlights show that game-theoretic approaches increase content access performance. It is also worth mentioning that even if naive LRU driven in-network caching permits to reduce latency, it does not accomplish as much one could expect, mostly because of the potentially high replication of contents in the network [126].

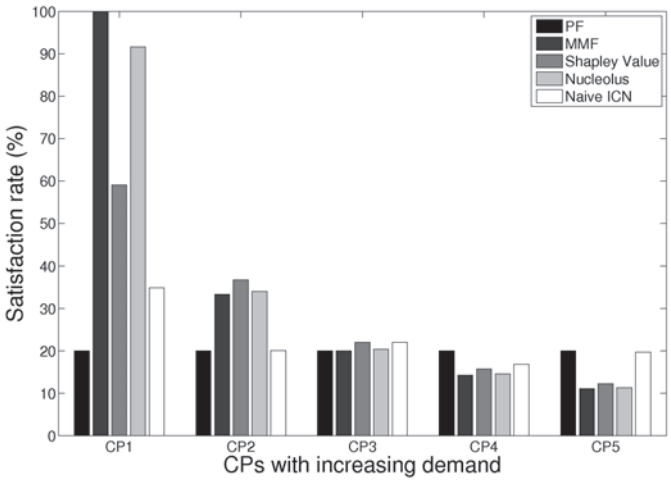
6.4.2 Fairness of Cache Imputations

In order to further investigate on the cache allocation results, Figure 6.5 shows the imputation distribution (i.e., the ratio of the cache each CP obtains as a function of the total available cache) as well as the satisfaction rate (i.e., the ratio of the cache each CP obtains as a function of its demand), for the different allocation cases (PF, MMF, Nucleolus, Shapley value, and naive ICN). The partial mesh topology with the RD metric case is considered (similar results are obtained for the tree topologies). We can observe that the Nucleolus and Shapley value give the lowest claimant (i.e., CP-1) an imputation in-between those obtained by PF and MMF: CP-1 gets by Nucleolus and Shapley value 18% and 11% respectively of the total estate, while PF and MMF give respectively 5% and 20% of the total estate (20% corresponds actually to the totality of its demand, indeed the satisfaction rate of CP-1 is 100% with MMF). The same behavior can be seen also for the highest claimant (CP-5) whose imputation by Nucleolus and Shapley value is in-between those of MMF and PF. This indicates that game-theoretic approaches do not favor low demands as MMF does, or high demands as PF does, but instead distribute the estate in a way that discourages too greedy demands at the benefit of lower demands.

It is also worth noting that the naive approach with ICN is closer to the PF approach than the others. Intuitively, this can be explained by the fact that as the claim increases, the probability of finding claimant's files in the network likely proportionally increases.



(a) Cache size distribution



(b) Satisfaction rates

Figure 6.5: Cache size distribution and satisfaction rates, as a function of the CP demand, for a partial mesh topology using the RD metric.

Furthermore, in order to qualify the fairness of the solutions, we evaluate them with respect to two notable fairness indexes: the Jain's fairness index (J_I) [89] that rates the fairness of a set of values and defined as:

$$J_I = \left(\sum_{i=1}^{n}(x_i/d_i))^2\right) / \left(n\sum_{i=1}^{n}(x_i/d_i)^2\right) \tag{6.9}$$

which in fact has been conceived to be better the closer the solution is to the PF, and the Atkinson's index (A_I) [127] which is one of the commonly used measure of inequality, computed as follows:

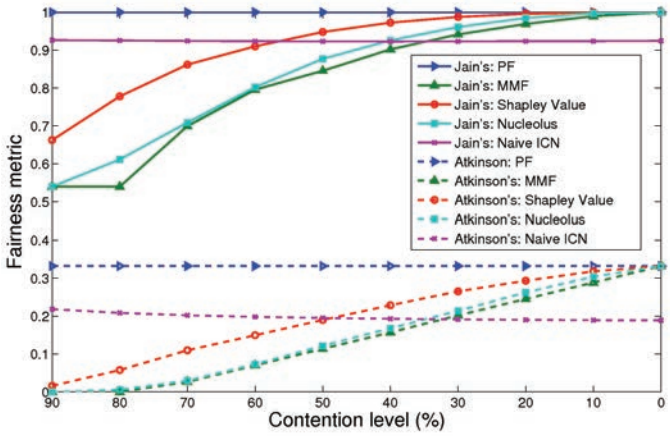


Figure 6.6: Fairness indexes as a function of the contention level (the lower the contention level, the higher the available cache size with respect to demands), for different allocation rules.

$$A_I = 1 - \frac{n}{\sum_{i=1}^{n} x_i} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{(1-\epsilon)} \right)^{1/(1-\epsilon)} \tag{6.10}$$

which conversely has been conceived to be better the closer the solution is to an even division ($A_I = 0$ means perfect equality while $A_I = 1$ expresses maximal inequality). ϵ is chosen in practice between 0.5 and 1.5 (we set a value of 1.5 in our case). Figure 6.6 shows the fairness index results, as a function of the contention level C_L. We can state that:

- Fairness indexes confirm the close behavior between naive ICN and PF. Both appear as independent of the contention level - PF gives the best for the Jain's index and the worst for the Atkinson's one, and naive ICN gives better Atkinson's index values than PF.
- Comparing the Nucleolus and the Shapley value for both metrics, the latter is strictly the closer one to the PF, while the former is closer to MMF. The gap between them, PF and MMF strictly decreases as the contention level decreases.

Overall, depending on the desired fairness behavior, PF or MMF, the network provider can refer to the Shapley as the one closer to PF, and the Nucleolus closer to MMF, being reassured about the fact that they bring a gain in terms of content access latency. Simply using the naive ICN approach would be a good approximation of the PF rule, with however a lower content access performance.

6.5 Summary

Novel technologies are difficult to adopt as it has to be proven that they are incentive compatible for all the involved stakeholders. In this chapter, we address a multi-

stakeholder situation (i.e., involving more than one provider) that appears as a win-win setting toward ICN deployment, i.e., the case of an Internet Network Service Provider deploying ICN for external content providers, offering a neutral interface and pricing to multiple content providers. The ICN provider hence allocates to external content providers spaces in its ICN routers' caches for content delivery.

In this context, we argue that the proper way the ICN provider shall design the ICN cache allocation framework and model the behavior of external content providers is game theory, so as to qualify and counter-balance their natural tendency to form oligopolies and to ally to have a stronger position in getting the available caching resources. We investigate the application of well-known concepts from cooperative game-theory showing desirable properties, the Nucleolus and the Shapley value, as well as other principles commonly adopted in networking research, the proportional fairness (PF) and the max-min fairness (MMF). We propose an ICN cache allocation algorithm able to incorporate these different allocation rules applying them to clusters of routers ordered with respect to centrality metrics suggested in the literature. Moreover, we propose a pricing framework that, taking advantages of the monotonicity of the presented cache allocation rules, correctly nullifies the threat of malicious behaviors in formulating content caching demands.

Results from simulations show that the game-theoretic approaches offer a (not so straightforward) sensible access latency gain with respect to both PF and MMF, and the naive ICN approach (without cache allocations and using least-recently-used cache management) to content providers. Among the Nucleolus and the Shapley value approaches, the former could be considered more interesting given its lower time complexity (asymptotically with the number of content providers). In terms of fairness the Nucleolus and the Shapley value sit in-between PF and MMF allocation rules, balancing their well-known weaknesses and strengths, so that the Shapley value is close to PF and the Nucleolus very close to MMF. It is also valuable to report that the naive ICN approach permits to approximate PF without having to compute cache imputation (at the expense, however, of worse content access performance).

Chapter 7

Conclusion and Perspectives

At present, mobile phones provide the best means of gathering information about user movements and content consumption behavior on a large scale. This dissertation aims to find solutions at the edge (frontier) between resource and traffic management, usage profiling and user mobility detection.

Using some mobility properties from big data log coming from cellular networks, we evaluate in Chapter 2 the appropriateness of using mobile data for estimating human mobility and behavior patterns. Overall this chapter clearly shows that the use of such mobility metrics allows an efficient qualification of user mobility patterns. Using the same mobility metrics, we characterize in Chapter 3 the correlation between mobility and usages in mobile networks and we present an in-depth study on how content and Cloud delivery points in an urban and peri-urban environments can be identified and estimated online. This study permits to estimate, with a high precision, the position of crowded cells in a cellular network, thus allowing to optimize the placement of content and computing resources and grant important traffic offloading, improve network efficiency and user quality of experience.

In this context, in order to enhance traffic offloading and better meet the increasing mobile Internet demand, while addressing the lack of available mobile spectrum and the expense of new infrastructure, we study in Chapter 4 and Chapter 5 two traffic offloading solutions, one based on small-cell networks and the other on WiFi-certified Passpoint hotspots. We study benefits and limitations of each solution and we propose a new algorithm based on cooperative game theory for resource allocation and interference management in small-cell networks. The results of Chapter 4 show important improvements and emphasize the necessity of referring to cooperative game theory concepts in the definition of spectrum contention situations. Furthermore, Chapter 5 clearly identifies the benefits of using the newly released WiFi Passpoint solution in increasing the spectrum capacity gain and decreasing the user energy consumption.

We also investigate methods to offload the backhauling wireline network in Chapter 5, where we propose an advanced approach to solve contention situations in

Information Centric Networking. This chapter shows the benefits of relying on coalitional game approaches to improve legacy solutions, by decreasing the content access latency in high contention situations.

Each proposition brings novelties to the corresponding research area. Chapters 2 and 3 find novel, scalable and effective ways to spot where to offload the access network as a function of user mobility and consumption patterns. Chapters 4, 5 and 6 brings original offloading solutions, horizontal and vertical wireless traffic offloading, and content offloading in backhauling network using pioneering solutions from cooperative game theory.

Despite the benefits of our proposed approaches in solving the current networking problems of traffic overloading and mobile data services quality, many open questions come to mind. Are these long-term solutions? Should mobile operators and vendors rely on such solutions for future networks?

As a matter of fact, to cater the explosive data traffic growth in 5G mobile networks while ensuring a high quality mobile data services, there exists nowadays a critical need for increased service flexibility. In order to enhance the system performance and reduce the cost associated with it, mobile operators started to shift their mindsets from the traditional Radio Access Network (RAN) architecture to the Cloud RAN architecture [128] using virtualization solutions with forms of centralization of the control plane and virtualization of the data plane.

Indeed, if virtualization is used for base station systems, service providers can dynamically allocate processing resources within a centralized baseband pool to different virtualized base stations and different air interface standards. This leads to multiple benefits: from one hand, it allows the operator to efficiently support the variety of air interfaces and to adjust to the tide effect in different areas and fluctuating demands; on the other hand, the common hardware platform can provide cost effectiveness to manage, maintain, expand and upgrade the base station [129] [130]. C-RAN is typically thought of as a large-scale urban macro solution, but the concept of pooled baseband serving a number of radio access nodes can also be applied to a variety of scenarios, such as small-cell underlays (using micro Remote Radio Units (RRUs), so-called Super Cells. These models, identified and defined partly through the Next Generation Mobile Networks (NGMN) Alliance, could prove an attractive way to introduce and develop C-RAN technology [131]. In the context of finding attractive solutions for catering the increasing data demand in mobile networks, we are interested in cloud-based access network solution, emerging nowadays as an important platform for next-generation radio access networks.

Furthermore, as mobile systems have limited resources such as battery life, network bandwidth, storage capacity and processor performance, we are interested in investigating cloud (computation) offloading solutions (e.g., CloudLet [132] [133]), enabling a significant saving of energy in mobile devices and adaptive virtual machine migration solutions enabling the migration of a virtual machine (VM) from

one physical server to another hence leading to more efficient resource utilization and load balancing. Besides these new technologies, the trend nowadays is moving towards a new digitalized world where users are associated with virtual resources, able to move and to resize, as a function their mobility habits.

It is worth mentioning that I passed more than sixteen months in different laboratories as a visiting researcher: nine months at MIT, six months at TU Berlin, almost one month at INRIA Sophia Antipolis, and many days at Orange labs. These visits allowed me to profit from the expertise of many persons for my researches, who helped in the preparation of the related works and publications. Therefore, my papers often include a high number of co-authors, to correctly acknowledge the contribution and time spent on my research of each co-author, while noting that the main contributions were originally proposed by myself for the first time, represented my own original contribution, which is formally confirmed by the first authorship accordingly to common international best practice in networking research.

Last words are needed to further acknowledge the researchers, and their respective teams, for their high added-value comments and suggestions to the dissertation. In order of content exposition:

Stanislav Sobolevsky and Carlo Ratti from Massachusetts Institute of Technology MIT on mobile data mining, presented in Chapter 2 and Chapter 3; Adam Wolisz from the Technical University of Berlin on traffic offloading over Passpoint hotspots presented in Chapter 5; Mahmoud el Chamie and Damien Saucez from INRIA Sophia Antipolis on the content distribution solution in ICN, presented in Chapter 6.

References

Related Own References

International Journals with peer review

[1] S. Hoteit, S. Secci, S. Sobolevsky, C. Ratti and G. Pujolle, "Estimating Human Trajectories and Hotspots through Mobile Phone Data", *Elsevier Computer Networks*, Vol. 4, pp. 296-307, 2014.

[2] S. Hoteit, S. Secci, R. Langar and G. Pujolle, "A Nucleolus-based Approach for Resource Allocation in OFDMA Wireless Mesh Networks", *IEEE Transactions on Mobile Computing*, Vol. 12, pp. 2145-2154, 2013.

International conferences with peer review

[3] S. Hoteit, S. Secci, G. Pujolle, S. Wietholter, A. Wolisz, C. Ziemlicki and Z. Smoreda, "Quantifying the Achievable Cellular Traffic Offloading Gain with Passpoint Hotspots", in *Proc. of the ACM International Workshop on Wireless and Mobile Technologies for Smart Cities*, Philadelphia, PA, USA, 2014.

[4] S. Hoteit, S. Secci, G. Pujolle, V. Hoa LA, C. Ziemlicki and Z. Smoreda, "Mobility-Aware Estimation of Content Consumption Hotspots for Urban Cellular Networks", in *Proc. of IEEE/IFIP Network Operations and Management Symposium NOMS*, Krakow, Poland, 2014

[5] S. Hoteit, S. Secci, S.Sobolevsky, G. Pujolle and C. Ratti, "Mobility-Aware Estimation of Content Consumption Hotspots for Urban Cellular Networks", in *Proc. of Int. conference on the Analysis of Mobile Phone Datasets (NetMob)*, Cambridge, MA, USA, 2013.

[6] S. Hoteit, S. Secci, S.Sobolevsky, G. Pujolle and C. Ratti, "Estimating Human Trajectories through Mobile Phone Data", in *Proc. of IEEE Int. Conf. Mobile Data Management- Human Mobility Computing Workshop*, Milan, Italy, 2013.

[7] S. Hoteit, S. Secci, G. Pujolle, Z. He, C. Ziemlicki, Z. Smoreda and C. Ratti, "Content Consumption Cartography of the Paris Urban Region using Cellular Probe Data", in *Proc. of ACM URBANE 2012, CoNext Workshop*, Nice, France, 2012.

[8] S. Hoteit, S. Secci, R. Langar, G. Pujolle and R. Boutaba, "Bankruptcy Game Approach for Resource Allocation in Cooperative Femtocell Networks", in *Proc. of IEEE Global Communications Conference (IEEE GLOBECOM)*, Anaheim, CA, USA, 2012.

[9] S. Hoteit, S. Secci, R. Langar and G. Pujolle, "Subchannel Resource Allocation for Cooperative OFDMA Wireless Mesh Networks", in *Proc. of IEEE Int. Conference on Communications (IEEE ICC)*, Ottawa, Canada, 2012.

Thesis

[10] S. Hoteit, "Strategic Resource Allocation in Cooperative Femtocell Networks", Master Thesis, Sept. 2011. Advisors: Stefano Secci and Guy Pujolle.

[11] S. Hoteit, "MaxMin Clusterisation Algorithm for Wireless Sensor Networks", Engineering Diploma Graduate Thesis, July 2010. Advisors: Mahmoud Doughan, Michel Marot and Monique Becker.

Submitted

[12] S. Hoteit, M. Premoli, S. Secci, G. Pujolle, C. Ziemlicki and Z. Smoreda, "Territory-based Hotspot Estimation for Urban Cellular Networks", *submitted to Journal of Network and Systems Management* `http://www-phare.lip6.fr/~hoteit/jnsm2014.pdf`

[13] S. Hoteit, S. Secci, G. Pujolle, S. Wietholter, A. Wolisz, C. Ziemlicki and Z. Smoreda, "Mobile Data Traffic Offloading over Passpoint Hotspots", *submitted to Computer Networks*. `http://www-phare.lip6.fr/~hoteit/computernetworks2014.pdf`

[14] S. Hoteit, M. El Chamie, D. Saucez and S. Secci, "Strategic Cache Allocation in Information Centric Networking", *submitted to IEEE Infocom 2015*. `http://www-phare.lip6.fr/~hoteit/infocom2015.pdf`

Other References

[15] R. Becker, R. Caceres, K. Hanson, S. Isaacman, J.M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, C. Volinsky, "Human Mobility Characterization from Cellular Network Data", *Magazine Communications of the ACM*, Vol.56, pp.74-82, 2013.

[16] "L'audience de l'Internet mobile en France", Mediametrie, http://www.mediametrie.fr/internet/communiques/l-audience-de-l-internet-mobile-en-france-en-mars-2014.php?id=1070#.U6Kh35S1Z3t, 2014.

[17] M. Meeker, S. Devitt and L. Wu, "Morgan Stanley", http://www.morganstanley.com/institutional/techresearch/pdfs/Internet_Trends_041210.pdf, 2011.

[18] J. Denny, "Search, Social and the Digital Marketing Revolution", http://johnhdenny.com/1474/primer-on-the-state-of-the-internet-2012, 2012.

[19] F. Calabrese, F. Pereira, G. Di Lorenzo, L. Liu and C. Ratti, "The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events", *In Proc. of International Conference Pervasive*, Helsinki, Finland, 2010.

[20] Cisco, San Jose, CA, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017, *http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf*, Feb. 2013.

[21] Cisco Visual Networking Index Forecast Projects "Cisco Press Release", http://newsroom.cisco.com/release/1135354, 2013.

[22] J. Giles, "Smartphone use makes cellular networks: collapse a real possibility", The Washington Post, http://www.washingtonpost.com/wpdyn/content/article/2010/11/29/AR2010112904854.html, 2010.

[23] D. Srinivasan, J. Dey, S. Kumar and R.N. Mukherjee, "Data Offload Approches for Mobile Operators: Improving Network Efficiency And Strengthening Quality Of Service", *Wipro Technologies 2012.*

[24] Femto Forum, "Femtocells- Natural Solution for Offload", *www.femtoforum.org*

[25] S. Eubank, H. Guclu, VS. Anil Kumar, MV. Marathe MV and A. Srinivasan, "Modelling disease outbreaks in realistic urban social networks", http://www.washingtonpost.com/wpdyn/content/article/2010/11/29/AR2010112904854.html, 2010.

[26] EL. Glaeser and ME. Kahn, "Sprawl and urban growth", *Handbook of Regional and Urban Economics. Elsevier*, pp. 2481-2527, 2004.

[27] T.Leighton, "Improving performance on the internet", *Communications Magazine of the ACM*, Vol.52, pp. 44-51, 2009.

[28] Q. Hao, R. Cai, C. Wang, R. Xiao and JM. Yang, "Equip tourists with knowledge mined from travelogues", *in Proc of the international conference on World Wide Web WWW*, New York, USA, 2009.

[29] Y. Zheng, L. Zhang, X. Xie and Y. Ma, "Mining interesting locations and travel sequences from gps trajectories", *in Proc of the international conference on World Wide Web. WWW*, New York, USA, 2009.

[30] M. Gonzalez, CA . Hidalgo and Al. Barabasi, "Understanding individual human mobility patterns", *Nature,* Vol. 458, pp. 238-238, 2008.

[31] H. Hohwald, E. Frias-Martinez and N. Oliver "User modeling for telecommunication applications: Experiences and practical implications", *in Proc. of the international conference on User Modeling, Adaptation and Personalization, UMAP*, Berlin, Germany, 2010.

[32] R. Huerta and L. Tsimring, "Contact tracing and epidemics control in social networks", *Physical Review E.66* , 056115, 2002.

[33] P. Wang, MC. Gonzalez, CA . Hidalgo and Al. Barabasi, "Understanding the spreading patterns of mobile phone viruses", *Science* Vol. 324, no. 5930, pp. 1071-1076, 2009.

[34] M. Turner, S. Love and M. Howell, "Understanding emotions experienced when using a mobile phone in public: The social usability of mobile (cellular) telephones", *Telematics and Informatics*, Vol. 25, no. 3, pp. 201-215, 2008.

[35] R.C. Nickerson, H. Isaac and B. Mak, "A multi-national study of attitudes about mobile phone use in social settings", *International Journal of Mobile Communications*, Vol. 6, no.5, pp. 541-563, 2008.

[36] M. R. Vieira, V. Frias-Martinez, N. Oliver and E. Frias-Martinez, "Characterizing dense urban areas from mobile phonecall data: Discovery and social dynamics", in *Proc. of the international conference on Social Computing - SocialCom*, 2010.

[37] H. Wang, F. Calabrese, G. Di Lorenzo and C. Ratti, "Transportation mode inference from anonymized and aggregated mobile phone call detail records", *Proc. of IEEE international Conference on Intelligent Transportation Systems (ITSC)*, 2010.

[38] H. Zang and J. Bolot, "Mining call and mobility data to improve paging efficiency in cellular networks", in *Proc. of ACM Int. Conf. on Mobile Computing and Networking (ACM MOBICOM)*, New York, USA, 2007

[39] L. Wei, Y. Zheng and W. Peng, "Constructing Popular Routes from Uncertain Trajectories", in *Proc. of the international conference on Knowledge discovery and data mining SIGKDD*, New York, USA, 2012.

[40] K. Zheng, Y. Zheng, X. Xie and X. Zhou "Reducing Uncertainty of LowSampling-Rate trajectories", *In IEEE International Conference on Data Engineering, ICDE, 2012.*

[41] T. Hagerstrand, "What about people in regional science?", *Papers in Regional Science*, Vol. 24, no. 1, pp. 6-21, 1970.

[42] S. Winter and Z.C. Yin, "Directed movements in probabilistic time geography", *International Journal of Geographical Information Science*, Vol. 24, no. 9, pp. 1349-1365, 2010.

[43] M. Ficek and L. Kencl, "Inter-Call Mobility Model: A Spatio-temporal Refinement of Call Data Records Using a Gaussian Mixture Model", *In Proc. of IEEE INFOCOM*, Orlando, FL, USA, 2012

[44] I. Rhee, M. Shin, S. Hong, K. Lee, S.J. Kim and S. Chong, "On the levy-walk nature of human mobility", in *Proc. of INFOCOM*, Phoenix, AZ, USA, April 2008.

[45] Airsage: Airsage WISE technology, http://www.airsage.com.

[46] C. Schneider, T. Couronne, Z. Smoreda and M. Gonzalez, "Are we in our travel decisions self-determined?", *Bulletin of the American Physical Society*, APS, 2012.

[47] R. H. Guting and M. Schneider, *Moving Objects Databases*, Morgan Kaufmann, 2005.

[48] C. S. Yang, S. P. Kao, F. B. Lee and P . S . Hung, "Twelve different interpolation methods: A case study of Surfer 8.0", in *Proc. of the ISPRS*, 2004.

[49] F.N. Fritsch and R. E Carlson, "Monotone piecewise cubic interpolation", *SIAM Journal of Numerical Analysis*, Vol. 17, no. 2, pp. 238-246, 1980.

[50] 5GPPP Association "What will the 5G-Infrastructure-PPP deliver?" `http://5g-ppp.eu/kpis/`.

[51] TK. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots", *Accident Analysis and Prevention*, Vol. 41, no. 3, pp. 359-364, 2009.

[52] J. Yuan, Y. Zheng and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs.", In *Proc. of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York, 2012.

[53] K. Seada, "Rendez-vous regions: a scalable architecture for service location and data-centric storage in large-scale wireless networks", in *Proc. of Parallel and Distributed Processing Symposium*, 2004.

[54] S.K. Das and S.K.S. Jayaram, "A novel load balancing scheme for the teletraffic hot spot problem in cellular networks", *Wireless Networks*, Vol. 4, no. 4, pp. 325-340, 2004.

[55] D. Ghosal and B. Mukherjee, "Exploiting user profiles to support differentiated services in next-generation wireless networks ", *IEEE Networks*, Vol. 18, no. 5, pp. 40-48, 2004.

[56] E. Oh and B. Krishnamachari, "Energy Savings through Dynamic Base Station Switching in Cellular Wireless Access Networks", in *Proc. of IEEE Globecom*, Miami, FL, USA, 2010.

[57] B. Nunes and K. Obraczka, "Modeling Spatial Node Density in Waypoint Mobility", in *Proc. of International Conference on Mobile Ad-hoc and Sensor Systems, MASS*, Las Vegas, NV, USA, 2012.

[58] US census Bureau, http://www2.census.gov/

[59] Femto forum, "Femtocell - Natural Solution for Offload", *White paper,* June, 2010.

[60] Femto forum, "Challenges in Deployment of UMTS/HSPA Femtocell", *White paper,* 2008.

[61] M.C. Reed, "Femtocells: Opportunities and Challenges", *IEEE ComSoc tutorial, 2010.*

[62] D. Calin, H. Claussen and H. Uzunalioglu, "On femto deployment architectures and macrocell offloading benefits in joint macro-femto deployments", *IEEE Communications Magazine*, Vol. 48, no. 1, pp. 26-32, 2010.

[63] D. Lopez-Perez , A. Ladanyi, A. Juttner and J. Zhang, "OFDMA femtocells: Intracell Handover for Interference and Handover Mitigation in Two-Tier Networks", *in Proc. of Wireless Communications and Networking Conference WCNC*, Sydney, NSW, 2010.

[64] J. Roh, Y. Ji, Y.G. Lee, I. Rhee and T. Ahn, "Femtocell Traffic Offload Scheme for Core Networks", *4th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Paris, France, 2011.

[65] D. Lopez-Perez, A. Valcarce, G. de la Roche and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance" *IEEE Comm. Magazine*, Vol.47, no. 9, pp. 41-48, 2009.

[66] V. Chandrasekhar, J.G. Andrews, T. Muharemovic, Z. Shen and A. Gatherer, "Power control in two-tier femtocell networks", *IEEE Trans. on Wireless Communications*, Vol. 8, no. 8, pp. 4316-4328, 2009.

[67] V. Chandrasekhar, J. Andrews, "Spectrum allocation in tiered cellular networks", *IEEE Trans. Communications*, Vol. 57, no. 10, pp. 3059-3068, 2009.

[68] K. Sundaresan, S. Rangarajan, "Efficienty resource management in OFDMA femtocells ", in *Proc. of ACM international symposium on Mobile ad hoc networking and computing (Mobihoc)*, 2009.

[69] H.C. Lee, D.C Oh, Y.H. Lee, "Mitigation of Inter-Femtocell Interference with Adaptive Fractional Frequency Reuse", in *Proc. of IEEE International Conference on Communications ICC*, Cape Town, 2010.

[70] A. Hatoum, N. Aitsaadi, R. Langar, R. Boutaba, and G. Pujolle, "FCRA: Femtocell Cluster-based Resource Allocation Scheme for OFDMA Networks", in *Proc. of IEEE International Conference on Communications ICC*, Kyoto, 2011.

[71] G. Owen, Game Theory, *3rd ed. London, U.K.:Academic*, 1995.

[72] T. Basar and G. J. Olsder, "Dynamic Noncooperative Game Theory", *SIAM Classics in Applied Mathematics*, end edition, 1999.

[73] G. He, M. Debbah and E. Altman, "A Bayesian game-theoretic approach for distributed resource allocation in fading multiple access channels", *EURASIP Journal on Wireless Communications and Networking*, Vol. 2010, no. 8, 2010.

[74] R.B. Myerson, *Game Theory, Analysis of Conflict*, Cambridge, Harvard Univ.Press, 1991.

[75] F. Pantisano, M. Bennis, W. Saad, R. Verdone, M. Latva-aho, "Coalition formation games for femtocell interference management: a recursive core approach", in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, Cancun, Quintana Roo, 2011.

[76] R. Thrall and W. Lucas, "N-person games in partition function form", *Naval Res. Logistics Quart*, Vol. 10, no. 1, pp. 281-298, 1963.

[77] W. Saad, Z. Han, M. Debbah, A. Hjrungnes and T. Basar, "Coalitional Game Theory in Wireless and Communications Networks", *IEEE Signal Processing Magazine*, Vol. 26, no. 5, pp. 77-97, Sept. 2009.

[78] F. Pantisano, K. Ghaboosi, M. Bennis and M. Latva-Aho, "Interference Avoidance via Resource Scheduling in TDD Underlay Femtocells", in *Proc. of IEEE nternational Symposium on Personal, Indoor and Mobile Radio Communications Workshops (PIMRC Workshops)*, Instanbul, 2010.

[79] R. J Aumann and M. Maschler, "Game Theoretic Analysis of a Bankruptcy Problem from the Talmud", *Journal of Economic Theory*, 1985.

[80] L. Shapley, "A value for n-person games", *H Kuhn and A Tucker, eds, Contributions to the Theory of Games*, Vol. 2 of Annals of Mathematics Studies, Princeton U Press., 1953.

[81] M. Mycek, S. Secci, M. Pioro and J. Rougier, A. Tomaszewski, and A. Pattavina, "Cooperative Multi-Provider Routing Optimization and Income Distribution", in *Proc. of International Workshop on Design of Reliable Communication Networks-DRCN*, Washington, DC, USA, 2009.

[82] T. Alpcan and T. Basar, "A game theoretic approach to decision and analysis in network intrusion detection", in *Proc. of Conference on Decision and Control*, 2003.

[83] D. Schmeidler, "The Nucleolus of a Characteristic Function Game", *SIAM Journal on Applied Mathematics*, Vol. 17, No. 6, 1969.

[84] Z. Han and V. Poor, "Coalition Games with Cooperative Transmission: A Cure for the Curse of Boundary Nodes in Selfish Packet-Forwarding Wireless Networks", *IEEE Trans. on Communications*, Vol. 57, no. 1, pp. 203-213, 2009.

[85] D. WINNER II, "Winner II Channel models", *IST-4-027756 WINNER II, D1.1.2 V1.2, Tech. Rep.*, Sept. 2007.

[86] 3rd Generation Partnership Project, "3GPP TS 32.500 V10.0.0 (2010-06) specification", 2010.

[87] 3rd Generation Partnership Project, "3GPP TS 32.781 V9.1.0 (2010-03) specification", 2010.

[88] T. Driessen, *Cooperative Games, Solutions and Applications*, Kluwer Academic Publishers, 1988.

[89] R. Jain, W. Hawe and D. Chiu, "A Quantitative measure of fairness and discrimination for resource allocation in Shared Computer Systems", *DEC Research Report TR-301*, 1984.

[90] W. Lehr and L.W. Mcknight, "Wireless Internet access: 3G vs. WiFi?", *Telecommunications Policy*, Vol. 27, no. 5, pp. 351 - 370, 2003.

[91] K. Lee , J. Lee, Y. Yi, I. Rhee and S. Chong, "Mobile Data Offloading: How Much Can Wi-Fi Deliver?", *IEEE/ACM Transactions on Networking*, Vol. 21, no. 2, pp. 536 -550, 2013.

[92] "Wi-Fi Certified Passpoint Architecture for Public Access," *White Paper, Aruba Networks*, 2012.

[93] Wi-Fi Alliance, "Launch of Wi-Fi CERTIFIED Passpoint Enables a New Era in Service Provider Wi-Fi Austin, TX, June, 2012.

[94] A. Schumacher and J. Schlien, "WLAN Traffic Offload in LTE White Paper, November 2012.

[95] A. Balasubramanian, R. Mahajan and A. Venkataramani, "Augmenting mobile 3G using WIFI", in*Proc. of the international conference on Mobile systems, applications, and services ACM MOBISYS,* 2010.

[96] A. Aijaz, O. Holland, P. Pangalos and H. Aghvami, "Energy savings for cellular access network through Wi-Fi offloading", in *Proc. of IEEE International Conference on Communications - ICC*, Ottawa, Canada, 2012.

[97] X. Hou, P. Deshpande and S.R. Das, "Moving bits from 3G to metro-scale WiFi for vehicular network access: An integrated transport layer solution", in*Proc. of IEEE International Conference on Network Protocols (ICNP)*, Vancouver, Canada, 2011.

[98] X. Zhuo, W. Gao, G. Cao and Y. Dai, "Win-coupon: An incentive framework for 3G traffic offloading", in *Proc. of IEEE International Conference on Network Protocols (ICNP)*, 2011.

[99] D. Zhang and C.K. Yeo, "Optimal handing-back point in mobile data offloading," in *Proc. of IEEE Vehicular Networking Conference (VNC)*, Seoul, 2012.

[100] COST Action 231, "Digital mobile radio towards future generation systems, annual report," *tech. rep., European Communities, EUR 18957*, 1999.

[101] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, vol. 29, pp. 317-325, 1981.

[102] Y. Cui, W. Li, X. Cheng, "Partially overlapping channel assignment based on node orthogonality for 802.11 wireless networks," in *Proc. of IEEE INFOCOM*, Shanghai, 2011.

[103] M. Burton, "Channel overlap calculations for 802.11b networks," *White Paper, Cirond Technologies Inc.*, 2002.

[104] A. Rahmati and L. Zhong "Context-for-wireless: context-sensitive energy-efficient wireless data transfer," *in MobiSys*, 2007.

[105] N. Ristanovic, J.-Y. L. Boudec, A. Chaintreau, and V. Erramilli "Energy Efficient Offloading of 3G Networks" *In Proc. of IEEE MASS, 2011.*, 2007.

[106] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet Inter-domain Traffic," *SIGCOMM Comput. Commun. Rev.*, Vol. 40, no. 4, pp. 75 - 86, 2010.

[107] G. Xylomenos, C.N. Ververidis, V.A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K.V. Katsaros, and G.C. Polyzos, "A Survey of Information-Centric Networking Research," *IEEE Communications Surveys and Tutorials*, Vol. 16, no. 99, pp. 1024 - 1049, 2014.

[108] E. Nygren, R. K. Sitaraman, and J. Sun, "The akamai network: A platform for high-performance internet applications," *SIGOPS Oper. Syst. Rev.*, Vol. 44, no. 3, pp. 2 - 19, 2010.

[109] V. Jacobson, D.K. Smetters, J.D. Thornton, M. Plass, N. Briggs and R. Braynard, "Networking Named Content", in *Proc. of the international conference on Emerging networking experiments and technologies CoNEXT*, New York, NY, USA, 2009

[110] A. Detti, M. Pomposini, N. Blefari-Melazzi, S. Salsano and A. Bragagnini, "Offloading cellular networks with Information-Centric Networking: The case of video streaming", In *Proc. of the Int. Symp. on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, San Francisco, CA, USA, 2012.

[111] G. Zhang, Y. Li and T. Lin, "Caching in Information Centric Networking: A survey," *Computer Networks*, Vol. 57, no. 16, pp. 3128 - 3141, 2013.

[112] J. Bulow and J. Roberts. "The Simple Economics of Optimal Auctions," *Journal of Political Economy*, Vol. 97, no. 5, pp. 1060 - 1090, 1989.

[113] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. "Algorithmic Game Theory," *Cambridge University Press*, New York, NY, USA, 2007.

[114] D. Rossi and G. Rossini, "On sizing ccn content stores by exploiting topological information" in *Proc. of IEEE Infocom Nomen Workshop*, Orlando, Florida, USA, 2012.

[115] Y. Wang, G. Tyson, S. Uhlig and G. Xie, "Optimal Cache Allocation for Content-Centric Networking" in *Proc. of IEEE ICNP*, Gottingen, Germany, 2013.

[116] S. Fayazbakhsh, Y. Lin, A. Tootoonchian, A. Ghodsi, T. Koponen, B. Maggs, K. Ng, V. Sekar, and S. Shenker, "Less pain, most of the gain: Incrementally deployable ICN" in *Proc. of ACM SIGCOMM*, Hong Kong, China, 2013.

[117] V. Pacifici and G. Dan ,"Selfish content replication on graphs" in*Proc of IEEE International TeletrafiňĄc Congress ITC*, San Francisco, CA, USA, 2011.

[118] B. Chun, K. Chaudhuri, H. Wee, M. Barreno, C. Papadimitriou, and J.D. Kubiatowicz,"Selfish Caching in Distributed Systems: A Game-Theoretic Analysis" in*Proc of ACM symposium on Principles Of Distributed Computing (PODC)*, Newfoundland, Canada, 2004.

[119] E. Jaho, M. Karaliopoulos, and I. Stavrakakis, "Social similarity favors cooperation: the distributed content replication case" *IEEE Transactions on Parallel and Distributed Systems*, Vol. 24, no. 3, pp. 601 - 613, 2013.

[120] D. Saucez, S. Secci, and C. Barakat, "On the Incentives and Incremental Deployments of ICN technologies for OTT Services" *IEEE Network Magazine*, Vol. 28, no. 3, pp. 20 - 25, 2014.

[121] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions" *IEEE Transactions on Wireless Communications*, Vol. 3, no. 4, pp. 1250 - 1259, 2004.

[122] S. Keshav, "An Engineering Approach to Computer Networking: ATM Networks, the Internet, and the Telephone Network" *Addison-Wesley Longman Publishing Co., Inc.,*, Boston, MA, USA, 1997.

[123] R. B. Myerson, "Optimal auction design. Mathematics of Operations Research" *Mathematics of Operations Research*, Vol. 6, no. 1, pp. 58 - 73, 1981.

[124] G. K. Zipf, "Human Behavior and the Principle of Least Effort" *Addison-Wesley Press*, Cambridge, MA, 1949.

[125] H. Che, Y. Tung, and Z. Wang, "Hierarchical web caching systems: modeling, design and experimental results." *IEEE journal on Selected Areas in Communications*, Vol. 20, no. 8, pp. 1305-1314, 2002.

[126] W. K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache âĂİless for moreâĂİ in information-centric networks" in *Proc. of the 11th International IFIP TC 6 Conference on Networking*, Berlin, Germany, 2012.

[127] A. B. Atkinson, "On the measurement of inequality," *Journal of Economic Theory*, Vol. 2, no. 3, pp. 244 -263, 1970.

[128] C. Chen, "C-RAN: the Road Towards Green Radio Access Network," *White paper,* 2011.

[129] T. Anderson, L. Peterson, S. Shenker, and J. Turner, "Overcoming the Internet impasse through virtualization," *IEEE Computer,* Vol. 38, no. 4, pp. 34-41, 2013.

[130] X. Liu, P. Juluri and D. Medhi, "An experimental study on dynamic network reconfiguration in a virtualized network environment using autonomic management," in *Proc. of IFIP/IEEE International Symposium on Integrated Network Management IM*, Ghent, Belgium, 2013.

[131] NGMN Alliance, "Suggestions on Potential Solutions to C-RAN," `http://www.ngmn.org,` 2013.

[132] M. Satyanarayanan, P. Bahl, R. Caceres and N. Davies, "The Case for VM-based Cloudlets in Mobile Computing", *IEEE Pervasive Computing*, Vol. 8, no. 4, pp. 14-23, 2009.

[133] Lei Jiao, Roy Friedman, Xiaoming Fu, Stefano Secci, Zbigniew Smoreda and Hannes Tschofenig,"Challenges and Opportunities for Cloud-based Computation Offloading for Mobile Devices", in *Proc. of Future Network and Mobile Summit*, Lisbon, Portugal, 2013.

[134] T. Basar and G.J. Olsder, "Dynamic Noncooperative Game Theory," *SIAM Series in Classics in Applied Mathematics*, Philadelphia, USA, 1999.

[135] J.v. Neumann and O. Morgenstern, "Theory of Games and Economic Behavior," *Princeton University Press*, Princeton, NJ, USA, Sep. 1944.

[136] R. J. Aumann and B. Peleg, "Von neumann-morgenstern solutions to cooperative games without side payments," *Bulletin of American Mathematical Society*, Vol. 6, pp. 173-179, 1960.

[137] M. Voorneveld and S. Grahn, "A Minimal Test for Convex Games and the Shapley Value," *Uppsala - Working Paper Series*, 2001

[138] J.G. Diaz, E.S. Rodriguez, "The Core-Center and the Shapley Value: A Comparative Study," *Reports in Statistics and Operations Research 03-10, University of Santiago de Compostela,* 2003.

[139] B. Peleg and P. Sudholter, "Introduction to the Theory of cooperative games," *Series: Theory and Decision Library C*, Vol. 34, 2003.

[140] P. Walker, "A Chronology of Game Theory," `http://www.econ.canterbury.ac.nz/personal_pages/paul_walker/gt/hist.htm`, 2005.

Part III

Appendix

Appendix A

Cooperative Game Theory

This appendix is used mainly for the resource allocation modeling in Chapter 4 and for cache allocation modeling in Chapter 6.

A.1 Introduction

Game theory is a decision-making approach considering the strategic interaction between participants as a key aspect to take into account in problem solving. It can be divided into two branches: non cooperative [134] and cooperative games [74] [71], differing in the presence or not of binding agreements among the players. Non-cooperative game theory studies the strategic choices resulting from the interaction among competing players while cooperative theory describes the outcomes that result when the players come together in different combinations. In this appendix, we only focus on cooperative game theory, the algorithms in Chapter 4 and 6 are based on.

A.2 Cooperative Game: Definition

A cooperative or coalitional game is uniquely defined by the pair ($\mathcal{N}$,v) where:

- $\mathcal{N}$=1, 2, . . . , n, a finite set of players who seek to form cooperative groups or coalitions,
- v is the coalition value that quantifies the worth of a coalition or a group S of players in a game when they cooperate and interact together.

The most common form of a coalitional game is the characteristic form with the transferable utility (TU) property [135], whereby the value of a coalition S depends solely on the members of that coalition, with no dependence on how the players in $\mathcal{N}\backslash S$ are structured [77].

The value of a game in characteristic form with TU can thus be seen as a function that associates with every coalition S of $\mathcal{N}$ a real number interpreted as the gain created when the members of S ally. The TU property implies that the total utility represented by this amount can be divided in any manner between the

coalition members [77]. The characteristic function with TU can thus be seen as: $v : 2^N \rightarrow R$.

We note that there are coalitional games with non-transferable utility (NTU) property [74] [136] where the payoff or the profit that each player in a coalition S receives is dependent from the joint actions that the players of coalition S select. This sort of games is out of the scope of this appendix. The games studied in **Chapter 4** and **Chapter 6** are TU coalitional games.

A.3 Some Definitions

Definition A.3.1. For a given cooperative game $(\mathcal{N}, v)$, a **payoff allocation** is the vector $x = (x_1, x_2, ...x_i) \in R^{|S|}$ where x_i is the amount of utility that a player $i \in S$ receives from the division of $v(S)$, x_i is called the player's payoff. $|S|$ represents the cardinality of the coalition S.

Definition A.3.2. An allocation $(x_1, x_2, ..., x_n)$ is **individually rational** if no player receives less than what he could get on his own:

$$x_i \geq v(\{i\}), \forall\ i \in \mathcal{N} \tag{A.1}$$

In other terms, the division of the overall value must give each player as much value as that player receives without interacting with the other players.

Definition A.3.3. An allocation is **collectively rational** if the sum of the players' payoffs is at least equal to the payoff of the grand coalition formed by all the players in the game:

$$\sum_{i \in \mathcal{N}} x_i \geq v(\mathcal{N}) \tag{A.2}$$

Definition A.3.4. An allocation is **feasible** if the sum of the players' payoffs is at most equal to the grand coalition utility $v(\mathcal{N})$:

$$\sum_{i \in \mathcal{N}} x_i \leq v(\mathcal{N}) \tag{A.3}$$

Definition A.3.5. An allocation is **efficient** if the payoff vector exactly splits the total value:

$$\sum_{i \in \mathcal{N}} x_i = v(\mathcal{N}) \tag{A.4}$$

We note that if an allocation is feasible and collectively rational at the same time, automatically it is efficient.

Definition A.3.6. An allocation is **symmetric** if it allocates equal payoffs $x_i = x_j$ to symmetric players $i, j \in \mathcal{N}$ where:

$$v(S \cup \{i\}) = v(S \cup \{j\}), \forall\ S \subseteq N \setminus \{i, j\} \tag{A.5}$$

In other terms, by exchanging one player i by the other j in any coalition containing only one of the players, we do not change the payoff of that coalition.

Definition A.3.7. An allocation is **additive** if the payoff of each player in a sum of two games is the sum of the allocations to the player in each individual game. Mathematically, if v and ω are games, the game $(v + \omega)$ simply assigns to any coalition the sum of the payoffs the coalition would get in the two individual games.

Definition A.3.8. An allocation is called a **zero allocation** if it attributes to a null or a dummy player a payoff equal to zero. A null player i that satisfies:

$$v(S \cup \{i\}) = v(S), \forall\ S \subseteq \mathcal{N} \setminus \{i\} \tag{A.6}$$

gets a payoff $x_i = 0$.

We note that an efficient payoff vector is called a pre-imputation, and an individually rational pre-imputation is called an imputation.

Definition A.3.9. An allocation is **consistent** if $\forall i \neq j$ the division of $x_i + x_j$, prescribed for claims d_i and d_j is $(x_i; x_j)$.

This means that no player or group of players can gain more by unilaterally deviating from a consistent solution since it will always obtain the same profit.

Definition A.3.10. The **marginal contribution** of player i in a coalition S, denoted by $MC_{i,S}$ is the amount by which the overall profit value created for the coalition would shrink if the player i leaves it:

$$MC_{i,S} = v(S) - v(S \setminus \{i\}) \tag{A.7}$$

Definition A.3.11. A coalitional game is said to be **superadditive** [71] if it guarantees the formation of the grand coalition (formed by all the players in the game). The motivation under this formation is the assumption that by forming large coalitions, the profit obtained by each player is enhanced. The mathematical formula reflecting this definition is given by:

$$v(S_1 \cup S_2) \geq v(S_1) + v(S_2),\ \forall S_1 \subset \mathcal{N}, S_2 \subset \mathcal{N}, s.\ t.\ S_1 \cap S_2 = \emptyset. \tag{A.8}$$

We note that the characteristic function of the coalitional game should always satisfy the superadditivity property if the desired goal is the formation of a grand coalition grouping all players. Otherwise, if Equation (A.8) is not satisfied, the coalitional game should be modeled as a game in partition form [77].

Definition A.3.12. A coalitional game is said to be **supermodular** [80] [88], if the marginal utility of increasing a player's strategy rises with the increase in other players' strategies:

$$v(S \cup T) + v(S \cap T) \geq v(S) + v(T), \forall\ S, T \subseteq \mathcal{N} \tag{A.9}$$

The authors in [88] show that the supermodularity of v is equivalent to:

$$v(S \cup \{i\}) - v(S) \leq v(T \cup \{i\}) - v(T), \forall\ S \subseteq T \subseteq \mathcal{N} \setminus \{i\}, \forall\ i \in \mathcal{N}; \tag{A.10}$$

We note that a game is said to be **convex** if it adopts this supermodularity property. Moreover, the supermodularity property is stronger than the superadditivity one, so a game satisfying the supermodularity implies that it satisfies the superadditivity.

A.4 Solutions for Coalitional Games

In this section, we present the main solution concepts of coalitional games, proposed so far in the literature.

A.4.1 Core

The core [74] [71] is considered as a key concept for coalitional games, it is related directly to the grand coalition's stability. In a coalitional game $(\mathcal{N}, v)$, due to superadditvity property, the players have an incentive to form the grand coalition $\mathcal{N}$. Thus, the core is the set of payoff allocations which guarantees that no group of players has an incentive to leave $\mathcal{N}$ in order to form another coalition $S \in \mathcal{N}$.

$$C(v) = \left\{ x \in \mathbb{R}^N : \sum_{i \in N} x_i = v(\mathcal{N}); \quad \sum_{i \in S} x_i \geq v(S), \forall\ S \subseteq \mathcal{N} \right\}. \tag{A.11}$$

The core of coalitional game is not always guaranteed to exist since in many games the core is empty and hence the grand coalition cannot be stabilized. In convex games that satisfy the supermodularity property, the core is always non empty [71].

For example, let us consider the following game in the characteristic form, where the players are: $\mathcal{N} = A, B, C$ and the payoff values are given in Table A.1

The core of this game $C(v)$ can thus be computed using Equation (A.11):

$C(v) = \left\{ x \in \mathbb{R}^N \right\}$ s.t. $x_1 \geq 0$; $x_2 \geq 0$; $x_3 \geq 0$; $x_1 + x_2 \geq 40$; $x_1 + x_3 \geq 30$; $x_2 + x_3 \geq 10$; $x_1 + x_2 + x3 = 100$. Where x_1, x_2 and x_3 are the profits of players A, B, C respectively. The group of imputations in the core is illustrated in Figure A.1.

A.4.2 Shapley Value

As stated before, the core of a coalitional game does not always exist and sometimes it can be quite large hence relying on the core concept is not a good option. This motivated the researchers to look for another solution concept that can associate

Table A.1: Coalitional payoffs.

Coalition	$v(S)$
$\emptyset$	0
A	0
B	0
C	0
$A \cup B$	40
$A \cup C$	30
$B \cup C$	10
$A \cup B \cup C$	100

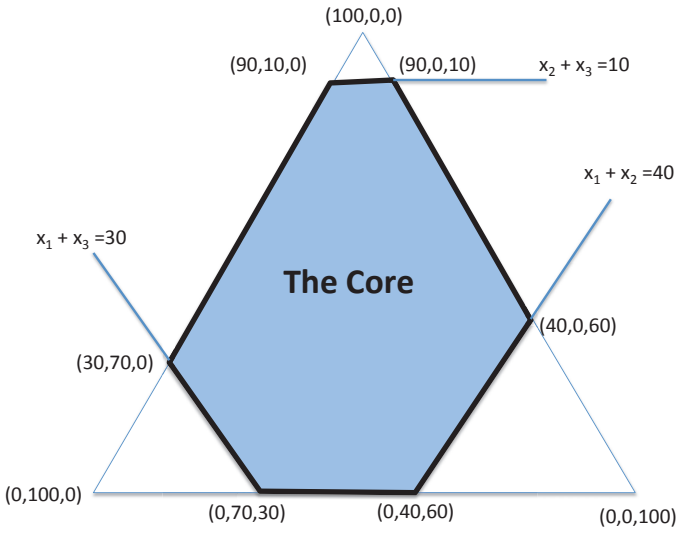


Figure A.1: The set of imputations in the core.

to every coalitional game $(\mathcal{N}, v)$ a unique payoff vector known as the value of the game. Shapley [74] [80] approached this problem axiomatically by defining a set of desirable properties and by characterizing a unique mapping that satisfies these axioms, later known as the Shapley value.

The Shapley value satisfies the efficiency (i.e, Def. A.3.5), the symmetry (i.e, Def. A.3.6, the dummy (i.e, Def. A.3.8) and the additive (i.e, Def. A.3.7) axioms. It is calculated as follows:

1. We first consider all the possible permutations of the players in the grand coalition (e.g., if we have three players as before A, B, C the permutations are

ABC, ACB, BAC, BCA, CAB, CBA). In general, in a game with $\mathcal{N}$, we have $N!$ permutations.

2. For each permutation and each player, we calculate the marginal contribution that the player grants if he joins the coalition formed by the predecessor players (i.e, Def. A.3.10).

3. For each player, we calculate the average of its marginal contributions on all the permutations. The obtained results give the Shapley value solution having a time complexity of O(N!).

Let us take again the same example of Section A.1, the Shapley value solution is reported in Table A.2.

Table A.2: Shapley Value computation.

Permutations	**Marginal Contribution of A**	**Marginal Contribution of B**	**Marginal Contribution of C**
A, B, C	$v(\{A\}) - v(\varnothing) = 0$	$v(\{A,B\}) - v(\{A\}) = 40$	$v(\{A,B,C\}) - v(\{A,B\}) = 60$
A, C, B	$v(\{A\}) - v(\varnothing) = 0$	$v(\{A,B,C\}) - v(\{A,C\}) = 70$	$v(\{A,C\}) - v(\{A\}) = 30$
B, A, C	$v(\{A,B\}) - v(\{B\}) = 40$	$v(\{B\}) - v(\varnothing) = 0$	$v(\{A,B,C\}) - v(\{A,B\}) = 60$
B, C, A	$v(\{A,B,C\}) - v(\{B,C\}) = 90$	$v(\{B\}) - v(\varnothing) = 0$	$v(\{B,C\}) - v(\{B\}) = 10$
C, A, B	$v(\{A,C\}) - v(\{C\}) = 30$	$v(\{A,B,C\}) - v(\{A,C\}) = 70$	$v(\{C\}) - v(\varnothing) = 0$
C, B, A	$v(\{A,B,C\}) - v(\{B,C\}) = 90$	$v(\{B,C\}) - v(\{C\}) = 10$	$v(\{C\}) - v(\varnothing) = 0$
Average	41.67	32.67	26.66

The formula for calculating the Shapley value is given by:

$$\Phi_i(v) = \sum_{S \subset \mathcal{N} \setminus \{i\}} \frac{|S|!(N - |S| - 1)!}{N!} [v(S \cup \{i\} - v(S)] \tag{A.12}$$

In general the Shapley value is unrelated to the core; however for convex games that satisfy the supermodularity property, the Shapley value lies in the core and it is considered as the barycenter of this core [137] [138].

A.4.3 Kernel

The "Kernel" solution concept groups the imputations such that no player has bargaining power over another one. It states that if players i and j are in the same coalition, then the highest excess that player i can make in a coalition without player j is equal to the highest excess that player j can make in a coalition without player i.

By definition we say that a player i has more bargaining power than player j with respect to an imputation $\overline{x} = (x_1, x_2, ..., x_n)$ if the maximum surplus he has on j is bigger than the maximum surplus that a player j has on him.

The maximum surplus $s_{i,j}(\overline{x})$ is defined as the maximal amount that a player i can gain without the cooperation of player j , it is given by:

$$s_{ij}(\overline{x}) = \max \left\{ \nu(S) - \sum_{k \in S} x_k : S \subseteq N \setminus \{j\}, S \ni i \right\}. \tag{A.13}$$

Intuitively, player i has more bargaining power than player j with respect to imputation x if the maximum surplus he has on j is bigger than the maximum surplus j has on him [i.e., if $s^{\nu}_{ij}(\overline{x}) > s^{\nu}_{ji}(\overline{x})$].

As a conclusion, the kernel is computed by:

$$[s^{\nu}_{ij}(\overline{x}) - s^{\nu}_{ji}(\overline{x})][x_j - \nu(j)] \leq 0 \ \wedge \ [s^{\nu}_{ji}(\overline{x}) - s^{\nu}_{ij}(\overline{x})][x_i - \nu(i)] \leq 0, \ \forall i, j \in N \quad \text{(A.14)}$$

In the case of a convex game that satisfies the supermodularity property, the kernel consists of a single point that is denoted thereafter as Nucleolus [139].

A.4.4 Nucleolus

Finally, another concept is worth being mentioned: the "Nucleolus". It is the imputation that minimizes the worst inequity. The Nucleolus is computed on the imputation set as follows:

1. For each coalition $S \in \mathcal{N}$, we determine the inequity of an imputation $\overline{x}$ as the coalitional surplus also called the excess that measures the amount by which coalition S falls short of its potential $v(S)$ in the allocation $\overline{x}$. The excess is computed as the difference between the coalition worth $v(S)$ and the received payoff $\sum_{j \in S} x_j$;

$$e(x, S) = v(S) - \sum_{j \in S} x_j \ \ , \forall S \subset \mathcal{N}. \quad \text{(A.15)}$$

2. We order the coalitions with respect to their excesses, decreasingly.

3. The Nucleolus is the lexicographic minimum of all possible excess vectors, it is the solution that minimizes the largest excess.

We note that a game has a unique nucleolus that is always in the kernel. If the core is non-empty, the nucleolus belongs to the core and can be used for selecting a core element.

If we return back to the example above with three players $\mathcal{N} = A, B, C$, with the coalition payoffs of Table A.1, the Nucleolus solution is obtained in Table A.3 and computed as follows:

1. We start at an arbitrary allocation vector such that $x_1 + x_2 + x_3 = 100$, e.g., $(50, 30, 20)$.

2. After computing the excesses of the different coalitions, we determine the largest excess, corresponding to coalition C in our case. Our goal is to minimize the largest excess, but we notice that by making x_3 larger, the excess of $A \cup B$ increases at the same rate and these excesses then meet at -30, when $x_3 = 30$. Clearly, no other allocation x can make the excess smaller than -30 since at least one of the coalitions C or $A \cup B$ can have at least an excess of -30. Hence, $x_3 = 30$ is the first component of the Nucleolus.

3. Proceeding in the same manner, one finally obtains the Nucleolus allocation $(35, 35, 30)$ for the considered game.

Table A.3: Nucleolus computation.

Coalition	$e(x, S)$	$(50, 30, 20)$	$(38, 32, 30)$	$(35, 35, 30)$
F_1	$-x_1$	-50	-38	-35
F_2	$-x_2$	-30	-32	-35
F_3	$-x_3$	-20	-30	-30
$F_1 \cup F_2$	40-x_1-x_2	-40	-30	-30
$F_1 \cup F_3$	30-x_1-x_3	-40	-38	-35
$F_2 \cup F_3$	10-x_2-x_3	-40	-52	-55

Visibly, the computation of the nucleolus is not trivial. Another way of computing it is to model the problem as mixed linear programming problem. The objective is to minimize the maximal excess α such that:

$$\nu(S) - \sum_{i \in S} x_i \leq \alpha, \ \forall S \subset \mathcal{N} \tag{A.16}$$

And the grand coalition has always null surplus:

$$\sum_{i \in N} x_i = \nu(N) \tag{A.17}$$

Solving this LP, many optimum solutions might be obtained for the same minimum α_0. If this is not the case, the solution vector $\overline{x}^*$ is the nucleolus, and α_0 is the maximal excess. Otherwise, one extracts the set of binding coalitions S_0 and iterates the optimization changing (A.16) with:

$$\nu(S) + \sum_{i \in S} x_i = \alpha_0, \ \forall S \in S_0 \tag{A.18}$$

when a coalition belongs to S_0. S_0 is to be built as set of coalitions that are likely to oscillate the least, typically one takes the least surplus set of coalitions. After at most n iterations the solution is unique and it is the nucleolus of the game.

A.5 Bankruptcy Game

The story of begins astonishingly enough with the Babylonian Talmud! The Talmud is a collection of commentaries on the (Hebrew) Bible that originated as an oral tradition and eventually was codified into written form. The Talmud discusses the bankruptcy problem in the context of a man offering debts to his three wives in excess of his assets. The Talmud answer is not immediately obvious, and in fact, the answer baffled academics for over almost 2,000 years [140].

A.5.1 Talmud Mystery

One problem discussed in the Talmud is the so called marriage contract problem: a man has three wives whose marriage contracts specify that in the case of this death

Figure A.2: Talmud answers.

they receive 100, 200 and 300 respectively. The Talmud offers answers through three examples. The text does not contain a general rule, which is what makes these answers seemingly contradictory. The three cases are when the estate size is 100, 200, and 300.

- When the man dies leaving an estate of only 100, the Talmud awards 33 1/3 to each wife, it seems an equal division of the estate but strangely this is not the same division used in the other cases.
- When the estate is 300, the Talmud recommends a proportional division of 50, 100 and 150 to the wives.
- For an estate of 200, the estate is supposed to be divided as 50, 75, and 75. The division does not seem an equal division nor a proportional division, but it is simply a curious decision altogether. Why should the second and third wives be given the same amount of money? And where do the numbers come from?

We summarize the Talmus answers in Figure A.2; the rows are the estate sizes, the columns are the claims or demands, and the table entries are the division size recommended by the Talmud. The answers defied a proper explanation for almost 2,000 years, filling volumes of critical review. Some scholars have essentially given up and suggested the 200 case might be an issue of faulty transcription.

A.5.2 Talmud Solution and Bankruptcy Game

In the 1980s the mystery has been cracked [79], the authors demonstrate that the Talmud solution is computed by the following seven steps algorithm:

1. Order the claimants from lowest to highest claims.
2. Divide the estate equally among all parties until the lowest claimant receives one half of the claim.
3. Divide the estate equally among all parties except the lowest claimant until the next lowest claimant receives one half of the claim.
4. Proceed until each claimant has reached one-half of the original claim.

5. Now, work in reverse. Start giving the highest-claim money from the estate until the loss, the difference between the claim and the award, equals the loss for the next highest claimant.

6. Then divide the estate equally among the highest claimants until the loss of the highest claimant is equal the loss of the next highest.

7. Continue until all the estate has been awarded.

They also demonstrate that the Talmud answer can be viewed as a consistent application of a game theory principle and that the Talmud answer is the solution of a properly defined coalitional game, called thereafter: *Bankruptcy Game* and defined as follows:

Definition A.5.1. A bankruptcy game [79] is defined as $G(\mathcal{N}, v)$ where $\mathcal{N}$ represents the claimants of the bankruptcy situation and v is the characteristic function that associates to each coalition its worth defined as the part of the estate not claimed by its complement:

$$v(S) = max(0, E - \sum_{i \in \mathcal{N} \setminus S} d_i) \, , \forall S \subseteq \mathcal{N} \setminus \{\emptyset\} \tag{A.19}$$

where $E \geq 0$ is an estate that has to be divided among the members of $\mathcal{N}$ (the claimants) and $d \in R_+^{|\mathcal{N}|}$ is the claim vector such that $E < \sum_{i \in N} d_i$.

Equation (A.19) has been proven to be superadditive [71]. Moreover, it satisfies the supermodularity property [80] [88].

The Nucleolus defined in Section A.4.4 has been shown [79] as the unique consistent solution (i.e., please refer to Definition A.3.9) in bankruptcy game.

UNIVERSITÉ PIERRE ET MARIE CURIE
LABORATOIRE D'INFORMATIQUE DE PARIS 6
4 PLACE JUSSIEU, 75005 PARIS