Allocation opportuniste de spectre pour les radios cognitives

Jean-Christophe Dunat

Thèse

présentée pour obtenir le grade de docteur

de l'Ecole Nationale Supérieure
des Télécommunications

Spécialité : Informatique et Réseaux

Jean-Christophe DUNAT

Allocation opportuniste de spectre pour les radios cognitives

Thèse soutenue le 07 Avril 2006 à l'ENST Paris, devant le jury composé de

Président	Jean-Claude Belfiore	(ENST Paris, France)
Rapporteurs	Klaus Moessner	(Université de Surrey, UK)
	Sami Tabbane	(Ecole Sup. des Comm. de Tunis, Tunisie)
Examinateurs	Xavier Lagrange	(ENST Bretagne, France)
	Jean-Claude Belfiore	(ENST Paris, France)
	David Grandblaise	(Motorola Labs Paris, France)
Directeur de thèse	Christian Bonnet	(Institut Eurécom, France)

Remerciements

Je tiens à remercier vivement les membres de mon jury de thèse pour le temps qu'ils ont pu y consacrer et notamment pour avoir été présents à ma soutenance de thèse. Au travers de leurs remarques et de leurs questions j'ai pu mesurer tout l'intérêt qu'ils portent à ce travail. A ce sujet, je tiens à remercier tout spécialement Xavier Lagrange pour tout le temps qu'il a investi dans la relecture minutieuse du rapport et dont les remarques ont fait beaucoup pour améliorer la clarté et la justesse du présent rapport final. Mes sincères remerciements vont donc à Jean-Claude Belfiore pour avoir présidé le jury de thèse et pour la justesse de ses remarques, à Xavier Lagrange, à Klaus Moessner et Sami Tabbane pour leur intérêt dans ce domaine de l'allocation flexible de spectre et pour le temps qu'ils ont passé à relire mon rapport initial.

Je tiens également à adresser mes remerciements à Christian Bonnet, mon directeur de thèse, pour son encadrement pendant toutes ces années de thèse, pour son écoute, ses conseils et son expérience qui ont été pour moi une aide précieuse afin d'orienter ma problématique de thèse. En outre, je remercie vivement Merouane Debbah pour ses précieux conseils et tout le temps consacré à m'aider.

Si ce travail a pu se construire et aboutir de façon fructueuse, malgré les quelques contraintes matérielles et structurelles que j'ai pu vivre pendant cette thèse, c'est en partie grâce au soutien de certaines personnes. Je remercie Lucas Elicegui avec qui j'ai pu travailler avec succès pendant la première partie de cette thèse et qui a permis d'initier le processus pour que je fasse cette thèse avec Motorola. Je tiens également à remercier Dragan Boscovic et Prakash Moorut pour m'avoir permis de commencer chez Motorola.

Je remercie très vivement David Grandblaise avec qui j'ai travaillé pendant la plus grande partie de ma thèse à Motorola, pour son attitude toujours positive, sa motivation, la qualité de ses conseils, ainsi que tout le temps passé à relire le rapport.

Un grand merci à Didier Bourse pour ses nombreux encouragements et sa motivation communicative, à Jean-Noël Patillon pour son soutien technique et logistique, à Guillaume Vivier pour la justesse de ses conseils.

Merci à Laurent Mazet, Mohammed Kammoun et CIS pour vos nombreuses aides en informatique. Merci à Pascale Griveaux pour son sourire, sa bonne humeur constante et son aide pour la partie administrative. Merci à Rémy Pintenet et Guillaume Arnaudet pour les longues discussions que nous avons pu avoir et qui ont contribué à la réussite de mon travail.

Je tiens à remercier également tous ceux que j'ai croisé durant ces années de thèse, nos échanges ont contribué à alimenter ma réflexion sur ma problématique de thèse.

Enfin, je remercie du fond du coeur Laurence, ma femme, pour ses encouragements constants, ses conseils, sa confiance en moi et sa patience pour supporter les journées à rallonges pendant des périodes à rallonges. Je remercie enfin également sincèrement ma famille qui m'a toujours soutenue dans cet effort, ce qui a beaucoup compté pour moi.

Abstract

Within a few years, Réseau Local sans fil (WLAN) systems such as IEEE 802.11 have become increasingly popular systems, and have been deployed in many public places, offices and homes. Aside costly cellular systems, they have opened up the possibility for everyone to cheaply and simply install, and own a WLAN network, offering high data rates to support data and voice services. In addition there has been an increase in the demand for wireless services, and several market studies have forecasted a sustained growth in the years to come. In the future, wireless networks will intensively use radio packet transmissions, with a single hop or multi-hop access to an infrastructure connected to the Internet. However, several current limitations need to be removed to unleash the full potential of WLANs. The first issue is the "spectrum scarcity": in many countries the spectrum allocation chart is almost full, without the allocated bands being necessarily used. Accordingly, the spectrum allocation regulatory rules should be more flexible to allow for their opportunistic use. The second issue is in the Contrôle d'Accès au Médium (MAC) layer limitations of WLANs with distributed access: more intelligence is needed to use the available spectrum more efficiently. As a result, new solutions (regulatory, technical) need to be found to turn the command-and-control policy of spectrum use into a more dynamic and beneficial model. This is being possible in a cognitive radio context, where more intelligence and environment-awareness is included at the user terminal level.

Our objective in this thesis is to propose innovative technical algorithms of dynamic spectrum allocation. Assuming smarter devices, as in the cognitive radio context, we have adapted the legacy technical framework for spectrum allocation and used it into a more efficient way. More specifically, we study algorithms of opportunistic spectrum access at the user level to achieve simultaneously the resource allocation and the users' scheduling, in a case of multiple users and multiple channels.

Our first contribution identifies the MAC limitations of IEEE 802.11 WLAN systems. We are especially interested in the IEEE 802.11a version. Deployments of several cells and users are studied by simulation. We evaluate the impact of the users' distribution on the spatial unfairness and the inter-cell interference. The conclusion is that the IEEE 802.11 MAC protocol is very sensitive to variations in users' distributions. In other words, the considered MAC protocol strongly impacts each user's channel access success as a function of his position in the cell (relative to the other users, as well as relative to the Point d'Accès (AP)).

We also tackle the spectrum allocation problem by comparing several channel access methods to find the best one. This is done using the queuing theory analysis tool. We show that the channel access policies achieving the best system throughput are the ones changing their choice of spectrum allocation, as a function of the different users' link qualities per channel (opportunistic policy), contrary to policies consisting in a static or even a dynamic but "blind" (regarding users' channel conditions) allocation.

Based on the previous results, we propose a new distributed MAC with more intelligence at the border of the network (located within the users' devices). In other words, we design a distributed opportunistic algorithm to control the spectrum access in Voie de communication Montante (UL) at the user level. This bottom-up approach uses (1) a dynamic available spectrum resource (bandwidth and position), (2) a user-based distributed control for spectrum access and (3) an opportunistic spectrum access (spectrum access contention collectively controlled). This is particularly suitable in dynamic environments and when the number of users is important. The algorithm uses, in a wireless context, an adaptation of the swarm intelligence meta-heuristic, a method inspired from the study of social insects (e.g.: ants, bees, wasps, termites). It brings flexibility, robustness and dynamicity in the spectrum allocation. Our Accès Multiple par Séparation de Fréquences Orthogonales (OFDMA) algorithm (multi-user Multiplexage par Division de Fréquences Orthogonales (OFDM)) brings an innovative self-spectrum planning method adapting the spectrum allocation (OFDM Sous-porteuses (SC)s) according to the needs. This approach is particularly suited for cognitive radios and would provide them with dynamic spectrum allocation algorithms. Indeed, such advanced radios have reconfigurable capabilities and can adapt their behavior as a function of the environment. Our solution is compared to other methods in terms of quality of the final reached solution. The speed of convergence towards a good allocation is evaluated. Our results demonstrate the real potential of such distributed collaborative methods for tackling the spectrum allocation problem. The complexity of a real wireless network is recreated at the global level, by means of simple (controlled) local interactions. This bottom-up approach requires completely new modeling rules, which is a real challenge to overcome, but the benefit is huge. We believe that future wireless networks will be more and more designed according to bottom-up approaches.

"The imagination of nature is far, far greater than the imagination of man."
Richard Feynman

Contents

List of Acronyms

2G 2nd Generation

3G 3rd Generation

AACh Alternate Adjacent Channel

ACh Adjacent Channel

ACK ACKnowledgment

ACO Ant Colony Optimization

AES Advanced Encryption Standard

AP Access Point

BC Backoff Counter

BPSK Binary Phase Shift Keying

BS Base Station

BSA Basic Service Area

BSS Basic Service Set

BW BandWidth

BWA Broadband Wireless Access

CCA Clear Channel Assessment

CDMA Code Division Multiple Access

CFP Contention Free Period

CIR Carrier to Interference Ratio

COFDM Coded Orthogonal Frequency Division Multiplexing

CP Contention Period

CR Cognitive Radio

CS Carrier Sense

CSMA/CA Carrier Sensing Multiple Access / Collision Avoidance

CSMA/CD Carrier Sensing Multiple Access / Collision Detection

CTS Clear To Send

CW Contention Window

DAB Digital Audio Broadcast

DARPA Defense Advanced Research Projects Agency

DATA DATA frame

DCA Dynamic Channel Assignment

DCF Distributed Coordination Function

DFS Dynamic Frequency Selection

DFT Discrete Fourier Transform

DI Diffuse Infrared

DIFS Distributed Inter-Frame Space

DL DownLink

DS Distribution System

DSA Dynamic Spectrum Allocation

DSSS Direct Sequence Spread Spectrum

DVB Digital Video Broadcast

EA Evolutionary Algorithm

EIFS Extended Inter-Frame Space

ESS Extended Service Set

ETSI European Telecommunications Standards Institute

FCA Fixed Channel Assignment

FCC Federal Communication Commission

FDD Frequency Division Duplex

FDMA Frequency Division Multiple Access

FFT Fast Fourier Transform

FHSS Frequency Hopping Spread Spectrum

FIFO First In First Out

GPRS General Packet Radio Service

GPS Global Positioning System

GSM Global System for Mobile communications

HCA Hybrid Channel Assignment

HCF Hybrid Coordination Function

HiperLAN High performance radio Local Area Network

HOL Head-Of-Line

IBSS Independent Basic Service Set

ICI Inter Carrier Interference

ID IDentifier

IDFT Inverse Discrete Fourier Transform

IEEE Institute of Electrical and Electronics Engineers

IFFT Inverse Fast Fourier Transform

IFS Inter Frame Space

IP Internet Protocol

ISI Inter Symbol Interference

ISM Industrial, Scientific and Medical

ITU International Telecommunications Union

LA Link Adaptation

LAN Local Area Network

LOS Line Of Sight

LRC Long Retry Counter

MAC Medium Access Control

MCL Minimum Coupling Loss

MIMO Multiple Input Multiple Output

MMS Multimedia Message Service

MPDU MAC Protocol Data Unit

MS Mobile Station

MSDU MAC Service Data Unit

MT Mobile Terminal

MVA Mean Value Analysis

NAV Network Allocation Vector

NLOS Non Line Of Sight

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

OFDM/FDMA OFDM with FDMA

OFDM/TDMA OFDM with TDMA

OSI Open Systems Interconnection

PAPR Peak-to-Average Power Ratio

PCF Point Coordination Function

PER Packet Error Rate

PHY PHYsical

PLCP PHYsical Layer Convergence Protocol

PMD Physical Medium Dependent

PPDU PLCP Protocol Data Unit

PSDU PLCP SDU

PSO Particle Swarm Optimization

QAM Quadrature Amplitude Modulation

QoS Quality of Service

QPSK Quadrature Phase Shift Keying

RAT Radio Access Technology

RCH Random access CHannel

RF Radio Frequency

RKRL Radio Knowledge Representation Language

RL Reinforcement Learning

RRM Radio Resource Management

RSS Received Signal Strength

RTS Request To Send

Rx Receiver / Received

SC Sub-Carriers

SDMA Space Division Multiple Access

SDR Software Defined Radio

SDU Service Data Unit

SFN Single Frequency Network

SI Swarm Intelligence

SIFS Short Inter-Frame Space

SINR Signal to Interference Noise Ratio

SMS Short Message Service

SNR Signal to Noise Ratio

SRC Short Retry Counter

SS Subscriber Station

STA STAtion

STL Standard Template Library

TDD Time Division Duplex

TDMA Time Division Multiple Access

TPC Transmission Power Control

TS Time Slot

Tx Transmitter / Transmitted

UE User Equipment

UL UpLink

UMTS Universal Mobile Telecommunications System

UTRA UMTS Terrestrial Radio Access

UWB Ultra Wide Band

VoIP Voice over IP

WAN Wide Area Network

WCDMA Wideband Code Division Multiple Access

WEP Wired Equivalent Privacy

WLAN Wireless Local Area Network

WM Wireless Medium

WPAN Wireless Private Area Network

XG neXt Generation

List of Figures

List of Tables

Chapter 1

Introduction

1.1 About Wireless Communication

We live in a world of ever expanding "wireless communication". Let us define the terms "communication" and "wireless", as well as a brief history of their evolution.

Why does people communicate? It seems that humans are social animals who need to communicate with each other. Can we postulate that: the easier the **communication**, the greater the need for communication? Since several years, the technological progresses achieved in the field of computers have permitted an exchange of information easier, faster, cheaper and more reliable than ever before. Such novel capabilities are creating new ways and reasons of communicating, thus transforming the communication habits between people. What information does people need to exchange? People exchange diverse kind of contents such as: movies, music, pictures, formatted documents (typed, written), voice (conversation), etc. All these information can nowadays be digitalized, ensuring a more reliable storing and exchange. In this thesis, we will be concentrating on the technological aspects of the communication, as described hereafter.

How does people communicate? Today, the Internet makes us live in a world-size network in which everyone can be a part of it and can be connected to millions of others people. The exchange of information has moved from a centralized topology into a decentralized topology.

A communication between a source and one or several destinations can be either wired (using a type of wire such as: copper wire, optic fiber) or wireless. Strictly speaking, a **wireless** communication is a transmission of an information from a source to a remote destination, without wires and without the need for the message source to move towards the destination. Very ancient and diverse examples of wireless communication techniques can be found in the history of humankind, each one trying to answer the following question: "how to transmit an information between a source and a remote destination, simply, quickly and reliably?". Many communication protocols and information coding techniques have been invented and used. Here is a list of several wireless communication techniques:

- Visible signals: smog (used by Indian tribes), Chappe telegraph.
- Lighthouse (for boats), traffic lights, laser communication.
- Acoustic signals: acoustic Morse coding.
- Electromagnetic signals: cellular radiocommunication system.

Due to its interesting wireless communication capabilities, let us describe the Chappe telegraph. The first successful experience of the Chappe telegraph was realized by Claude Chappe (1763-1805) in 1791. Then, his invention was deployed all over France and even in some countries close to the frontiers. In 1790, it was taking four days by the road to deliver a message from Paris to Strasbourg (distant by 480 km). In 1799, using the Chappe telegraph, it was taking only two hours, leading to a speed of message propagation of 240 km/h (much faster than any other known means of communication at that time) !!!

The Chappe telegraph was composed of several towers acting as transmitters, receivers and relays to exchange information from the source to the final destination. All the towers used the same communication protocol and relayed the information to its closest neighboring tower. The message was coded and visually transmitted using a set of moving mechanical parts located at the top of the tower.

The system was already working as a network composed of geographically distant nodes, and included the following features: a communication protocol, information coding, message relaying, uplink and downlink communication, message collision resolution and error recovery. Accordingly, the Chappe telegraph was the first asynchronous and distributed European wireless telecommunication system, and can be considered as the ancestor of the Internet system.

Figure 1.1 shows a tower and the map of all the towers of the network used by the Chappe telegraph.

Wireless communication using electromagnetic waves is only one technique out of the other possible ones, but it is the most promising one for large scale networks. Note however, that even though the Chappe telegraph is an old wireless communication technique, it was even more "wireless" than any current cellular system (e.g.: GSM). Indeed, any information sent from a GSM mobile located in a cell to another mobile located in another cell would usually travel a longer distance using wires than using a wireless communication. Cellular systems are in fact "last mile" wireless communication systems between a user equipment and a radio access infrastructure equipment, whereas the communication between infrastructure equipments is usually wired.

For the remaining of this document, we will refer to "wireless communication" as a communication achieved without wire using the propagation of electromagnetic waves, and excluding visible light signals.

The early days of wireless communication using electromagnetic waves date back to the pioneering works of James Clerk Maxwell (1831-1879) and Heinrich Hertz (1857-1894). They gained a better understanding of the possibilities to control the underlying physical rules of electromagnetic wave

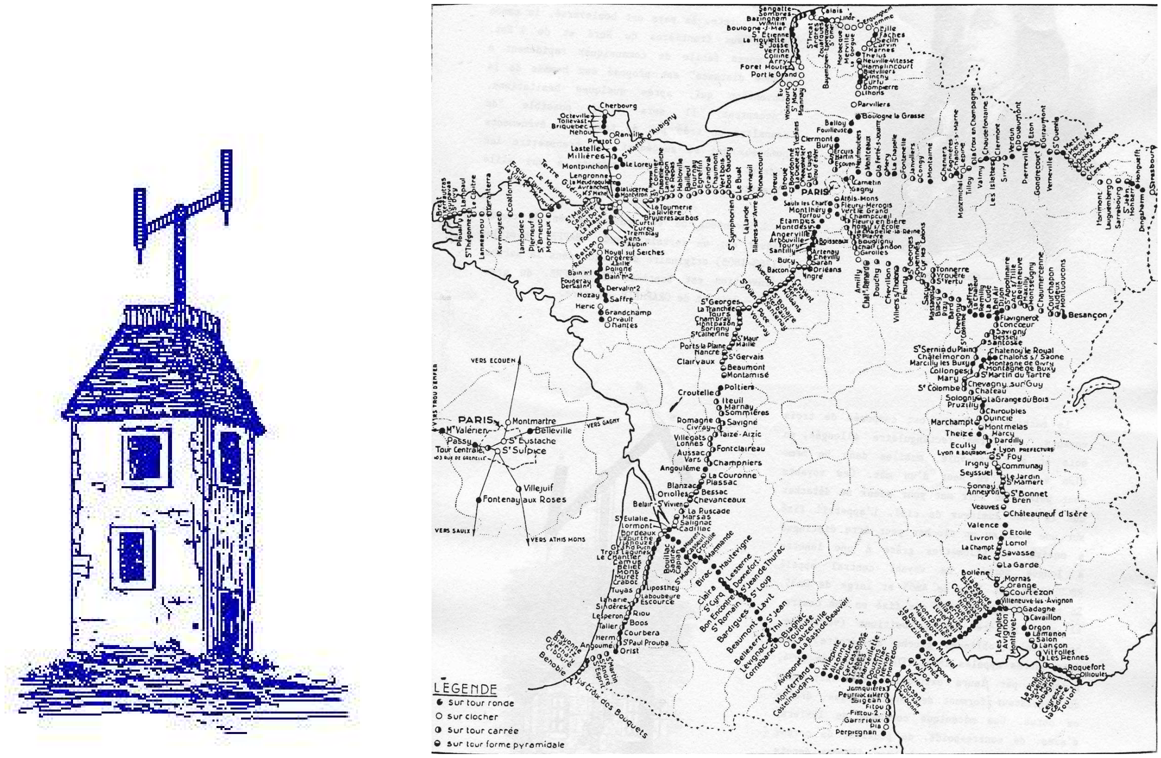


Figure 1.1: Example of a Chappe Telegraph Tower and Network Map (1793-1852)

propagation, in order to send and receive information using these radio waves. Before them, electromagnetic waves propagation was not used as a mean of wireless communication.

However, the "spectrum resource" was already present everywhere, it was abundant in any location on earth and at anytime, and its use was not limited by any license giving somebody a greater right-of-use than anybody else on a part of the spectrum. Since then, new wireless systems have emerged (e.g.: radio, television, mobile phone, wireless Internet connectivity, remote control, etc) relying on ever improving technologies. Here are some examples showing the growing wireless market size over time: in the 1960's - 1970's, for commercial use, the Bell laboratories developed the cellular concept. In 1999, the number of wireless subscribers was greater than the number of wireline subscribers in several European countries (Finland, Norway, Italy).

Wireless communication technologies have changed the way we communicate in our societies. However, the mobile cellular market is quickly evolving and has very diverse requirements depending on the users. Accordingly, predicting the future profitable services for the next generations of wireless systems is challenging. As an example, who would have predicted that many users might be willing to transmit billions of unformatted text messages? Now, the success of Short Message Service (SMS) is recognized and exploited by operators in their commercial offers, while the market forecasts for the SMSs (and Multimedia Message Service (MMS)s) show an explosion in their use in the years to come.

Our research objective is focused on identifying the approach and the algorithms that would reveal the hidden potential of spectrum use for the next generations of wireless systems.

1.2 Background and Motivation

1.2.1 Heterogeneous Wireless World

Since a few years there has been an increasing number of new radio communication standards created for commercial use. A large variety of new services are offered to the users with an ultimate goal of service availability and continuity of service anytime and anywhere. Most of the expected services require high data rates (such as music, image or videos downloads) with high reliability (good quality), high disponibility (in time and space), high continuity (always available even when moving and changing environment) and high integrity (to maintain ones privacy), low cost (cheap terminals and/or licenses), low communication costs, and upgradeable terminals.

A currently discussed question is: to achieve a continuity of service for a user, where should the complexity be located? in the user equipment or in the network or in both? Technological progresses do not require anymore the user to have one device per standard. Rather, more than a single Radio Access Technology (RAT) is being included into the same device. Also, the Software Defined Radio (SDR) allows a reconfigurability of the equipments, with sometimes over-the-air download capabilities to upgrade an equipment or correct some bugs. However, to this date, not all the created standards have turned into a commercial mass market success. Actually, only a few of them are clearly being adopted by a majority of the users. Here are some examples: Global System for Mobile communications (GSM), WiFi©, Bluetooth©.

It might appear that even if more RATs become available with tremendous capabilities, people still need some time to adopt a new technology, and to accept to turn the new offered capabilities into a need. An example of commercial failure is the High performance radio Local Area Network (HiperLAN)/2 standard which was completed until the end, but never adopted by the users, because simultaneously the WiFi© technology was cheaper to build, leading to cheaper devices and first available on the market.

Even if several RATs are simultaneously available into the same device, still each wireless technology is designed to optimally offer a limited set of services for a given range of conditions. To this date, no single wireless technology has been able to better serve all the services than any specific RAT. Accordingly, the key to achieve a continuity of service on a global basis is to have interoperable technologies. Figure 1.2 illustrates the relationships between various systems in terms of mobility versus data rates.

An example of a foreseen convergence is between 3G (the Universal Mobile Telecommunications System (UMTS) standard) and WLAN (the IEEE 802.11 standard), where 3G would allow for a larger coverage area and mobility than WLAN, and WLANs would consist in an affordable and efficient (cheap and quickly installed radio equipments) wireless technology to provide high data bit rates within a limited coverage area (hot spot) for voice and data services.

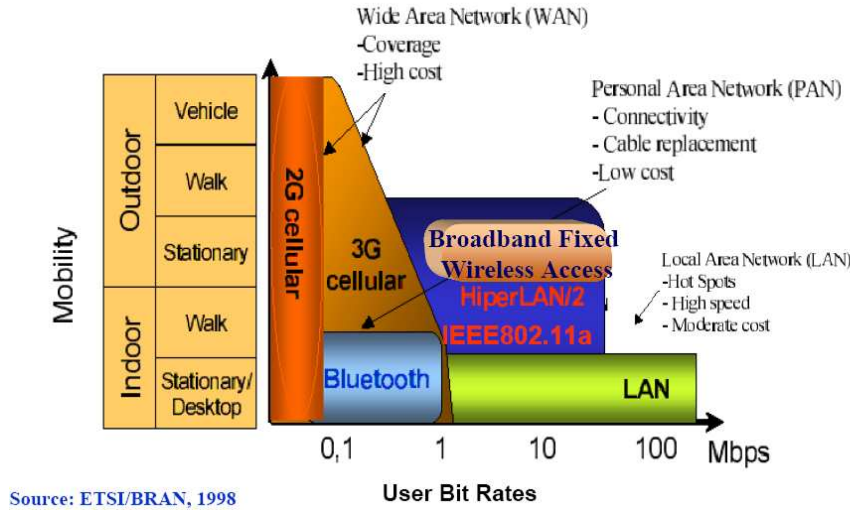


Figure 1.2: Mobility Versus Data Rates for Different Communication Systems

1.2.2 Interference Versus Convergence

Regulators allocate a spectrum band for each new system, and inter-system protection is ensured by setting a spectrum separation (guard bands) between close systems. Even though copping for inter-system interference in 100% of the cases is a very difficult task, there exists a trade-off between over-protecting systems from interference by enforcing over-dimensioned guard bands, and wasting spectrum.

Also, there is another trade-off to consider when dealing with real spatial deployments of several wireless technologies in an area. Indeed, spatially overlapping networks provide a better diversity of service as well as a continuity of service to the users, but at the same time, it increases inter-system interference. For example, WLAN systems' deployments could raise spectral coexistence problems with other systems especially with 3rd Generation (3G) systems, even though they are considered as complementary in terms of offered services.

By deploying more wireless technologies in the same area, the level of radio emissions increases. This is especially true if users are using multi-RAT equipments with multi-homing (using simultaneously several RATs) capabilities: the result is an increase in the density of radio emissions, potentially leading to interference problems. Indeed, the level of protection in the standards are usually designed according to a certain level of aggregated signals corresponding to common densities of users per environment. However, multi-homing capabilities change these densities, while the standard would remain unchanged...

A solution is to allow more dynamic allocation of the spectrum, in order to take advantage of the unused frequencies to mitigate the impact of inter-system and intra-system interference.

1.2.3 Overall Assumptions for the PhD Work

This PhD work tries to propose innovative algorithms able to allocate the spectrum resource more dynamically over time and space to the users, according to their varying needs (varying traffic demand). We assume that a user chooses a RAT for the service offered. In case of congestion, our approach is that the infrastructure tries to acquire more spectrum instead of having the user change RAT (inter-system handover). Accordingly, we always consider the point of view of a single RAT. The example used to validate our algorithm is a WLAN system.

Given the overall convergence of systems toward an IP-world, given that 802.11 systems use a packet-based communication for both data transfer and voice (VoIP) services, given that packet transmission does not reserve a radio resource when not using it, the system we study uses the transmission of radio packets.

We assume more flexible (available spectrum discovery, non-contiguous spectrum for a system, etc) regulatory rules and smarter equipments with reconfigurable capabilities in frequencies.

We will be considering unlicensed spectrum bands as the playground for spectrum reallocation. Such bands are spectrum pools shared by several operators and equipments, with no spectrum license required. In case of spectrum shortage in the initial spectrum band, there might be a possibility to temporarily use spectrum from another band, even from a licensed spectrum band allocated to another technology.

Using a distributed control, user equipments are supporting the task of locally managing the network radio resource. This is similar to the Internet philosophy, where a significant part of the network intelligence is located at the border. Our proposed algorithm to appropriately allocate spectrum follows a bottom-up approach: from the user level to the cell level, and then up to the network level. The idea is to propagate the optimization from the user level to the overall network.

Using an OFDMA context (multi-user OFDM), our algorithms is tested to optimize the OFDM SC allocation over the users. Both the PHY (OFDM) and the MAC (intelligent coordination) layers are jointly optimized (cross-layer optimization), especially we have concentrated with an emphasis put on the MAC layer. As a result, flexibility in spectrum allocation is made possible.

The general conclusions in chapter 8 describes the possible extensions of this work to other scenarios, also explaining what necessary improvements can be done.

1.3 Problem Statement: Challenges to Overcome / Objectives

Our objective is to propose new algorithms for dynamically allocating spectrum over time and space, as well as for solving some of the current limitations in spectrum use.

Here are some of the current challenges to overcome:

- How to achieve a greater flexibility in spectrum resource access and use, while controlling interference level?
- How to go from an under-used spectrum (network resources are dimensioned to absorb the busy hour traffic) to a flexible use?
- How to appropriately adapt the spectrum resource in accordance with the traffic demand?
- How to be more aware of the spatial dimension? In other words, how to more efficiently reuse the spectrum by taking into account the variations in the propagation conditions over space and time?
- How to identify and characterize spectrum holes?
- According to which model should we develop a smarter MAC with cooperation between the users?
- Which regulatory models for spectrum allocation would value spectrum the most?

1.4 Accomplishments and Contributions

Our thesis concentrates on designing new spectrum access algorithms for use by cognitive radios in a WLAN context.

More specifically, we study opportunistic spectrum access at the user level achieving simultaneously the resource allocation and the users' scheduling in a case of multiple users and multiple channels. Our bottom-up chosen approach uses (1) a dynamic bandwidth and position in the band of the spectrum resource, (2) a user-based distributed control for spectrum access and (3) an opportunistic spectrum access (spectrum access contention collectively controlled). This is particularly suitable in dynamic environments and when the number of users is important.

Among others accomplishments, the main ones are listed below:

- We first study the MAC performance of the IEEE 802.11 WLAN systems due to their simplicity of use, great success and the high throughput they theoretically offer. We are especially interested in the IEEE 802.11a version, which uses an OFDM PHYsical (PHY) layer and operates at 5 GHz in an unlicensed spectrum band.
- We coded the PHY and MAC parts of an IEEE 802.11a system into a C++ program. Note that only the parts of the standard having an impact from a radio standpoint were implemented. Deployments of several cells and users can then be investigated.
- WLAN systems are known to suffer several limitations which could be studied using our program such as: capture effect, hidden node problem, inter-channel interference. Usually, the articles found in the literature analyze mean performance at a cell level, or make important assumptions to simplify the calculations, thus removing some of the above limitations in the study. We study the impact of the users' distribution on the spatial unfairness and

the inter-cell interference. The 802.11 MAC mechanisms are very sensitive to the deployments of users and APs. The system performance is studied in terms of goodput and delay both at the cell and user levels.

- We then try to find solutions to solve these WLAN observed problems by proposing innovative MAC mechanisms. Accordingly, our work proposes a new distributed MAC in a cognitive radio perspective. More intelligence and autonomy are given to the users' terminals in the control of their radio access. Assuming a distributed control, the optimization problem of finding the best allocation of OFDM SCs to the cell users that maximizes the UL throughput according to some constraints in spectrum budget per user, turned out to be NP-complete. To find a sub-optimal solution to the problem, a biologically inspired meta-heuristic called "swarm intelligence" is used. This method is able to find a global solution to a problem without the need for a central controller, but only using simple local rules and interactions between geographically distributed individuals. Such an approach, as observed for social insects, has several built-in interesting properties for problem solving such as: scalability, robustness, adaptability to varying conditions.

- After modeling and implementing our algorithm into a C++ program, it is successfully applied to obtain results into an OFDMA WLAN context, starting from the study of a single cell containing several distributed users. Accordingly, we manage to design an innovative OFDMA spectrum access algorithm. It consists in an opportunistic, thus dynamic, self-spectrum planning method adapting the spectrum allocation according to the needs.

- I have published 4 papers in major IEEE conferences (PIMRC, DySPAN, EUROCON) and 2 papers in other conferences/workshops (S3M, WSR), and a journal paper will soon be published in JCM, all as a first author. In addition, I have technically contributed within the E^2R European project in the area of flexible spectrum allocation. Refer to appendix D to find the list of all my written contributions during my PhD.

1.5 Document Organization

This thesis focuses on the design of an opportunistic spectrum access method for cognitive radios. In this chapter, we introduce the context of the PhD work, the assumptions and the challenges to overcome. A brief description of our contributions is also given. Chapter 2 presents the problem of Dynamic Spectrum Allocation (DSA) as well as the current spectrum scarcity in the allocation. Then, chapter 3 reviews the early attempts to use the spectrum more efficiently (spectrum reuse concept), and presents the new foreseen approaches to move to an increased flexibility in spectrum allocation. Our preliminary investigations in trying to find new spectrum allocation algorithms are described in chapter 4. These results have oriented our researches towards the use of opportunistic algorithms taking into account the channel conditions in the spectrum allocation. Then, we have decided to use a user-based approach (bottom-up) based on the swarm intelligence meta-heuristic to allocate spectrum more efficiently and dynamically. More details can be found in chapter 5 that presents our method of opportunistic spectrum allocation using a distributed coordination. An example of OFDM-based WLAN system is studied in chapter 6. This chapter presents the scenarios and results obtained concerning the IEEE 802.11a and IEEE 802.11h systems. Several limitations are pointed out. In chapter 7 we improve the performance of OFDM-based WLAN system by applying the method presented in chapter 5. It consists in achieving a distributed and collaborative UL OFDMA SC allocation. Finally, chapter 8 concludes the achieved work by providing a summary of the main achievements, and suggests directions for future research.

Chapter 2

About Dynamic Spectrum Allocation

2.1 Introduction

In this chapter we will explain what we call a "spectrum resource", what is spectrum allocation, what is dynamic spectrum allocation, and why it is becoming an increasing issue for the entire wireless world.

The important growth in the number of available wireless services confirms the increasing demand of spectrum resource. Thus, understanding the problem of spectrum allocation is of major importance for appropriately using the spectrum for future generations of wireless systems.

Many spectrum holes exist in the currently "fully" allocated spectrum, thus opening new perspectives in spectrum use, as shown in this chapter.

2.2 Definition and Properties of a Spectrum Resource

A resource is often referred to a material element that can be used, exploited, exchanged, and whose use can be controlled by access rules (exclusive use by somebody or shared access by several people, temporally or forever, etc). It cannot be duplicated, and has precise coordinates in space and time.

The spectrum in itself is not a material resource, it is a set of frequencies. By controlling the use of electromagnetic waves propagation, the spectrum becomes a resource capable of remotely transporting energy over space and time, to exchange information. Accordingly, **we define a spectrum resource as a resource available for use in transmission, disregarding of what is going on in reception**.

Figure 2.1 shows the available frequency bands for wireless systems.

We define a spectrum resource according to three **orthogonal axis**: frequency, space and

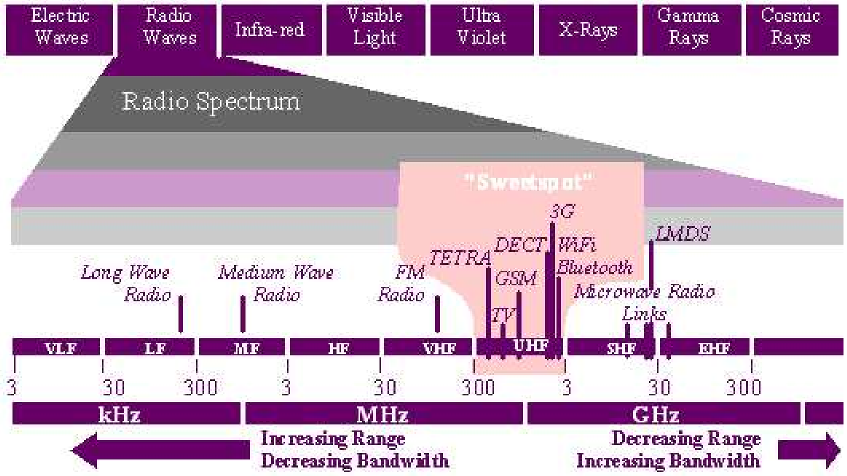


Figure 2.1: Available Spectrum Bands

time. Orthogonal is important because it indicates that the frequencies are independent from time and space.

The spectrum resource has additional specificities compared to a material resource:

- Immediately, always and entirely available anywhere and at anytime (even if using it or not).
- Not consumable.
- Infinite possibility of reuse (simultaneous use) anywhere and at anytime.

There is no spectrum scarcity in transmission as they can all transmit on the same channel, at the same time and from the same location. The problem only comes in reception when trying to extract one message out of the others [3]. Nowadays we do not have the science and the technology sufficient to fully exploit the potential of the spectrum resource, we are only able to design "dumb receivers" [4].Currently, the potential of spectrum reuse over time and space is limited by the receivers, not able to extract some signals from others, leading to interference. However, "interference is only a receiver problem" [4] it is not an inherent property of the spectrum. To limit the interference problems between different systems, regulators have decided to rule the spectrum access. In addition, in order to help the receivers succeed in receiving a transmitted signal, transmitters have restricted their access to the spectrum by applying reuse constraints. The consequence is a reduction in the available spectrum over time and space, resulting in the creation of an artificial scarcity in the spectrum availability.

The current ways of limiting the spectrum access is only intended to allow a correct communication using dumb receivers, but this is not at all linked with a property of the spectrum. When scientists will finally be able to build receivers fully integrating the intimate properties of the spectrum, then the restrictions on spectrum access should not stand anymore.

As an example of the great progresses in digital communications, who could have imagined a successful reception of several overlapping wireless signals in time, space and frequency? This was made possible by the spread spectrum technique (now used in the Code Division Multiple Access (CDMA)). The Ultra Wide Band (UWB) technology is another step in gaining a better understanding of the properties of the spectrum use for communication. However, even if one day we were able to build intelligent interference-free receivers, we would still have to limit interference towards existing dumb receivers.

A conclusion is: spectrum is currently being underused. Nowadays, the spectrum is said to be scarce in reception, but not in transmission. Also, a lot of care should be taken when applying the analogy of rules used for the allocation of material resources to the spectrum. As seen above, the spectrum resource has few common properties with material resources. To solve the problem of spectrum resource sharing, some people have "materialized" the spectrum resource, applying some models taken from the economy or property rules, thus introducing some scarcity in the spectrum. More flexibility should be introduced in spectrum allocation rules in order to stimulate more flexibility in its use.

2.3 "The Tragedy of the Commons"

In "the Tragedy of the Commons" [5], Hardin deals with the problem of a common material resource sharing. Hardin argues in favor of a regulated access to a common resource to avoid the chaos. As an example, consider several users each one having several cows all willing to graze the grass of the same limited field. He claims that in the absence of rules or of an external control, each user (due to the human selfish nature) would try to maximize his own interest, thus overusing the grass resource, resulting in the chaos.

Nowadays, the "spectrum scarcity" argument can be heard everywhere! This vision might come from:

- The regulators: by partitioning the spectrum into dedicated bands allocated to a given RAT, they have created an artificial scarcity.
- The operators: with only a limited access to the spectrum included in their license and dumb receivers, they try to fit in the same space, at the same time and using the same frequencies, an increasing number of users asking for always increasing data rates.

For a given band, the spectrum access rules and constraints strongly influence the available system capacity and then the capacity per user [6]. Applying the same rules as for material resources, the regulators have created a spectrum scarcity, then in turn imposed onto the operators.

The holy grail of wireless communications is to provide seamless availability of applications and thus of spectrum anywhere and anytime for all users. Most of the legacy spectrum allocation techniques have difficulty adapting spectrum fast enough to the traffic demand because the spectrum allocation is not flexible enough compared to traffic variations. Such rigidity comes from: the regulation, the frequency planning parceling the radio resource into rigid contiguous channels, the multiplexing methods and the MAC mechanisms.

As a result of this materialization of the spectrum as a resource, Hardin proposes several options to control the spectrum access as a property (sell as a private property, public property with right-of-use, auction system, merit, lottery, first-come, first-served basis), some of which are currently used by spectrum regulators.

However, owing to the current way of partitioning the spectrum, as well as the increasing number of new wireless standards, the amount of spectrum available for new systems is becoming incredibly small. In addition, spectrum re-farming (the operation of reassigning a spectrum band from one technology to another) is not usually taking place. Thus, once assigned a spectrum band will not be available for other systems before a long period of time.

The following section illustrates by real field measurements the current status of the spectrum bands occupation.

2.4 Current Status of Spectrum Allocation

Several field measurements have been performed by the Shared Spectrum Company [7]) in the USA. They show that the so-called "spectrum scarcity" only applies to the regulatory allocation of the frequency bands (see Section 2.4.1), but not to its use (see section 2.4.2).

2.4.1 Map of Allocation: Spectrum Almost Fully Allocated

This section presents the current status of the spectrum allocation in several countries. Figures E.1 and 2.3 show respectively the radio frequency allocation tables in the USA and in the UK for a range of frequencies between 3kHz and 300GHz. The reader should be aware that we present these figures to illustrate the current saturation of the bands in the allocation, but we do not expect him to be able to read the name of each system indicated on the figure.

As seen in figures E.1 and 2.3, almost 100% of the available spectrum bands are already allocated to existing radio systems, making the allocation of new systems always more complicated. Note that this saturation problem is the same in many other countries.

Accordingly, almost no new system can be allocated spectrum, unless another one is declared obsolete and removed from the allocation chart, or another one shares or reduces its allocated spectrum with new comers. As indicated in the next chapter, several similar initiatives are taking place today.

To emphasize the difficulty of new systems trying to fit in the existing spectrum allocation chart, note that not to harm all the already allocated systems, the new systems are imposed stronger restrictions than the early allocated ones.

Fortunately, as shown in the following section, not all the allocated spectrum is actually being used.

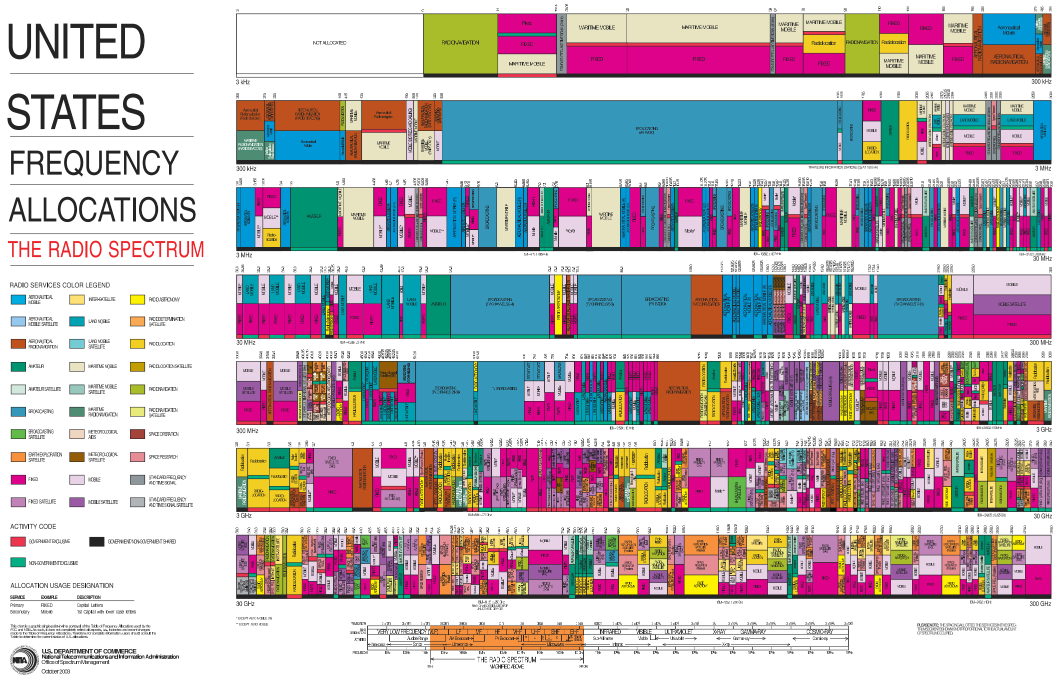


Figure 2.2: Radio Frequency Allocations in USA

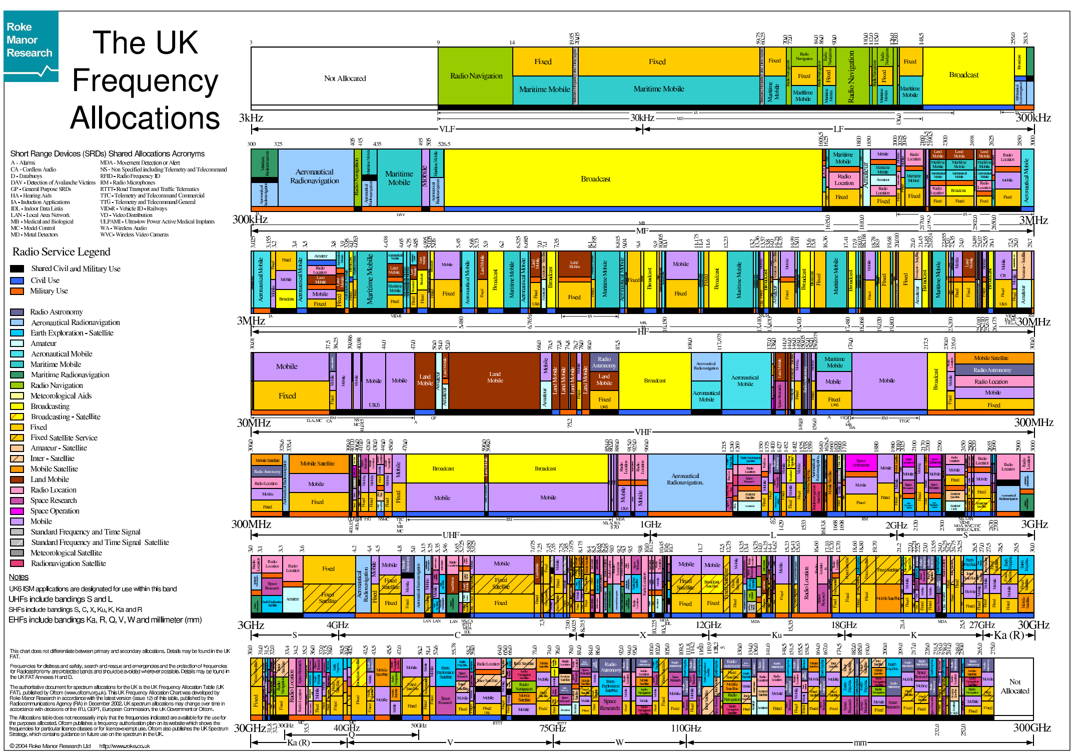


Figure 2.3: Radio Frequency Allocations in UK

2.4.2 Measurements Campaigns: Spectrum Not Fully Used

This section presents spectrum occupancy field measurements obtained in the USA by the Shared Spectrum Company from January 2004 until August 2005 for the National Science Foundation (NSF) [7]. The experience included to measure each band from 30 MHz to 3,000 MHz and at multiple locations with different densities of users. Their obtained results are striking and full of hope regarding the "spectrum scarcity" argument.

Figures 2.4 and 2.5 show the measured spectrum occupancy for different spectrum bands and geographical locations in the USA.

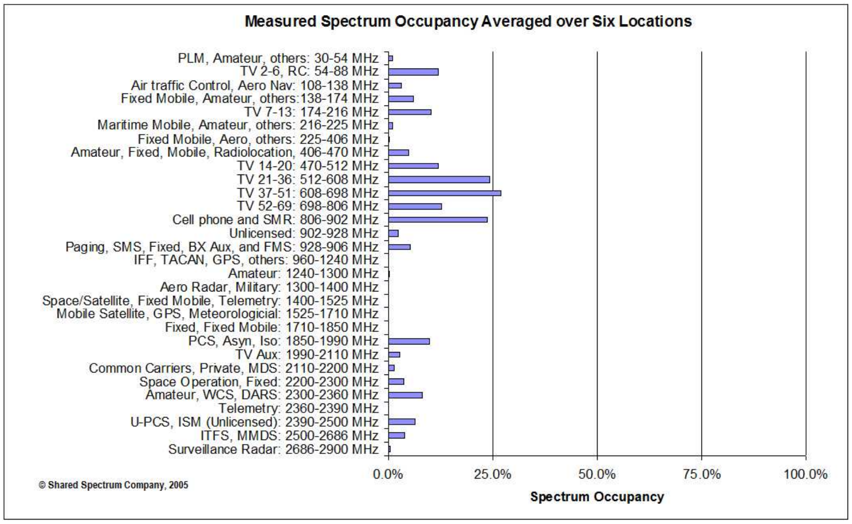


Figure 2.4: Measured Spectrum Occupancy Averaged over Six Locations

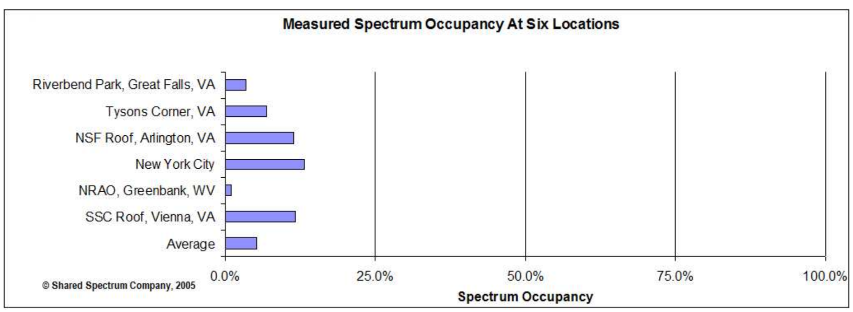


Figure 2.5: Measured Spectrum Occupancy at Six Locations

The study finds significant available spectrum in most bands, showing that only a small portion of the allocated spectrum is used everywhere and all the time. Each bar in the graphs corresponds to the average of the occupancy in each band (figure 2.4) and at each location (figure 2.5). The main results are:

- The average occupancy over all locations is 5.2%.
- The maximum total spectrum occupancy is 13.1% (New York City) and corresponds to the highest density of users.

- The minimum total spectrum occupancy is 1% (National Radio Astronomy Observatory).

These low occupancy levels show that a significant part of the spectrum can be reused to provide additional services. The Shared Spectrum Company concludes that in rural areas, there is enough unused spectrum for a radio reusing the spectrum to provide ten times the capacity of all existing wireless devices together [7].

In the USA, as in many other countries, many "spectrum holes" exist showing that even though allocated, the spectrum bands are only sparsely used by the license holders: the "spectrum scarcity" mostly exists in the regulatory spectrum allocation.

This calls for new frameworks of spectrum access:

- Technical perspectives [8]: use more dynamic algorithms to opportunistically identify and use the spectrum holes.
- Business model perspectives [9]: define appropriate business models of spectrum use for future technologies and services.
- Regulatory perspectives [10]: allocate the spectrum to those "who value it the most", in order to encourage its intensive use.

2.4.3 Exploiting the Spectrum Holes

Now that the existence of many "spectrum holes" have been shown in the "scarce" spectrum allocation, regulators should change the way the spectrum is considered, otherwise no gain will result from this discovery. Indeed, legally speaking, for a given licensed spectrum band, a non-license holder is not allowed to use this licensed band, even if unused. Also, note that in most countries the licensed systems occupy the major part of the entire allocated spectrum.

Nowadays, it is technologically possible to detect spectrum holes and occupy them, even for a short period of time. However, before allowing such an opportunistic spectrum use, regulators expect insurances that no harm (interference) will affect the primary users (the license holders of a considered spectrum band) coming from the secondary users (the non-licensed holders willing to opportunistically use the considered licensed spectrum band).

The next section shows the urgency of undertaking the task of redefining the rules of spectrum allocation as well as of finding adaptive algorithms of spectrum use.

2.5 Forecasted Increase in Spectrum Use

As technology progresses, the following indicators have been noticed:

- Increase in the number of wireless subscribers.

- The volume of data exchanged per user using wireless communication is increasing due to multimedia content.

- Mobile wireless users are always demanding for new services available anywhere and at anytime, assuming a complete transparency of the underlying wireless technology.

According to eTForecasts [11], a market research and consulting company for the computer and Internet industries, shows the growth of cellular subscribers for the main six regions of the world (indicated in figure 2.6). In a nutshell, the number of worldwide cellular subscribers topped 750M in 2000 and is forecasted to surpass 3.2B by 2010 showing a continued importance of the wireless industry.

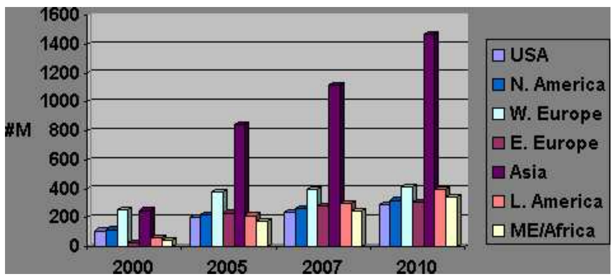


Figure 2.6: Growth of Cellular Subscribers for the Main Regions of the World until 2010

The projections show Asia Pacific became the largest region in 2001, and will grow from 847M in 2005 to 1.47B cellular subscribers in 2010. Asia Pacific will grow its market share from 33% in 2000 to 41% in 2005 and over 45% in 2010. Western Europe was the largest region in 2000 with 254M cellular subscribers, which will grow to 413M in 2010. North America will have 321M cellular subscribers in 2010. Latin America is projected to surpass N. America with nearly 370M cellular subscribers in 2010.

The developing countries will also make tremendous strides in using wireless technology. The cellular phone will be especially important in the countries that have limited land-based telecommunications infrastructure. The handsets will see penetration surpassing 50% of the population in the next decade in the developing countries. Some subscribers will have a low-end handset, but owing to the incredible technological progresses in handsets, by 2015 the low-end handset will far surpass the most advanced smart phone currently available.

Considering the WLAN market, since the IEEE 802.11b (WiFi©) standard creation in 1999, more than 50000 hotspots were already deployed around the world at the beginning of 2005, with a growth of 100% in 2005. The WiFi© is becoming an incredibly popular technology, mainly because it offers an affordable access to high data rates without the need for a spectrum license (it operates in an unlicensed spectrum band) and is easy and quick to install.

Accordingly, the need for more spectrum band is firstly motivated by an explosion in the demand for wireless services and wireless technology. Where will the new spectrum bands come from?

2.6 Conclusion

Waiting for the technological breakthrough to replace the "dumb receivers" by smarter ones, the amount of spectrum allocated to a given technology is limited by regulations. Once allocated, a licensed band becomes the property of the license holder, usually having an exclusive right-of-use of this spectrum band.

However, uncorrelated services usage in time and/or space lead to spectrum holes. Such unused parts of the spectrum could be exploited by other services requiring more spectrum at the same time and space. Depending on the degree of temporal and spatial correlation of the services, the gain of using flexible rules in spectrum access could be important and fulfill users' demand. Many field measurements have shown that the spectrum use is not uniformly spread over the entire spectrum band but it appears as concentrated in some parts of the spectrum (especially for commercial use radio communication systems). As a consequence, spectrum saturation is no more a problem if more flexibility is allowed in its use. However, an evolution of both the regulatory framework and the technological capabilities are required.

In this thesis, we investigate innovative dynamic algorithms to allocate spectrum, thus creating a dynamic spectrum reuse over space and time.

Hopefully, there is no reason for the technological progresses to stop, and each technological breakthrough will open up a new world of applications. The next chapter describes the evolutions of the methods used to regulate the spectrum access.

Chapter 3

Towards an Increased Flexibility in Spectrum Allocation

3.1 Introduction

This chapter deals with the evolution of the spectrum allocation framework at the regulator level as well as at the operator level.

Indeed, as presented hereafter, the way the different wireless systems are allowed to use spectrum is becoming of major importance. In other words, the success and the efficiency of future wireless systems heavily rely on the framework ruling spectrum access in time and space.

This chapter is chronologically organized showing how more flexibility has been introduced over time in the regulatory framework of spectrum allocation, starting from the first attempts of spectrum allocation, pursuing with the DSA techniques and concluding by the cognitive radio initiative.

3.2 First Attempts in Flexible Spectrum Allocation in Cellular Systems

The wireless medium is difficult to control when compared to the wired one. Accordingly, to avoid interference between systems and/or within a system, protection measures have been designed when using spectrum:

- By the regulators: guard bands between systems.
- By the operators: reuse distance between interfering channels.

Regulators have artificially created spectrum scarcity by segmenting the whole spectrum resource, with a right to access each segment based on license costs for commercial use, and preventing not license holders to use the spectrum, even when unused (exclusive access [12]).

This status is amplified in areas of high traffic demand such as urban areas. A large part of the spectrum is allocated to government agencies with only a sparse use compared to commercially used bands.

There exists two types of spectrum:

1. Licensed: a license is required to operate in this band, and is obtained according to various rules. The license holder has an exclusive right of use for this band (but no obligation to use it intensively).
2. Unlicensed: even if this part of the spectrum is regulated, no license is required to operate in it. There is no exclusive use but rather a shared use.

The first wireless equipments were band limited and unable to avoid co-channel interference. This motivated the first regulatory rules based on considering the spectrum as a finite natural resource. Since then, regulators have always centrally and rigidly ensured inter-system protection from interference.

When sharing a finite material resource between an increasing number of users, the per-capita share decrease. To avoid the chaos situation where each user seeks only his own interest in exploiting the shared resource, rules are needed to decide how to access the resource and how much each one receives: sell each good as private property, keep all the resource as a public property but allocate the right to enter them (this is the context for spectrum regulation). The allocation might be on the basis of wealth, by the use of an auction system. It might be on the basis of merit, as defined by some agreed-upon standards. It might be by lottery. Or it might be on a first-come, first-served basis, administered to long queues.

Determining the level of flexibility allowed in access protocol designs is one key element in establishing rules of coexistence, spectrum sharing or even interoperability of several smart devices. Consider the following range of possibilities within a given band, from most to least restrictive [6]:

- A full standard is established, allowing all devices to interoperate.
- All devices are required to follow a minimal standard. Examples include television transmission in USA.
- A more detailed set of constraints is imposed on access protocols. Examples include: Unlicensed Personal Communications Services (UPCS) band.
- The only constraint is a maximum power level. Examples include the US National Information Infrastructure (NII) band.
- There are no constraints on spectrum reuse. Examples include some unlicensed bands in Haiti.

Regulations have an important impact on the landscape of spectrum allocations and communications' system performance. Any set of rules and constraints imposed on a spectrum band will strongly influence the available capacity per user.

Channel allocation schemes can be divided into a number of different categories depending on the comparison basis. For example, when channel assignment algorithms are compared based on the manner, in which co-channels are separated, they can be divided into fixed Fixed Channel Assignment (FCA), Dynamic Channel Assignment (DCA), and Hybrid Channel Assignment (HCA). These techniques are presented hereafter [13]:

- **FCA**:

  - Co-Channel constraints: a set of nominal channels is permanently allocated to each cell for its exclusive use. Here a definite relationship is assumed between each cell and each channel, in accordance to the co-channel reuse constraints.

    In the simple FCA strategy, the same number of nominal channels is allocated to each cell. In order to confine the interference between adjacent cells at minimum levels, cells are organized in clusters. Within a cluster of cells, the whole available radio spectrum is exploited. A portion of the total number of channels is allocated to each cell, while adjacent cells within the same cluster are assigned different groups of channels. The minimum number of channel sets (also represents the minimum number of cells per cluster) required to serve the entire coverage area is related to the co-channel reuse distance.

    An increasing capacity demand can lead to cell splitting, so that a cell contains several smaller size cells. Each smaller size cell provides the same number of channels as the former cell, without the need of a wider spectrum.

  - Adjacent Channel Constraints: interference is more severe in urban areas due to the greater radio frequency noise floor and the large number of base stations and mobiles. Recognizing co-channel interference as one of the two major types of system-generating cellular interference, adjacent channel interference also constitutes a bottleneck in the efficient network operation.

    Adjacent channel interference results from imperfect receiver filters, which allow nearby frequencies to leak into the passband. Such an interference can severely affect the transmission between a receiver and a base station, if an adjacent channel user is transmitting either in a very close range to the receiving antenna or very close to the base station (in this case, on a channel close to the one being used by a "weak" mobile). Careful filtering and channel assignment between adjacent cells can minimize adjacent channel interference which is worse in small cell clusters and heavily loaded cells.

  Even though there exists possibilities of channel exchange between cells, FCA schemes do not adapt to changing conditions and user distributions. Unfortunately, non uniform user distributions happen in reality to some extent. The interest in FCA schemes mostly resides in their simplicity of implementation.

- **DCA**: in contrast to FCA schemes, there is no fixed relationship between channels and cells in DCA. All channels are kept in a central pool and are assigned dynamically to radio cells as new calls arrive in the system. After a call is completed, its channel is returned to the central pool.

The main idea of all DCA schemes is to evaluate the cost of using each candidate channel, and select the one with the minimum cost provided that certain interference constraints are satisfied. The selection of the cost function is what differentiates DCA schemes.

Based on information used for channel assignment, DCA strategies could be classified either as call-by-call DCA or adaptive DCA schemes.

DCA schemes have higher complexity than FCA schemes, but they provide flexibility and traffic adaptability. However, DCA strategies are less efficient than FCA under high load conditions.

DCA schemes consist in a good model for spectrum allocation of channel in unlicensed bands.

- **HCA**: HCA strategies combine advantages of both FCA and DCA schemes. In HCA, the total number of channels available for service is divided into fixed (belonging to cell as in FCA) and dynamic (dynamically requested by cell as in DCA) sets.

3.3 DSA at a Glance

The idea is to use idle spectrum for another service than the legacy one. The DSA has been studied in the projects DRiVE [14] and OverDRiVe [15]. There are three types of DSA:

- Temporal DSA: can provide gain by taking advantage of the temporal orthogonality between services to use the unused spectrum for another service.
- Spatial DSA: can provide gain by taking advantage of the spatial orthogonality between services to reuse the unused spectrum for anoter service.
- Combined temporal and spatial DSA.

The achieved gain is highly dependent on the considered temporal and spatial orthogonality between services. The DSA is interesting because it opens the way to a more dynamic spectrum management with spectrum exchange between systems.

3.4 Unlicensed Spectrum Success

The unlicensed bands are identified by the International Telecommunications Union (ITU) Radio Regulations as an Industrial, Scientific and Medical (ISM) band: between 2400 and 2483.5 MHz, 4.90 GHz and 5.90 GHz as well as between 868 and 928 MHz. Nowadays, an increasing number of wireless technologies operate in these bands, sometimes creating inter-system interference due to the saturation of these bands. Some examples of these systems include:

- WLAN systems: WiFi© (IEEE 802.11b/g/a/...).
- Wireless Private Area Network (WPAN) systems: Bluetooth^{TM} (IEEE 802.15.1).
- Etc.

Such systems do not require any license for emission (no license cost) and are easy to deploy. Some initiatives also mention the development of unlicensed operations in licensed bands as a mean to increase their available spectrum [12].

In unlicensed bands all the channels are shared ("spectrum pool") and can be accessed by any entity. Accordingly, the spectrum use is adapted to the needs. Unlicensed bands are often described as saturated bands, thus showing the success of the spectrum pooling.

3.5 Open Spectrum: a New Vision of the Spectrum Resource

Open spectrum (leaded by David Reed [16]) proposes "to free the airwaves" [17] from regulatory rigid "command-and-control" framework. It consists in a futuristic vision of spectrum use and network organization, incorporating a much more accurate view of the relationship between bits, their physical representations as electromagnetic waves in space, and some tools for manipulating signals (including the ability to build distributed, adaptive, interoperable and self-organizing [18] architectures for communication). In [3], David Reed describes what are the limitations of the current regulatory spectrum allocation rules, and in which direction to concentrate the research efforts to build innovative and efficient systems. In essence, the idea is to have a really dynamic sharing of the spectrum resource between users, and between systems.

Indeed, spectrum shall not be looked as a thing that could be owned as a property (something of value to which someone, by legal right, has exclusive access), but as an open standard. The current policy, however, treats spectrum as if it were a physical thing to be carved up. By focusing on open standards rather than on spectrum-as-thing, the medium becomes far more efficient to transport information and offer far greater capacity. In this revolution, the open spectrum policy will allow anyone to send signals across any range of spectrum without prior permission, with the minimum set of rules required to enable the success of a "wireless commons".

Open spectrum vision supporters claim that interference is not a law of nature [19] but rather "interference is a receiver problem" [4] not being able to separate signal from noise and extract the useful information. But as receivers become smarter, there is less need for a regulatory rule that is the equivalent of licensing only one person to talk at a time, because interference is less and less becoming an issue. Today's receivers are capable of separating signal from noise well enough that they do not anymore require "exclusive" use of the frequency bands they are tuned onto. Some examples include very wideband modulation techniques such as DSSS (IEEE 802.11b), OFDM (IEEE 802.11a/g), UWB and many others use new technologies to spread information across many frequency bands, creating very high transmission rates at low cost with very little degradation even in noisy environments. This is even more true considering future SDR or Cognitive Radio (CR)s because of their live reconfigurability capabilities.

As a result, open spectrum is likely to enable innovation in the wireless world by permitting a new approach [20] to spectrum allocation. While unlicensing more spectrum would certainly help the development and deployment of new technologies, it would not allow the open and ubiquitous access that is expected. However, open spectrum will not create overnight, it is a slowly growing transformation. Open spectrum will not create more bandwidth (as a natural resource) but as more people join a wireless network, there can be a cooperative gain effect whereby the network actually increases its capacity [4]. Open spectrum is expected to do for wireless communications what the Internet has done for networking computers. Indeed, the Internet (an end-to-end network) architecture teaches us that decentralization works. In fact, central control and regulation would have kept the Internet from becoming the force that it has: put the intelligence at the border [21].

3.6 New Opportunities with the Cognitive Radios

The Cognitive Radio vision was first introduced by Mitola [22]. It embraces many different concepts [22] [23] [24] including a new way of managing spectrum allocation. It is based on an SDR architecture.

3.6.1 SDR

The SDR concept, envisaged in the early 90s, and researched in the military and commercial domains, is an intense research domain. The introduction of SDR equipments in future wireless networks will be one of their major evolution, allowing convergence toward an IP-based core network and ubiquitous seamless access (2G, 3G, broadband, broadcast) in a context of hierarchical and self-organizing networks. The Federal Communication Commission (FCC) issued in 2001 formulated the problems associated to regulating SDR equipment in a "First Report and Order". This has also recently addressed in [25]. The docket is a first official approach to capture the regulatory issues related to SDR equipment and tries to outline the type of changes that might be done to equipment without infringing the rules associated. In this attempt, the FCC defines a Software Radio as:

".. a radio that includes a transmitter in which the operating parameters of the transmitter, including the frequency range, modulation type or maximum output power (either radiated or conducted) can be altered by making a change in software without making any hardware changes components that affect the radio frequency emissions."

The SDR equipment reconfiguration is being driven by a flexible spectrum allocation algorithm. As such, the SDR can be seen as the technological aspect of the flexible spectrum allocation. SDR radios can be used in both contexts: licensed and unlicensed spectrum.

Using a SDR is a mean to achieve radio communication equipments (user and infrastructure equipments) reconfigurability to bring more flexibility in a single equipment. Accordingly, an SDR equipment can synthesize over time new functionalities to become another equipment. It can be successively several equipments in once. An ideal software radio architecture is presented in Figure 3.1.

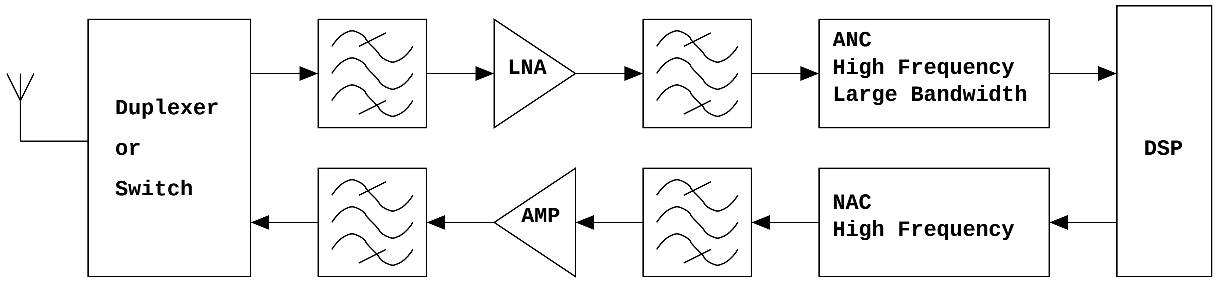


Figure 3.1: Ideal SDR Architecture

Such a radio is digital after the antenna and has a large bandwidth. Also, the architecture is fully reconfigurable. However, such a software defined radio is not technologically possible today. Indeed, several limitations are still creating obstacles such as: conception of ADC/DAC converters (simultaneously incorporating the requirements from several technologies), and antenna capabilities (trade-off between having a small antenna size and a broadband antenna).

However, different levels of reconfiguration (depending on the equipment class) can be considered and some of them are already a reality today. The FCC classically foresees three classes of permissive changes that may be undertaken (on authorized equipment only) without the need for filing/new certification:

- First class: changes that do not affect or degrade RF emissions.
- Second class: changes other than frequency, modulation or power and that does not degrade RF emissions. This class foresees the software alteration of frequency ranges, but the changes must not alter the device hardware).
- Third class: still under discussion, but it foresees a simplified registration for new software (that may change/degrade the RF emissions) on certified equipment, alterations to the hardware would not be permitted.

In addition, to the technological barriers against SDR, there is an economical question: will be the cost of an SDR equipment (which is likely to be higher than current equipment) be compensated by the gain offered in terms of performance. This addresses the cost of reconfiguration: how often, how rapidly and how much extra power is spent in this operation, will a SDR terminal reconfigure some of its parts and what is the total benefit brought? The gain of equipment reconfiguration will be high if the DSA algorithm driving it is effectively providing gain.

3.6.2 CR

From the users' perspective, SDR equipments will introduce a lot of flexibility, in a seamless and transparent way. Nevertheless, SDR terminals envisaged today, do not allow the users to manage their radio profiles in the most efficient ways. The cognitive radio principle proposes the introduction of a language Radio Knowledge Representation Language (RKRL) [22] enabling intelligent machine-to-machine communication within a wireless network and machine learning, also allowing an interpretation of the User Equipment (UE) environment, thanks to the structure of the language. Using this interpretation, the UE is then able to take decisions (machine-learning) in order to optimize its profile depending on its context. The cognitive radio embraces many different concepts, including a new way of managing the use of spectrum (spectrum awareness) [26] [27] [28]. Indeed, current spectrum regulation trends are to disconnect spectrum and technology. Progresses in location- and environment-aware computing are vital for the introduction of a cognitive radio.

The cognitive radio initiative assumes smarter terminals with reconfigurable capabilities especially in the frequency domain but also in terms of functionalities (SDR equipments) than current mobile phones. A cognitive radio terminal is context-aware [22] (location, mobility, environment) and even more, it is self-aware. Thus, it has the knowledge of its own state (hardware and software configuration) and is able to reason and take decisions to trigger changes and to learn and adapt to its changing environment to maintain a given level of performance.

The term CR identifies the point at which "wireless personal digital assistants" (PDAs) and the related networks are sufficiently computationally intelligent about radio resources and related computer-to-computer communications to:

1. Detect user communications needs as a function of use context,
2. Provide radio resources and wireless services most appropriate to those needs.

As the radio has to be context-aware, it must interact with the external world. The interactions are accomplished thanks to the cognition cycle consisting in a series of related action such as: observe, orient, plan, decide, act and learn. The cognition cycle implies a large scope of hard research problems for Cognitive Radio in all technical domains (speech processing, scheduler, machine learning).

One fundamental issue in the application of the CR technology is the trade-off between intelligence in the infrastructure and intelligence in the mobile devices. Spectrum Pooling is a novel approach to Radio Resource Management (RRM) enabled by CR. Pooling is the rental of public and government spectrum by the present owners to cellular service providers for a time as brief as one second since federal, state and local governments could generate revenue streams by renting channels that are not currently in use. Spectrum that could be made available for pooling includes all the bands allocated for mobile terrestrial uses, and not satellites, radar and radio navigation bands.

In order to enable spectrum pooling, a new class of protocol named Radio Etiquette has been defined. This protocol provides for the advertising and rental of spectrum. Etiquette includes the spectrum renting process, assured back off to authorized legacy radio, assured conformance to precedence criteria and an order-wire network. The etiquette allows to designate a user (by international mobile subscriber identification), a channel, or any combination of [user x time x space x frequency] with a specific precedence (1 Emergencies, 2 Government, 3 Public interest, 4 Commerce, 5 Other). To rent some spectrum from legacy radios, a CR must transmit all these parameters when accessing free resources. If a legacy radio enters the band with a higher precedence than the one who occupies the band, the radio etiquette would specify that CR has to change bands and modes. If each renting participant is location aware, the pooling efficiency will be increased. Knowing the propagation conditions, the location, distributed power management could be introduced. If each radio is environment aware, the efficiency of pooling is also increased, for instance, if the radio detects that it is entering a building, it tries to connect to a WLAN. Examples of spectrum pooling studies include [29] and [30].

The next section describes the XG DARPA project aiming at building smart radios.

3.6.3 XG DARPA Technology

The XG Darpa initiatives [31] [32] can be classified in the cognitive radio context.

The Defense Advanced Research Projects Agency (DARPA) neXt Generation (XG) communications program is developing a new generation of spectrum access technology [33]. In order to address the scarcity and deployment difficulty problems, XG is pursuing an approach wherein static allotment of spectrum is complemented by the opportunistic use of unused spectrum on an instant-by-instant basis, in a manner that limits interference to primary users. Indeed, measurements campaign lead by the Shared Company [7] have shown the existence of "spectrum holes": some parts of the spectrum remain unused at some time and/or place. The management of spectrum is now placed in each radio, where it can assess the actual situation at each instant in time, rather than have to be deconflicted in advance for any possible situation of time, position, signal, propagation, etc.

The XG approach, termed opportunistic spectrum access, consists in two new regimes:

- Spectrum agility: a new spectrum access behavioral regime consisting of technologies that senses and characterizes the environment, identifies and distributes spectrum opportunity information, and allocates and uses these opportunities commensurate with the demand, in an interference-limiting manner.
- Policy agility: a new regulatory control regime consisting of methods and technologies for controlling such opportunistic spectrum access behaviors in a highly flexible, traceable manner using machine understandable policies. A key aspect of the XG approach is that it is policy controlled and that there is a decoupling of policies, behaviors and protocols. Decoupling allows adaptation to policies that vary over time and geography. Technology can be developed in advance of policies, and worldwide deployment would be greatly simplified. It is a totally new "software-based" policy regime that allows policies to be decoupled from the implementation and changed dynamically.

While conceptually simple, the realization of opportunistic spectrum access is highly challenging. Although recent years have seen some of the components for opportunistic spectrum access mature (e.g. software radios), there is still a long way from a prototypical system. Further, no work exists in the area of decoupling the policies from the implementation. This yawning gap between current state-of-the-art and what is required for opportunistic spectrum access is the motivation behind XG. Indeed, several problems must be solved: sensing over a wide frequency band; identifying the presence of primaries and characterizing available opportunities; communication among devices to coordinate use of identified opportunities; and most importantly, definition and application of interference-limiting policies, and utilization of the opportunities while adhering to such policies.

3.7 Spectrum Trading

Spectrum trading is one of the tools available to flexibly allocate spectrum for or between systems. It consists in a transfer of the spectrum rights or in a change of the spectrum use. Favoring regular spectrum exchanges between operators consists in a smooth transition from a fully regulated spectrum towards an opportunistic use of the entire spectrum resource. The spectrum trading initiative is lead by the UK (OFCOM), the USA (FCC) and New Zealand. It consists in a vision where the spectrum resource is treated as an economic good and can be sold, leased, etc. It favors a more dynamic spectrum sharing and exchange framework using rules based on market laws. In this context, it is necessary to assign a price to each portion of the spectrum before being able to exchange spectrum parts. Spectrum exchanges between licensed holders are likely to consist in between 1-10 trades a year across Europe.

Economists are very interested in applying these models to any resource exchange, including spectrum. However, any model of spectrum allocation and sharing should be regularly updated or replaced by a more appropriate one, really taking into account the intimate properties of spectrum, as discovered by spectrum research. Accordingly, the spectrum trading model, if applied, should not be the last model for ruling spectrum access.

3.8 Conclusion

How spectrum should be best used, according to which model of sharing between competing systems or users? This question is still not solved today. Many new initiatives of spectrum access greatly differ from the way first wireless systems were allocated spectrum.

Motivated by the current congestion of the spectrum allocation charts, regulators are trying to find new models for ruling spectrum access by wireless systems. The applications from the wireless business are increasingly spreading to many domains, thus raising the interest of economists. Indeed, the artificially rare spectrum resource is becoming a major challenge in the future, as we are today witnessing a common effort from regulators, engineers and economists to develop a common framework.

Chapter 4

Preliminary Analysis

4.1 List of Notations for this Chapter

Table E.1 presents the notations used in this chapter.

Note that these notations are further explained in the remaining of this chapter.

4.2 Introduction

In this chapter, several combined routing and scheduling policies are compared. Our objective is to identify the best combined routing and scheduling policies to maximize a system throughput. The system throughput is the sum of all the contributions from all the active users over a given period of time. We define "routing policy" as the rule according to which a user chooses to use another channel than the previous one. We define "scheduling policy" as the rule according to which a channel is assigned to a given user for a period of time. Accordingly, a combined routing and scheduling technique consists of the set of rules assigning users to channels and channels to users. We define "channel" as a spectrum part exclusively used by a user for information exchange during a period of time.

After defining a spectrum (radio) resource (also called channel), the modeled and simulated results from several combined routing and scheduling methods are compared. The final objective is, given an amount of spectrum resources (several channels), how can we optimally assign the channels to the users over time in order to maximize the system throughput, is there a best method?

Notation	Description
β	Parameter used to weight the service rate (e.g.: $\beta\mu$)
μ	Mean service rate (exponential law)
μ_1	Mean service rate of the reservation phase (refer to method b) hereafter)
μ_2	Mean service rate of the data transmission phase (refer to method b) hereafter)
μ_i	Mean service rate for station i
σ	Duration of a backoff counter decrement (refer to [2])
a, b	Integer parameters used to decompose M into several groups
e_i	Rate of visit of station i
$p_i(k, m)$	Marginal probabilities at station i for a network of m clients
C_i	i^{th} server for a station with multiple servers
CW_{min}	Minimum contention window size using the IEEE 802.11a standard
L	Payload of the transmitted packets (in bytes)
M	Total number of clients (geographically distributed users) in the network
N	Total number of servers (radio channels) in the network
$Q_i(M)$	Mean number of clients at station i
$Q(M)$	Mean number of clients in the network
$R_i(M)$	Mean traveling time through station i
$R(M)$	Mean time per client between, after his service and the end of his next service
T_{ACK}	Duration of an ACKnowledgment (ACK) control message (refer to [2])
T_{CTS}	Duration of a Clear To Send (CTS) control message (refer to [2])
T_{DATA}	Duration of a DATA message (refer to [2])
T_{DIFS}	Duration of a Distributed Inter-Frame Space (DIFS) waiting period (refer to [2])
T_{RTS}	Duration of a Request To Send (RTS) control message (refer to [2])
T_{SIFS}	Duration of a Short Inter-Frame Space (SIFS) waiting period (refer to [2])
$X_i(M)$	Throughput of station i
$X(M)$	Throughput of the reference station

Table 4.1: List of Notations for this Chapter

4.3 Formally Defining a Radio Resource

We are interested in the methods of appropriate spectrum sharing between several users, in order to allow a successful reception of the transmitted information, while maximizing the system throughput.

We assume a cell model, with a radio access infrastructure equipment (e.g.: AP, Base Station (BS), etc) and several geographically distributed user equipments attached to each cell. In this context, the transmission of UL messages is a "many to one" operation, whereas the transmission of DownLink (DL) messages is a "one to many" operation.

For each transmission direction (UL or DL) several multiple access methods exist (Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), CDMA, Space Division Multiple Access (SDMA), OFDMA, etc) to ensure a successful:

- Reception of the transmitted messages.
- Spectrum sharing.
- User separation.

Accordingly, we will refer to "radio resource" as a volume occupied by each message in the following orthogonal system of axis: frequency, time, (code,) and space. The challenge is to achieve a maximum use (in transmission) of the spectrum shared by several users while maximizing the successful reception of the messages. All along our work, we assume the transmission of radio packets.

4.4 Performance Analysis of Combined Routing and Scheduling Methods

4.4.1 Introduction

The problem of channel access and scheduling between several users and cells in a radio system is a complex problem to solve while trying to avoid interference. Classical resource allocation and scheduling techniques consider the cell as a whole and do not go below cell area. However, we will show in the following sections that the solution leading to the best system throughput must consider the users' channel conditions in the channel allocation.

For a given period of time (let us say a frame), the total number of data bits that can be transported by all the packets is fixed by the multiple access technique used, the modulation used, the total available spectrum, the received power, etc. However, it is not possible to go higher than that, unless another dimension is added. Thus, the total available resource is limited. An important note is that the maximum density of users per area (in a real world) is also a finite dimension! And the maximum transmit demand per area is also limited!

Packet transmit durations can be varied due to the reasons presented below:

- Variation in the amount of payload bits of information included in the packet.
- Variation in the amount of spectrum used for the transmission,
- Variation in the modulation (according to the current Signal to Interference Noise Ratio (SINR) level). This is the Link Adaptation (LA) mechanism.

However, in the remaining of this chapter, we will assume that the variable packet transmit durations come from different modulations used.

4.4.2 Queuing Model

The goal of this chapter is to model and compare several policies of channel exchange and access, in terms of system throughput. We assume a group of N channels, M users geographically distributed. We assume non-interfering channels (perfect isolation between different channels), as well as no collision: each user and infrastructure equipment has the same and perfect knowledge of the system. It is an ideal system. Our scenario corresponds to a single cell, with some groups of users. The results from this study can be used for UL modeling. Each of the following method proposes a different approach to the following question: "which user transmits (resp. receives) using (resp. on) which channel?". Also, all the users are assumed to follow the same mean service rate following an exponential law of rate μ. Accordingly, statistically, all the users are equal to a mean user. As a result, even though none of the presented policies control the fairness in service time between users, there is no unfairness problem. But in a more realistic scenarios spatial unfairness could arise and should be controlled. The variation in service duration accounts for variations in channel conditions and various modulations: high SINR levels would lead to a short packet duration, whereas low SINR levels would lead to a long packet duration. All the channels are assumed to occupy the same bandwidth. Our queuing model assumes a First In First Out (FIFO) queuing policy in order for the clients (users) to access servers (channels). A station refers to a FIFO queue and one or several servers. Figure E.2 presents the notations for the current study.

The parameters used in the above Mean Value Analysis (MVA) algorithm [34] correspond to:

$$\begin{array}{ll}
\mu_i & = \text{Service rate for station i} \\
e_i & = \text{rate of visit of station i} \\
M & = \text{Total number of clients (users) in the network} \\
N & = \text{Total number of servers (channels) in the network} \\
p_i(k,m) & = \text{Marginal probabilities at station } i \text{ for a network of } m \text{ clients} \\
Q_i(M) & = \text{Mean number of clients at station } i \\
Q(M) & = \text{Mean number of clients in the network} = \text{M} \\
R_i(M) & = \text{Mean traveling time through station } i \\
R(M) & = \text{Mean time per client between, after his service and the end of his next service} \\
X_i(M) & = \text{Throughput of station } i \\
X(M) & = \text{Throughput of the reference station}
\end{array} \tag{4.1}$$

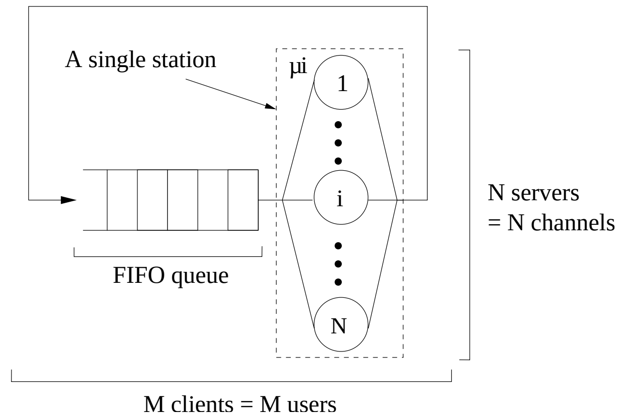


Figure 4.1: Queue Model Notations

Several methods are compared in the following paragraphs:

- **a) Fixed:** users are organized in groups of more than 1 user, and there is no more than 1 channel available per group. Within each group, all users must use the same channel for both transmission and reception.
- **b) Common and data Channels:** 1 common channel and $N-1$ data channels. Before each transmission, a user must win the contention (for the data channel reservation) on the common channel in order to find an idle data channel. Then, the data transmission is done using an idle data channel.
- **c) Single Channel:** a single channel is available for the M users, but it has a service rate of $\beta\mu$. This single channel consists of all the N channels put together.
- **d) Sequential:** a user randomly chooses a channel out of N and after each transmission completion he sequentially chooses the next channel index.
- **e) Random Pool:** the system consists of N channels organized as a common pool. All channels are accessible by all the users as soon as a channel becomes idle.
- **f) Probabilistic Routing:** the system consists of N channels that can be accessed by any user. A user randomly chooses a channel before a transmission and then waits for this channel to become idle, before being able to Tx.
- **g) Opportunistic:** the system, consisting of N channels, always schedules the user with the best SINR on a given channel in order to minimize the transmission duration.

This study takes into account the SINR levels variations over time, included in the variable packet (service) duration. However, methods a) to f) do not consider SINR levels to influence the channel choice, rather they do it in blind or following a given fixed allocation pattern. On the contrary, Method g) is expected to give the best system throughput results because it always schedules the user with the best channel conditions, leading to the best throughput at any given time.

The system aggregated throughput is defined as the collection of all the payload transmitted by all the users during a given period of time, considering all the wait time and the time for transmitting control messages.

Our study will make use of the queuing theory, a powerful analytical tool to evaluate the performance of queuing networks. We modeled the service rate as an exponential law of rate μ because it leads to tractable calculations. When considering more realistic models, the obtained results should not change the rank of performance between the different methods.

A user transmitting a packet of random duration, leads to a random wait time between users' transmissions. All servers have the same service rate, thus $1/\mu$ represents an average packet transmit duration. Clients represent flows. Transmitted packets do not need to be acknowledged as each packet transmission is assumed to be perfect (i.e.: correct reception).

We assumed a closed queue network with a constant number of users (clients) in the system, and we compared the different methods in saturated conditions: each user has always something to transmit (UL). In such systems the stability is always ensured. The users' channel conditions are distributed according to the same exponential law. The packet payload is the same for all users and equal to $L = 1500$ bytes.

In our study we have used the following two constraints: no spatial channel reuse (only N channels can be used), and a user can simultaneously only be assigned a single channel.

To solve (routing and scheduling) the following closed network of queues we used the MVA algorithm [34]. It is an iterative algorithm applicable to calculate the performance of a network of connected queues. We have considered two cases in our study:

- **Network composed of one or several single server stations**: in this case the MVA algorithm is expressed as follows, using our notations:

$$\boxed{\begin{array}{ll} \text{Initialization: } Q_i(0) = 0 & i = 1, ..., N \\ & \\ \text{For } m \text{ from 1 to } M \text{ do} & \\ \quad R_i(m) = \dfrac{1}{\mu_i}(1 + Q_i(m-1)) & i = 1, ..., N \\ \quad X(m) = \dfrac{m}{\sum_{i=1}^{N} e_i R_i(m)} & \\ \quad X_i(m) = e_i X(m) & i = 1, ..., N \\ \quad Q_i(m) = R_i(m) X_i(m) & i = 1, ..., N \end{array}} \tag{4.2}$$

- **Network composed of one or several multiple server stations**: in this case, we use the property that a multiple server station i with C_i servers is equivalent to a station with a rate μ_i depending on the state m_i such that:

$$\mu_i(m_i) = \begin{cases} m_i \mu_i & \text{if } 0 \leq m_i \leq C_i \\ C_i \mu_i & \text{if } C_i \leq m_i \leq M \end{cases} \tag{4.3}$$

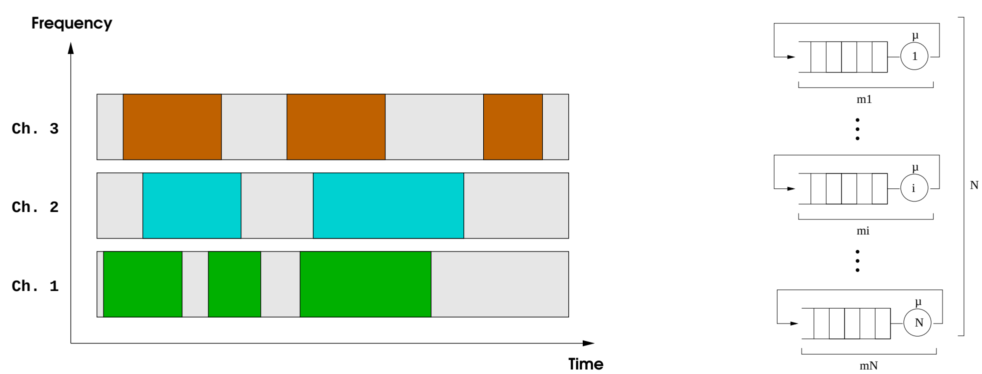
Thus, in this case the MVA algorithm is expressed as follows, using our notations:

$$
\boxed{
\begin{array}{ll}
\text{Initialization: } p_i(0,0) = 0 & i = 1, ..., N \\
& \\
\text{For } m \text{ from 1 to } M \text{ do} & \\
\quad R_i(m) = \displaystyle\sum_{k=1}^{m} \frac{k}{\mu_i(k)} p_i(k-1, m-1) & i = 1, ..., N \\
\quad X(m) = \dfrac{m}{\displaystyle\sum_{i=1}^{N} e_i R_i(m)} & \\
\quad X_i(m) = e_i X(m) & i = 1, ..., N \\
\quad Q_i(m) = R_i(m) X_i(m) & i = 1, ..., N \\
\quad p_i(k,m) = \dfrac{X_i(m)}{\mu_i(k)} p_i(k-1, m-1) & i = 1, ..., N \text{ and } k = 1, ..., m \\
\quad p_i(0,m) = 1 - \displaystyle\sum_{k=1}^{m} p_i(k,m) & i = 1, ..., N
\end{array}
}
\qquad (4.4)
$$

The following sections describe the different policies to compare and provide the corresponding analytical expressions of their performance. In the following figures' scenarios, each user is given a different color, and each colored rectangle corresponds to the operation of a user to transmit.

4.4.3 Method a)

Using this method, each group of users acts separately and a group is assigned only 1 channel out of N. On each channel there is a contention between all the users of the same group. There is total of M users in the system equally distributed in each group leading to approximately the same load per group. Once a user is assigned to a group, it will then stay in this group forever. Figure 4.2 shows the scenario and queue model for this method.



Scenario with
3 Channels and 3 Users

Queue Model with
M Channels and N Users

Figure 4.2: Method a)

The system consists of $\Delta(M,N)$ servers effectively used such that:

$$\Delta(M,N) = \begin{cases} M \text{ if } M < N \\ N \text{ if } M \geq N \end{cases} \tag{4.5}$$

Assuming an equal load between servers, there are m_i clients per server such that for $i = 1, ..., N$, we have: m_i clients per server. Servers are independent.

For a given station i $(i = 1, ..., N)$ we have:

$$\begin{aligned} &\sum_{i=1}^{N} m_i = M \\ &Q_i(m_i) = m_i \\ &X_i(m_i) = \mu \\ &R_i(m_i) = \frac{m_i}{\mu} \end{aligned} \tag{4.6}$$

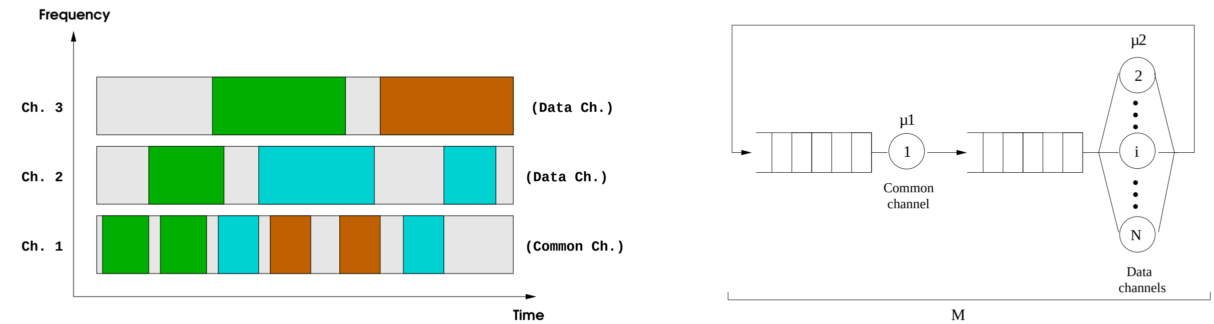
Considering the entire system (composed of the N cells) we obtain:

$$\boxed{\begin{aligned} &Q(M) = \sum_{i=1}^{N} Q_i(m_i) = M \\ &X(M) = \sum_{i=1}^{N} X_i(m_i) = \mu\Delta(M,N) \\ &R(M) = \frac{\sum_{i=1}^{N} m_i R_i(m_i)}{\sum_{i=1}^{N} m_i} = \frac{(N-b)a^2 + b(a+1)^2}{\mu M} \\ &\text{System aggregated throughput} = X(M) \times L = \mu L \Delta(M,N) \\ &\text{Mean wait time per user to get access to a channel} = R(M) = \frac{(N-b)a^2 + b(a+1)^2}{\mu M} \\ &\text{with} \begin{cases} M = aN + b, \ \{a,b\} \in \mathbb{N} \\ a = \text{int}\left(\dfrac{M}{N}\right) \\ b = M - N \cdot \text{int}\left(\dfrac{M}{N}\right) \end{cases} \end{aligned}} \tag{4.7}$$

4.4.4 Method b)

The system uses a total of N channels decomposed between 1 common channel and $N-1$ data channels. Transmissions of the data frames are achieved on separate data channels, but the contention is achieved on a single common channel for all the M users. Figure 4.3 shows the scenario and queue model for this method.

The queue model consists of 1 station with a single server and 1 station with $N-1$ servers. Only transmissions from station 2 are useful data transmissions, while station 1 is used as a channel reservation phase.



Scenario with
3 Channels and 3 Users

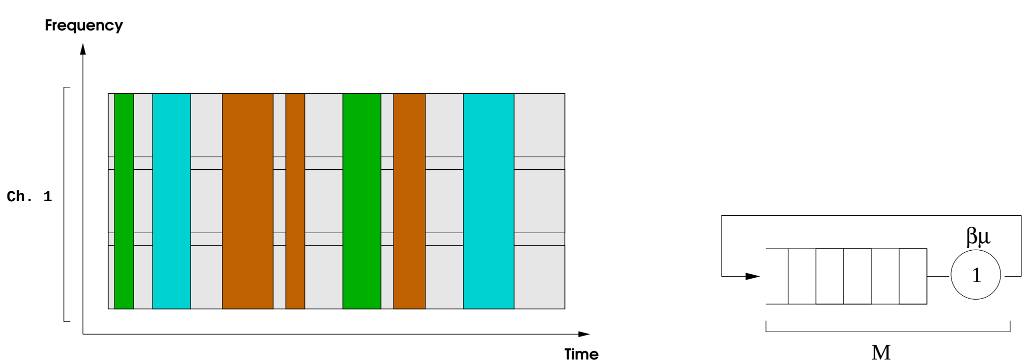
Queue Model with
M Channels and N Users

Figure 4.3: Method b)

This method was proposed in [35] as an improvement to the standard 802.11a MAC protocol (using only omnidirectional antennas) for ad-hoc networks. A greater sharing of the channels is achieved with this method compared to the classical one, thus maximizing the use of every idle channel. However, as every transmission has to negotiate a data channel through the common channel, the common channel becomes as congested as there are contending users. This can become a bottleneck for the system thus reducing the system throughput compared to other methods.

4.4.5 Method c)

There is only a single channel composed of all the N channels put together, and operating at service rate $\beta\mu$ (with $\beta \leq 1$). All the M users in the system must contend through this channel before transmission. Figure 4.4 shows the scenario and queue model for this method.



Scenario with
3 Channels and 3 Users

Queue Model with
M Channels and N Users

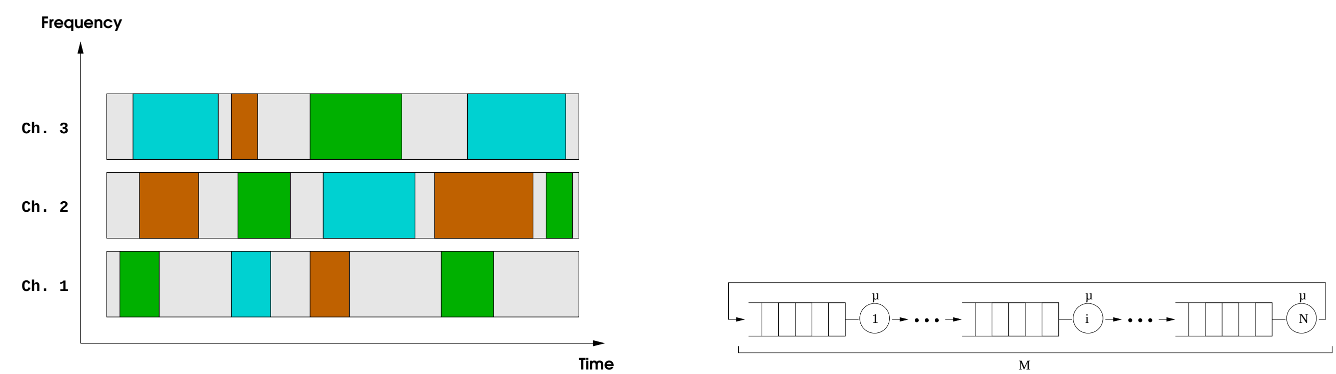
Figure 4.4: Method c)

Considering the above system we have:

$$\boxed{\begin{array}{l} Q(M) = M \\ X(M) = \beta\mu \\ R(M) = \dfrac{M}{\beta\mu} \\ \text{System aggregated throughput} = X(M) \times L = L\beta\mu \\ \text{Mean wait time per user to access the channel} = R(M) = \dfrac{M}{\beta\mu} \end{array}} \tag{4.8}$$

4.4.6 Method d)

The system is composed of N channels and M users. Each user first randomly chooses a channel. After completing a transmission, it will then contend for the next channel in the list of channel indexes $1, 2, ..., N$ ordered by increasing order. Even though the next channel in the list is busy, he waits in the queue for this channel. Accordingly, there is a rotation on the channels used by a given user. Figure 4.5 shows the scenario and queue model for this method.



Scenario with
3 Channels and 3 Users

Queue Model with
M Channels and N Users

Figure 4.5: Method d)

In this system we have: $e_i = 1, \forall i = 1, ..., N$.

For a given server i we have:

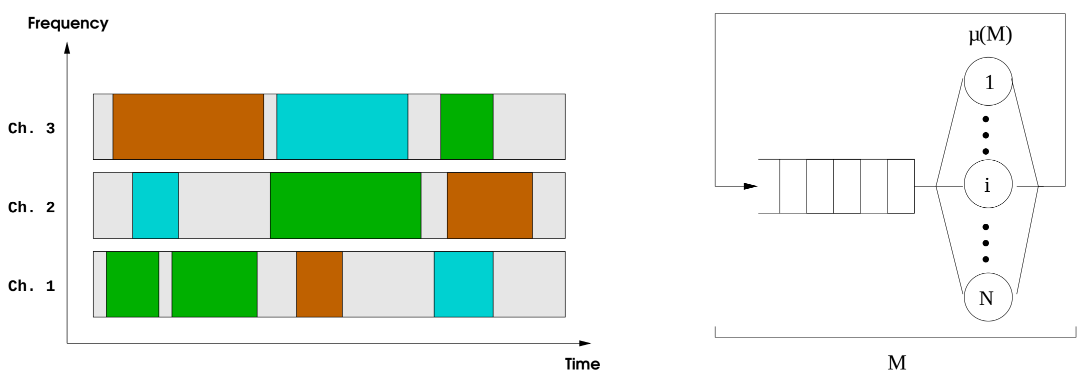
$$\begin{array}{l} Q_i(M) = \dfrac{M}{N} \\ X_i(M) = \dfrac{\mu M}{N + M - 1} \\ R_i(M) = \dfrac{N + M - 1}{\mu N} \end{array} \tag{4.9}$$

Considering the entire system (composed of the N servers) we obtain:

$$\boxed{\begin{aligned}
&Q(M) = \sum_{i=1}^{N} Q_i(M) = M \\
&X(M) = \sum_{i=1}^{N} X_i(M) = \frac{\mu M N}{N+M-1} \\
&R(M) = \frac{1}{N}\sum_{i=1}^{N} R_i(M) = \frac{N+M-1}{\mu N} \\
&\text{System aggregated throughput} = X(M) \times L = L\frac{\mu M N}{N+M-1} \\
&\text{Mean wait time per user to access a channel} = R(M) = \frac{N+M-1}{\mu N}
\end{aligned}} \tag{4.10}$$

4.4.7 Method e)

The system is composed of N channels organized as a common pool. All channels are accessible by all the M users as soon as a channel becomes idle. Figure 4.6 shows the scenario and queue model for this method.



Scenario with
3 Channels and 3 Users

Queue Model with
M Channels and N Users

Figure 4.6: Method e)

The law of service for the servers has the following expression:

$$\mu(M) = \begin{cases} M\mu & \text{if } 0 \leq M < N \\ N\mu & \text{if } M \geq N \end{cases} \tag{4.11}$$

This case is equivalent to a network of queues made of a single station with N servers.

- **<u>If $0 \leq M < N$:</u>**

For the single station we obtain:

$$\begin{aligned}
Q_1(M) &= M \\
X_1(M) &= M\mu \\
R_1(M) &= \frac{1}{\mu}
\end{aligned} \tag{4.12}$$

Considering the entire system we obtain:

$$\begin{aligned}
&Q(M) = Q_1(M) = M \\
&X(M) = X_1(M) = M\mu \\
&R(M) = R_1(M) = \frac{1}{\mu} \\
&\text{System aggregated throughput} = X(M) \times L = LM\mu \\
&\text{Mean wait time per user to access a channel} = R(M) = \frac{1}{\mu}
\end{aligned} \tag{4.13}$$

- **If $M \geq N$:**
  For the single station we obtain:

$$\begin{aligned}
Q_1(M) &= M \\
X_1(M) &= N\mu \\
R_1(M) &= \frac{M}{N\mu}
\end{aligned} \tag{4.14}$$

Considering the entire system we obtain:

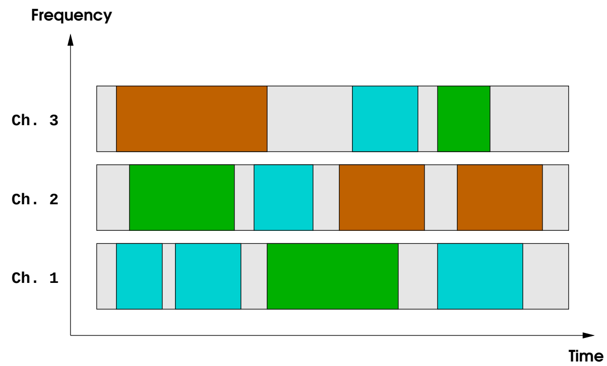
$$\begin{aligned}
&Q(M) = Q_1(M) = M \\
&X(M) = X_1(M) = N\mu \\
&R(M) = R_1(M) = \frac{M}{N\mu} \\
&\text{System aggregated throughput} = X(M) \times L = LN\mu \\
&\text{Mean wait time per user to access a channel} = R(M) = \frac{M}{N\mu}
\end{aligned} \tag{4.15}$$

Accordingly we obtain the following results for the entire system:

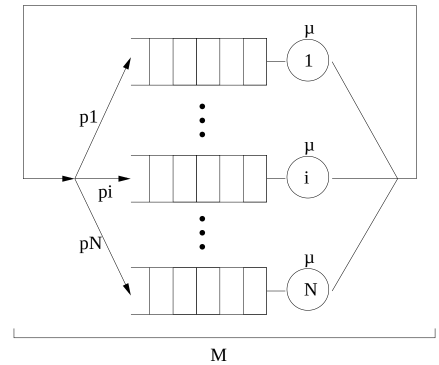
$$\boxed{\begin{aligned}
&Q(M) = M \\
&X(M) = \mu(M) \\
&R(M) = \frac{M}{\mu(M)} \\
&\text{System aggregated throughput} = X(M) \times L = L\mu(M) \\
&\text{Mean wait time per user to access a channel} = R(M) = \frac{M}{\mu(M)}
\end{aligned}} \tag{4.16}$$

4.4.8 Method f)

The system is composed of N channels that can be accessed by any user. A user randomly chooses a channel and then enters the corresponding FIFO waiting queue. The user then waits



Scenario with
3 Channels and 3 Users



Queue Model with
M Channels and N Users

Figure 4.7: Method f)

for this channel to become idle before being able to transmit. It is a probabilistic routing of users to channels. Figure 4.7 shows the scenario and queue model for this method.

In this system we have: $e_i = 1$, $\forall i = 1, ..., N$.

For a given server i we have:

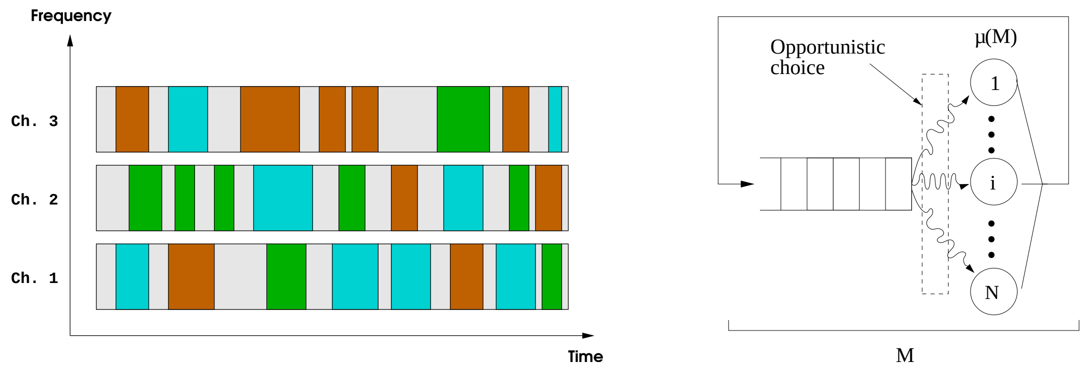
$$\begin{aligned} Q_i(M) &= \frac{M}{N} \\ X_i(M) &= \frac{\mu M}{N+M-1} \\ R_i(M) &= \frac{N+M-1}{\mu N} \end{aligned} \tag{4.17}$$

Considering the entire system (composed of the N servers) we obtain:

$$\boxed{\begin{aligned} &Q(M) = \sum_{i=1}^{N} Q_i(M) = M \\ &X(M) = \sum_{i=1}^{N} X_i(M) = \frac{\mu M N}{N+M-1} \\ &R(M) = \frac{1}{N}\sum_{i=1}^{N} R_i(M) = \frac{N+M-1}{\mu N} \\ &\text{System aggregated throughput} = X(M) \times L = L\frac{\mu M N}{N+M-1} \\ &\text{Mean wait time per user to access a channel} = R(M) = \frac{N+M-1}{\mu N} \end{aligned}} \tag{4.18}$$

4.4.9 Method g)

This method introduces some diversity (multi-channel diversity) by considering channel conditions (via transmit duration) in the channel choice: given a unique FIFO waiting queue, each pending Head-Of-Line (HOL) user chooses the idle channel that he sees with the best channel conditions, which corresponds to the smallest transmission duration for the same data payload to transmit. When several transmit modes are available, for a given data payload, transmitting a packet with a higher transmit mode leads to a smaller packet duration compared to using a smaller transmit mode. This method consists in an opportunistic routing of users to channels based on channel conditions. The larger the set of channels to choose from, the better the benefit from this method and thus the better the throughput. Figure 4.8 shows the scenario and queue model for this method.



Scenario with
3 Channels and 3 Users

Queue Model with
M Channels and N Users

Figure 4.8: Method g)

The performance of this method is studied by simulation in section E.4.4.

4.4.10 Comparison of Methods

The analytical expressions for methods a) to f) have been validated by simulations, in terms of system aggregated goodput and mean wait time. For each method the analytical expressions from a) to f) very well reflect the system behavior as shown by the very good match observed during the validation phase, between simulations and analytical expressions. The number of available channels is varied N from 1 to 20 by step of 1 (except for method b) starting at $N = 2$) while the number of users is kept constant ($M = 10$). The simulated time is long enough and equals 10 seconds.

The other parameters used are:

- For methods a), d), e) and f): $\mu = \frac{1}{2361.5}$ user$\cdot(s)^{-1}$ (corresponding to a mean service duration of: $\frac{1}{\mu} = T_{DIFS}+\sigma\frac{CW_{min}}{2}+T_{RTS}+T_{SIFS}+T_{CTS}+T_{SIFS}+T_{DATA}+T_{SIFS}+T_{ACK} = 2361.5\ s$, all packets (control and data) transmitted in mode 1 and a fixed payload of $L = 1500$ bytes), but the resulting use duration is random to model various channel conditions and transmit modes. All the T_x previous values are taken from the IEEE 802.11a standard [2] as a reference.
- For method b): μ_1 accounts for the reservation phase while μ_2 accounts for the data transmission phase, all packets (control and data) transmitted in mode 1 and a fixed

payload of $L = 1500$ bytes), but the resulting use duration is random to model various channel conditions and transmit modes.

- For method c): $\beta\mu$ varies with the number of available channels, all packets (control and data) transmitted in mode 1 and a fixed payload of $L = 1500$ bytes), but the resulting use duration is random to model various channel conditions and transmit modes.

Using the same assumptions and conditions, we compare the different methods in terms of system aggregated throughput and mean wait time. The curves represented are obtained by implementing all the above analytical expressions (except for method g) which is simulated).

Figure E.3 represents the system aggregated throughput of all the methods (from a) to f)).

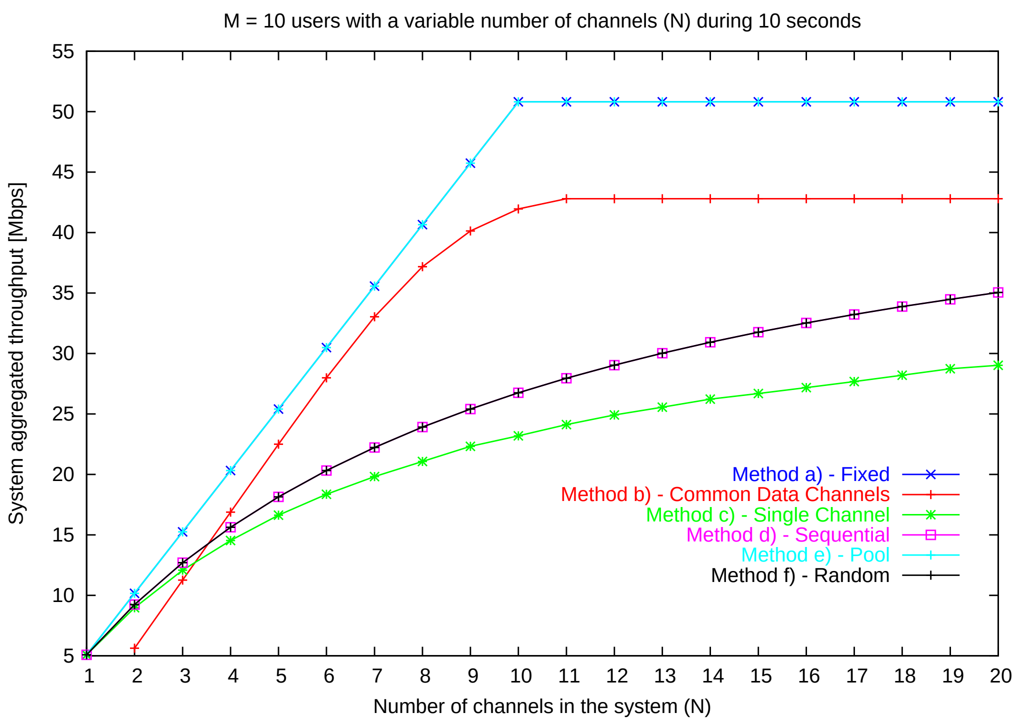


Figure 4.9: Comparison of Methods a) to f) in terms of System Aggregated Throughput [Mbps]

The main conclusions are:

1. The different methods do not all lead to the same throughput.
2. For a given number of users M, an increase in the number of channels lead for all the methods to an increase in throughput, but at a different rate per method.
3. The methods providing the best throughput, and giving the exact same results, are the fixed and the pool methods. The reason is that these methods best spread the users on the channels to leave the minimum number of idle channels at any given time when compared to the other methods. While there are less channels than users, the throughput increase

is linear. Then when there are more channels than users, there is no more increase in throughput as all the users are always assigned a channel for transmission.

4. The common data channel method is less effective than the pool method because it creates a bottleneck with the common channel, leading to one data channel less than the pool method.

5. For $N \geq 4$, the single channel method provides the smallest throughput because it does not parallelizes the wait time (the time during which each server is used not for payload transmission).

Figure 4.10 represents the mean wait time of all the methods (from a) to f)). The same reversed ordering hold true as for the throughput.

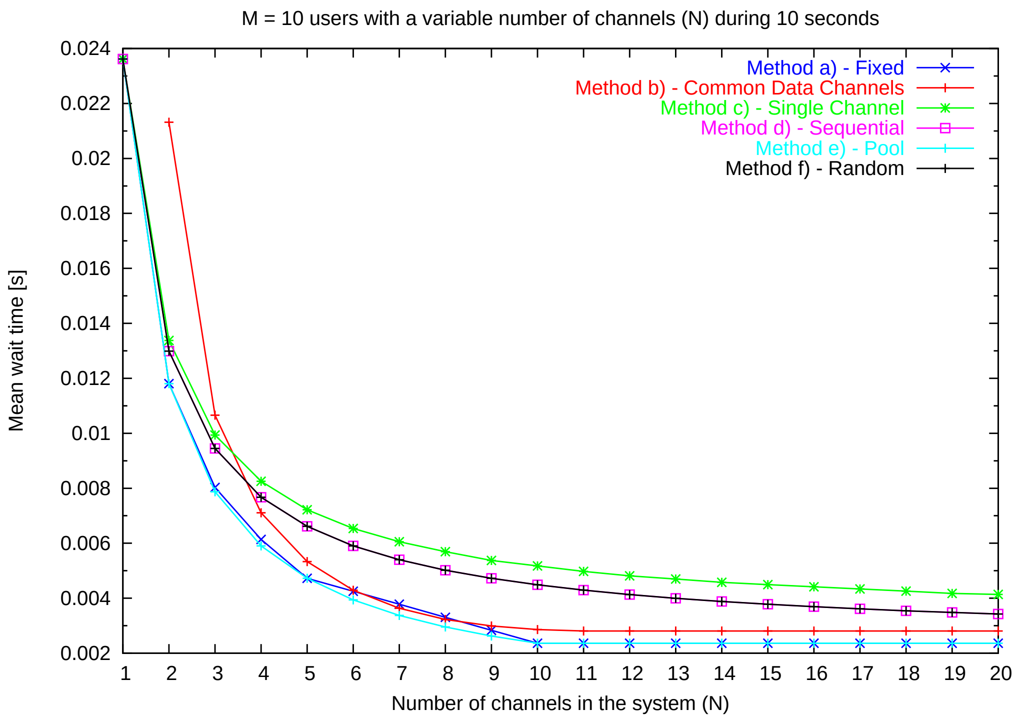


Figure 4.10: Comparison of Methods a) to f) in terms of Mean Wait Time [s]

Figure E.4 represents a comparison between method e) (the best method out of a) to f)) and method g) for the system aggregated throughput and the mean wait time. As seen in figure E.4, when HOL users using method g) have less than 1 channel to choose from (case where $N <= M$), the throughput is the same as the pool method. However, as soon as $N > M$, the throughput from method g) increases linearly with N whereas it remains constant for method e). Also, when using our assumptions, method g) provides G times more system throughput than the pool method ($N > M$), with G expressed as: $G = max\{0; N - M\} + 1$. In fact when using method g), the rate becomes M times greater when $N > M$ than the rate when $N \leq M$. The increase is linear because each point is considered as an average from the same law and channels are independent. The reason comes from the opportunistic choice that each user can do before using a new channel: he chooses to use the idle channel with the best channel conditions.

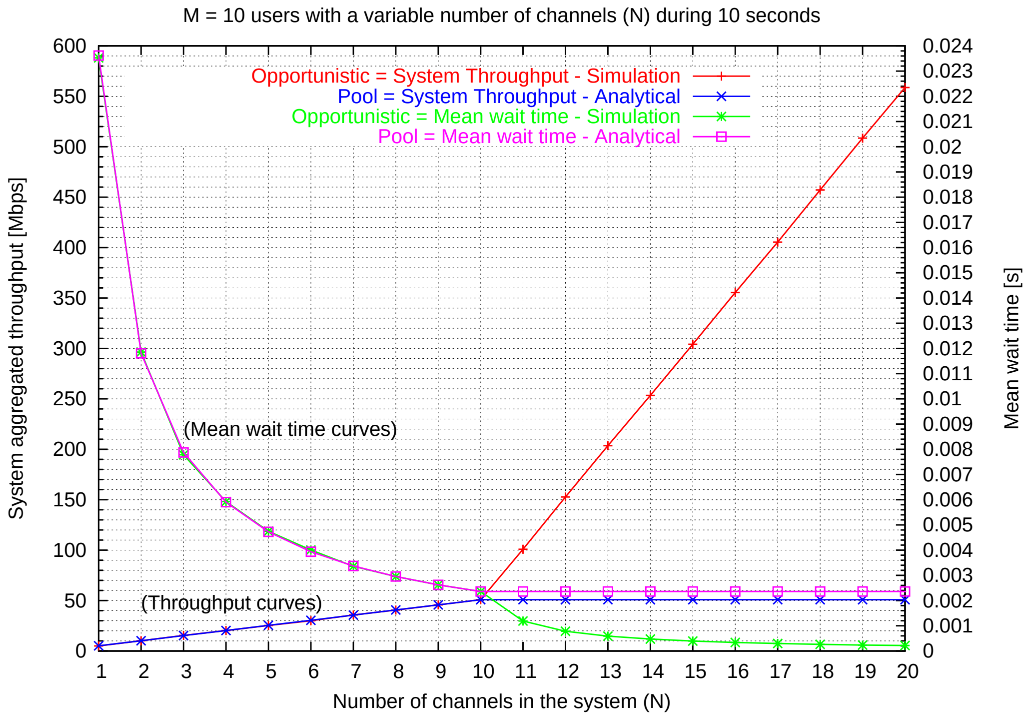


Figure 4.11: Comparison of Methods e) and g) in terms of System Aggregated Throughput [Mbps] and Mean Wait Time [s]

4.5 Conclusion

As a conclusion, considering all the studied methods (based on a FIFO queuing policy), the method leading to the highest system throughput is provided by method g). Indeed, the multi-channel diversity (opportunistic channel choice compared to a systematic or a random channel choice) introduces some gain in the system throughput. Thus, controlling the channel access at the user level as a function of the channel conditions increases the throughput for the whole system. The benefit comes from a dynamic and negotiated ordering of the users access to idle channels leading to the best possible system throughput. An even better method, not presented in this chapter (because it uses a non-FIFO queuing policy) but presented in chapter 5, would provide an even greater gain in system throughput by introducing multi-user multi-channel diversity. However, in a more realistic distributed context, collisions in channel access cannot be avoided. Inter-user communication is necessary for users to coordinate and negotiate their spectrum access. However, the communication overhead should be minimized not to reduce too much the system throughput.

Chapter 5

Framework for Solving the Spectrum Allocation Optimization Problem

5.1 List of Notations for this Chapter

Table E.2 presents the notations used in this chapter.

Notation	Description
a_{ij}	Binary (0 or 1) value indicating whether or not task i has selected resource j
f_{ij}	Force value (resulting benefit) obtained from the use of resource j by task i
i	Index used to refer to tasks, especially task i
j	Index used to refer to resources, especially resource j
n_i	Number of allocated resources to task i
n_{max}	Maximum allowed number of simultaneously used resources per task
n_{min}	Minimum allowed number of simultaneously used resources per task
M	Number of tasks
N	Number of available resources to allocate
R_{NA}	Number of idle resources as a result of the problem constraints
R_A	Number of resources that will be allocated due to the problem constraints
$S(N,i)$	Number of partitions of the ensemble N into i groups
$S(N;k_1,k_2,...,k_N)$	Number of partitions of N in k_i blocks of size i
T_A	Number of groups of resources that could be allocated

Table 5.1: List of Notations for this Chapter

Note that these notations are further explained in the remaining of this chapter.

5.2 Introduction

The past two decades have witnessed tremendous research activities in optimization methods with applications to various areas of academic science and industry. Optimization problems, either of practical or of theoretical importance, are made up of three basic ingredients:

- An **objective function** which we want to minimize or maximize. The idea consisting in choosing a "best" configuration or set of parameters to achieve some goal. For instance, various sectors in the industry are facing the challenge of minimizing waste, maximizing throughput or profit, optimizing the use of resources, etc.
- A set of unknowns or **variables** which affect the value of the objective function. For example, in a manufacturing problem, the variables might include the amounts of different resources used or the time spent on each activity.
- A set of **constraints**, restricting the unknowns to take on certain values but exclude others. For example, in a task scheduling problem, we could be interested in not exceeding a given waiting delay before tasks are scheduled.

Accordingly, an optimization problem can be reformulated as: find the values of the variables that minimize or maximize the objective function while satisfying the constraints.

In this thesis, we are interested in a constraint-based combinatorial (discrete variables and discrete solution) optimization problem, in a wireless context with the objective to find the best allocation of spectrum resources to users. Discrete optimization problems are ubiquitous in industrial processing and manufacturing. Such problems are computationally hard, and currently no efficient universal method exists to solve them. The most common approaches that have been proposed for solving discrete optimization problems are:

- Integer Programming Methods: they are quite powerful, but require that the problem be formulated as a maximization problem of a linear objective function over a set of linear inequalities. This is unfortunately not possible when the problem at hand contains disjunctions, as graph coloring problems and many scheduling problems do.
- Heuristic (or Meta-Heuristic) Search Techniques: some examples include: simulated annealing, tabu search, genetic algorithms, and neural networks. These methods can solve large problems with very good results, but often lack precise predictable behavior and a tuning of their parameters is required to reach the best values for the problem at hand.
- Global Search Techniques: they guarantee optimality but mostly for small size optimization problems. However, constraint-based search techniques provide means of aggressively pruning the search space, and can make global search a feasible alternative for large problems.

We are looking for the global optimum of a spectrum resource allocation optimization problem. In general terms, we assume a set of resources and tasks to be mapped on the resources in order to be executed.

In the next sections we first describe the wireless context of the problem, then formalize the problem and look at its complexity. Some possible methods for solving this problem, as well as the chosen approach will be presented before concluding the chapter.

5.3 Description of the Optimization Problem

5.3.1 Context

In the considered resource allocation optimization problem, tasks have different efficiency (force) when using resources. We define a "force" f_{ij} as the resulting benefit obtained from the use of resource j by task i. Assuming no overlap in data Tx: a resource can only be used by a single task at a time, thus the considered resource is limited.

Assuming a minimum resource unit, the smaller the resource group size to be assigned to each task without contiguity constraint, the better the opportunity to allocate the tasks to the resources that would value them the most, thus the best the final result. As a consequence, having no constraint of contiguous resources allocated to a task might increase the overall system sum force.

This is best explained through an example presented in table 5.2 with 2 resources and 2 tasks. Table 5.2 indicates the force of each task when using each resource separately. We assume a linear summation of the forces for a given task over several resources, and between all allocated tasks.

Table 5.2: Task and Resource Allocation Example

	Resource 1	**Resource 2**
Task 1	4	3
Task 2	5	1

As indicated in table 5.2, allocating the 2 resources to the same task leads to a total sum force of 7 (task 1) or 6 (task 2). This is lower than finding the best combination for the system: task 1 using resource 2 and task 2 using resource 1, leading to a total sum force of 8. In other words, there is an interest for the system in opportunistically assigning the best tasks to each resource, which is not necessarily the same as maximizing a given task's total force. In a spectrum allocation context, this can be interpreted as taking into account channel conditions (resp. the force) and assign the spectrum (resp. the resources) to the users (resp. the tasks) having the best channel conditions.

The next chapter further details the wireless context in which this optimization problem finds an application. The optimization problem will be formalized in general terms, as trying to map tasks on resources in order to reach the best final outcome for the whole system.

5.3.2 Mathematical Formulation of the Optimization Problem

The considered optimization problem is a resource allocation problem. Given N available resources to allocate among M tasks, each task i having a force value f_{ij} on resource j, we are looking for the best allocation of tasks on resources in order to maximize the total summation of the forces for the whole allocated system. Let us define a_{ij} as the value indicating whether or not task i has selected resource j, with $a_{ij} = \{0; 1\}$. If we define n_i as the number of allocated resources to task i, an "allocated task" has $n_i \neq 0$.

Using all these notations, the optimization problem objective can be formulated in mathematical terms as follows:

$$\text{Maximize} \sum_i \sum_j a_{ij} f_{ij}$$
$$\text{s.t.:} \begin{cases} (1) & \{n_{min}, n_{max}\} \in [1; N] \\ (2) & \forall i,\ n_i = 0 \Leftrightarrow \sum_j a_{ij} = 0 \\ (3) & \forall i,\ n_i \neq 0 \Leftrightarrow \\ & n_{min} \leq \sum_j a_{ij} \leq n_{max} \\ (4) & \forall j,\ \sum_i a_{ij} \leq 1 \end{cases} \tag{5.1}$$

The above constraints can be interpreted as follows:

- Constraint (1): the limits in number of resources allocated per task are bounded by 1 and the total number of available resources N.
- Constraints (2) and (3): n_i, the number of resources allocated per task is either 0 when the task is not taking part in the resource allocation, or between n_{min} and n_{max} resources.
- Constraint (4): no two tasks will be allocated the same resource at the same time.

In other words, the objective is to find the final system allocation vector maximizing the system sum force, assuming all the constraints are respected. In our problem, all the contending tasks simultaneously negotiate for their resources, and then all the winners simultaneously use the won resources.

5.3.3 Complexity of the Optimization Problem

This section presents the complexity of the considered optimization problem. First, let us give a small example of the optimization problem to solve. We assume $N = 5$ resources, and $M = 3$ tasks. In this example, resources will be indicated by letters: A, B, C, D and E, whereas tasks will be indicated by numbers: 1, 2, 3. We assume also $n_{min} = 2$ and $n_{max} = 3$, meaning that a task requires to use simultaneously at least 2 resources, but cannot use simultaneously more than 3 resources. Using this small example, the problem search space in which to find the best total allocation of tasks on resources is described following these steps:

1. Find the number of **possible decompositions of N in groups** of a given size (between n_{min} and n_{max}). There are only 2 possibilities:
   - Case a): 1 group of size 3 and 1 group of size 2.
   - Case b): 2 groups of size 2 and a rest of 1 group of size 1.
2. Find all the achievable **partitions** with the above decomposition (with no notion of order):
   - Using the grouping of the resources proposed in Case a), 1 task would be allocated 3 resources and another one would be allocated 2 resources, whereas a task would remain unallocated. There are 10 possible partitions: {ABC,DE}, {ABD/CE}, {ACD/BE}, {ACE/BD}, {ABE/CD}, {ADE/BC}, {BCD/AE}, {BCE/AD}, {BDE/AC}, {CDE/AB}.
   - Using the grouping of the resources proposed in Case b), only 2 tasks would each be allocated 2 resources, whereas 1 resource would remain idle and a task would remain unallocated. There are 15 possible partitions: {AB/CD/E}, {AB/CE/D}, {AB/DE/C}, {AC/BD/E}, {AC/BE/D}, {AC/DE/B}, {AD/BC/E}, {AD/BE/C}, {AD/CE/B}, {AE/BC/D}, {AE/BD/C}, {AE/CD/B}, {BC/DE/A}, {BD/CE/A}, {BE/CD/A}.
3. For each Case a) and b), find the total **number of ways of assigning the usable groups** of resources to the tasks.
   - For Case a), the number of ways of choosing 2 tasks out of 3 with a notion of order, i.e. $A_3^2 = 6$.
   - For Case b), the number of ways of choosing 2 tasks out of 3 with a notion of order, i.e. $A_3^2 = 6$.
4. The **total number of possibilities** for all cases put together is: $10 * 6 + 15 * 6 = 150$.

Accordingly, for this simple example of only 5 resources and 3 tasks, finding the best combination of tasks on resources to maximize the total sum of forces subject to the constraints requires to extract one solution out of a total of 150 possible solutions. Next we show that the search space size exponentially increases with N and M, n_{min} and n_{max}.

We calculate the complexity of the optimization problem as a function of the system parameters N, M, n_{min} and n_{max}. The **first task** is to calculate the number of possible decompositions of N in groups of size between n_{min} and n_{max}.

This is achieved by successive divisions as shown in figure 5.1.

For each integer value within $[n_{min}; n_{max}]$, we calculate the value $Q_i = floor\left(\frac{N}{i}\right)$, which requires $n_{max} - n_{min} + 1$ divisions.

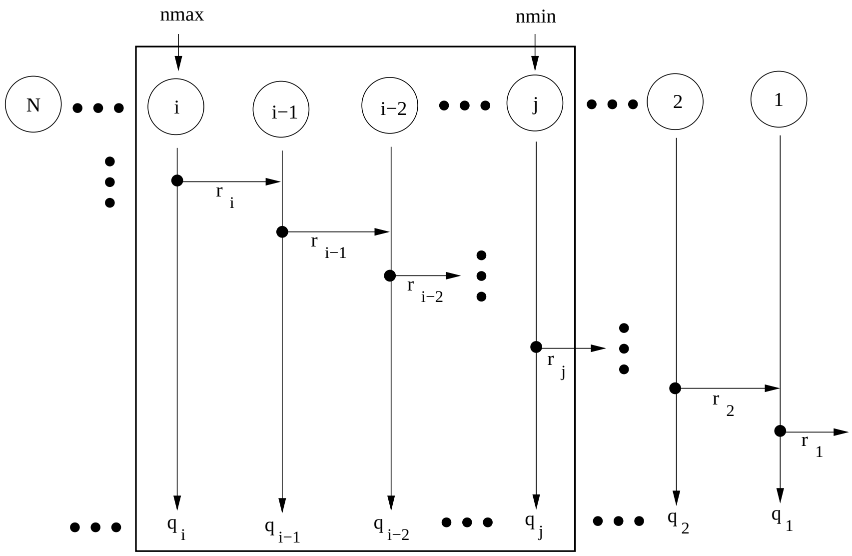


Figure 5.1: Decomposition of N into Groups of Integers

Then, we write :

$$\begin{cases} N & = q_{n_{max}} & \times n_{max} & + r_{n_{max}} \\ r_{n_{max}} & = q_{n_{max}-1} & \times (n_{max}-1) & + r_{n_{max}-1} \\ \cdots & = \cdots & & \\ r_{i+1} & = q_i & \times i & + r_i \\ r_i & = q_{i-1} & \times (i-1) & + r_{i-1} \\ \cdots & = \cdots & & \\ r_{n_{min}+2} & = q_{n_{min}+1} & \times (n_{min}+1) & + r_{n_{min}+1} \\ r_{n_{min}+1} & = q_{n_{min}} & \times (n_{min}) & + r_{n_{min}} \end{cases} \tag{5.2}$$

Each time, the division rest from the previous upper integer is propagated at the lower level until the n_{nmin} level, where the final rest $r_{n_{min}}$ is collected. As a result, all the resources will be allocated ($r_{n_{min}} = 0$) or not ($r_{n_{min}} \neq 0$).

Equations 5.2 will be written: $\forall i, q_i = Q_i \rightarrow 0$. The number of divisions performed to decompose N will be: $\prod_{i=n_{min}}^{n_{max}} (Q_i + 1) - Q_{n_{min}}$, which is also the final number of possible decompositions of N into groups of constrained size. A decomposition of N can be defined as finding the number of groups of a defined size whose sum sizes equal N.

The **second task**, to achieve for each decomposition of N is to find the number of possible partitions of the ensemble with N elements (collection of disjoint blocks whose union forms the ensemble). For example for $X = 1, 2, 3$, the possible partitions are: $\{1/2/3\}$, $\{1/2\}\{3\}$, $\{1/3\}\{2\}$, $\{1\}\{2/3\}$ and $\{1\}\{2\}\{3\}$. The order has no importance for the partitions.

Let us find $S(N; k_1, k_2, ..., k_N)$ the number of partitions of N in k_i blocks of size i, with $1 \leq i \leq N$ and $\sum_{i=1}^{N} ik_i = N$. Note that if $r_{min} \neq 0$ (with $1 \leq r_{min} \leq N$), it creates a single group of size $k_{r_{min}}$ because we cannot have at the end of the decomposition, more than one group with a size comprised between n_{max} and n_{min}.

We have:

$$S(N; k_1, k_2, ..., k_N) = \frac{N!}{k_1!k_2!...k_N!(1!)^{(k_1)}(2!)^{(k_2)}...(N!)^{(k_N)}} \tag{5.3}$$

For each decomposition of N we have:

- The number R_A of resources that will be allocated is: $R_A = \sum_{i=1, i \neq r_{min}}^{N} ik_i$, with $R_A \leq N$.
- The number R_{NA} of idle resources would be: $R_{NA} = N - R_A$.
- The number T_A of groups of resource that could be allocated is: $T_A = \sum_{i=1, i \neq r_{min}}^{N} k_i$ with $T_A \leq N$.

Note that the maximum number of idle resources is: $Max(0; N - n_{max} \times Min(M; floor\left(\frac{N}{n_{max}}\right)))$. **Finally**, for each partition out of $S(N; k_1, k_2, ..., k_N)$, the number of possible ways to assign the groups of resources to each task is given by the formula: $A_{max(T_A,M)}^{min(T_A,M)}$. If $M \leq T_A$, it corresponds to the number of ways to put M tasks into T_A resources with a maximum of one task per resource group. If $M \geq T_A$, it corresponds to the number of ways of taking T_A tasks out of M with a notion of order.

When putting together all the previous steps it is easy to see that this optimization problem is NP-complete. If we take the example of $n_{min} = 1$ and $n_{max} = N$, then all possible group sizes are allowed from 1 to N, leading to all possible partitions of the ensemble N. In this case, the number of partitions of N is given by the Stirling number of second kind $S(N) = \sum_{i=1}^{N} S(N, i)$, where $S(N, i)$ represents the number of partitions of the ensemble into i groups and is calculated using an iterative formula. However, a simple asymptotic approximation of $S(N)$ was proposed in 1918 by Ramanujan and Hardy such that:

$$S(N) \approx \frac{e^{\pi\sqrt{\frac{2N}{3}}}}{4N\sqrt{3}} \tag{5.4}$$

For example, for $N = 10$ and $M = 10$, the search space size in order to find the best allocation would be: $\sum_{i=1}^{N} S(N, i) A_M^i = \approx 10^{10}$.

For the following figures, we use the parameters: $N = 10$ and $M = 10$ and we vary n_{min} and n_{max} from 1 to N. Figure 5.2 represents the total search space size. As seen in figure 5.2, the considered optimization problem is NP-complete, and is very dependent on the values of n_{min} and n_{max}.

Figure 5.3 represents the proportion of total resources allocated. As seen in figure 5.3, some combinations of n_{min} and n_{max} values introduce a scarcity in the number of possible allocated resources. Especially, in this example, only 60% of the resources are allocated in the worst case, when $n_{min} = n_{max} = 6$. Accordingly, some combined values of n_{min} and n_{max}, constraint the system in leaving much more idle resources than in the other combinations.

As a consequence, several conclusions are drawn from this complexity study:

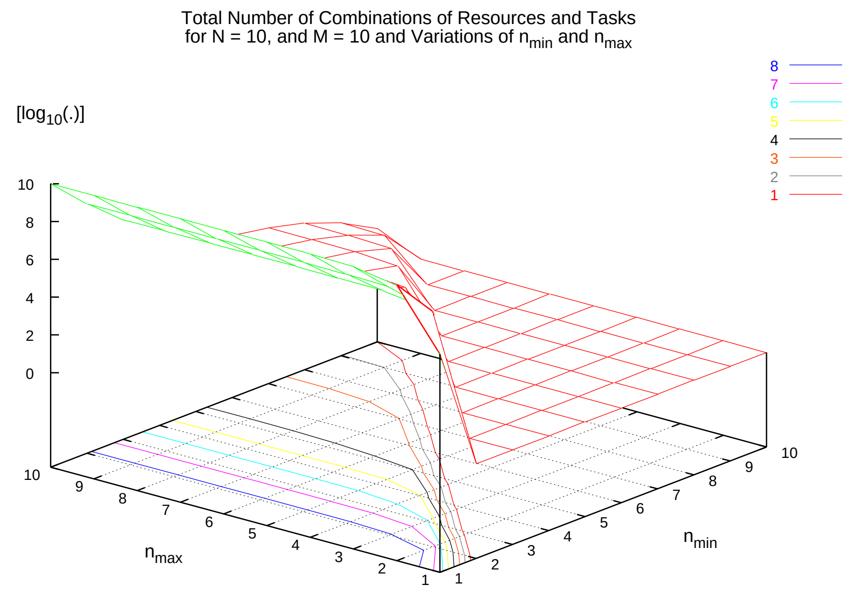


Figure 5.2: Total Search Space Size

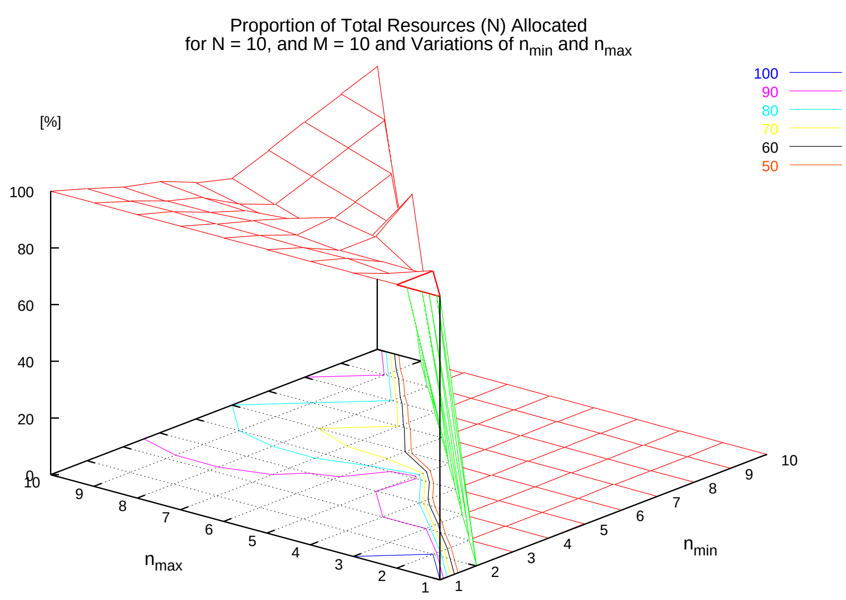


Figure 5.3: Proportion of Allocated Resources

- The current resource optimization problem with its constraints, is NP-complete and requires the use of an heuristic or meta-heuristic method in order to:

  1. Guide the search within the large search space.
  2. Reach a good solution in a "reasonable" amount of time.
  3. Allow to trade between quality of the solution and convergence speed.

- The n_{min} and n_{max} constraints have an impact on the search space size, as well as on the final achievable system sum force.

Note that the optimization problem's constraints do not specify that the final grouping of resources shall be contiguous, indeed it can be fragmented. Adding the constraint of contiguous groups of resources would greatly reduce the problem size, but also the achievable performances.

5.3.4 Optimal Solution at a Reasonable Cost

For some optimization problems involving real systems, the search for the global optimum is not possible, and even not desirable. For large and NP-complete optimization problems, the search for an optimal solution can be long and costly.

Sometimes, the input data include errors, or the current system model is not accurate enough, or the input data hold true only for a short period of time. In this context, searching for the optimal solution is not desirable because it would not correspond to the necessary real-life optimum.

Accordingly, for real-life applications, the major question is: "how good should the solution be?". In other words, considering all the many parameters to consider in an optimization problem: quality of the final solution, optimization cost, speed of convergence, etc, is the quality of the final solution the *only important* parameter? In fact, what is usually expected at the end is a quick acceptable solution.

For example, for some applications a quick convergence would prevail on the quality of the final achievable solution. In emergency situations (accidents, disaster relief, etc), the main constraint might rather be to "find in 1 hour the best solution to the problem", instead of "find the best achievable solution for this problem with no time constraint".

Depending on whether we are facing a static (fixed input data) or dynamic (variable input data) optimization problem, different types of algorithms should be used. In dynamic environments, robust and fast algorithms are required instead of fine and slow algorithms. Especially, the final solution should be found before an important change occurs in the input data.

What if the input data variations cannot be precisely known or predicted? To cope with such imperfect information, robust optimization algorithms are required, which should work under several (or any) kind of environment variations. We insist on the simplicity (cheap and fast) and the efficiency of the algorithms for applications to a real-world wireless context. Speed of convergence is important, especially if optimization tasks have to be regularly performed.

Owing to variations in channel conditions, number of contending users, available spectrum resource, etc, our spectrum resource allocation problem is dynamic. As a consequence, given the dynamicity of the optimization problem at hand, as shown in the following chapters, we try to develop an algorithm able to tradeoff, on-demand, between the speed of convergence and the quality of the final solution.

5.4 Possible Methods to Solve the Problem

Section 5.3.3 has shown that the currently studied problem is a combinatorial constraints satisfaction NP-complete optimization problem. Heuristics or meta-heuristics can be used to guide the search of a good solution within the large search space. Heuristics are computational methods using a trial and error approach to find an acceptable solution in a reasonable (polynomial) time, for computationally difficult problems. Metaheuristic methods are high-level heuristics that can be specialized to solve different types of optimization problems.

Several metaheuristics are inspired by natural systems such as:

- Simulated Annealing: inspired by physics. It uses a path approach (a single solution is explored at a time) to move through the search space.
- Genetic Algorithms: inspired by biology and especially the evolution of genetic information through individuals and time. It uses a population approach (several solutions are evaluated in parallel) to move through the search space.
- Neural Networks: inspired by biology and especially brain connections. It was introduced by the artificial intelligence domain. It uses a population approach to move through the search space.
- Collective Intelligence: inspired by biology and especially ethology (science of individuals' behaviors in nature). Several methods already exist such as: Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO). It uses a population approach to move through the search space.

Many other highly efficient methods or theories can be used in optimization such as:

- Graph Algorithms: proposes a symbolic representation (a graph) of a network and of its connectivity, and algorithms have been proposed to find the best sub-graph relevant to the problem.
- Markov Decision Processes: proposes to find the best strategy to choose for the next step, given the actions (and their result) from the previous steps.
- Fuzzy Logic: proposes techniques for reasoning under uncertainty. It has been widely and successfully used to solve difficult industrial control problems. It is especially useful when the system model's behavior is either too complex to model and put into equations, or not attainable.

- Game Theory: studies situations where players adapt their strategy and choose different actions in an attempt to maximize their returns. Applications are found in many domains: economy, ethics, politics, optimization, etc. Different types of games and assumptions on the players are used to model and solve various types of problems. This approach is currently becoming increasingly studied by the wireless community.

Note that each method has its own sets of advantages and drawbacks, and is better suited to solve a given category of problems. Many methods already exist and others will be developed in the future in order to improve the currently known solutions. Sometimes, combining the advantages of several approaches within the same algorithm lead to better results.

An application of the spectrum resource allocation optimization problem is to create an adaptive frequency planning in a network. In this thesis, we have chosen to contribute to the development of the collective intelligence metaheuristic because of its bottom-up approach of system modeling. This approach belongs to the domain of ethology and artificial intelligence. It uses a population approach, where several solutions are simultaneously tested in parallel, thus greatly increasing the convergence towards a good solution. In the next chapter, it will be shown that this method can easily trade between speed of convergence and quality of the solution.

Usually, the following steps are iteratively followed when using a metaheuristic:

1. Diversification (exploration): new solutions than the best currently known are explored by some individuals, in order to discover a better solution to the problem.
2. Intensification (exploitation): some individuals will maintain the best currently known solutions.
3. Memory and learning: memory refers to remembering good solutions and forgetting bad solutions. Thus, the algorithm learns about the problem and becomes faster in finding out the best solutions than at start time.

The following section will justify, for future generations of wireless systems, our metaheuristic choice, and better explain its many interesting properties.

5.5 Opportunistic Spectrum Allocation with Collective Intelligence

5.5.1 Interests of Distributed Optimization

Historically, wireless networks have long been designed using centralized, hierarchical command and control techniques. However, this is no more possible when dealing with systems composed of thousands or even millions of dynamically changing, communicating, heterogeneous entities. In addition, the design of current wireless networks will have to change in the future, considering (1) the forecasted increase in wireless subscribers, (2) the increasing volume of data exchanged per subscriber, (3) users expecting seamless heterogeneous wireless access anywhere and at anytime.

Future communication networks should be able: to scale to billions of users, to adapt to unexpected diverse and dynamic conditions in the network, to require minimal human configuration and possess a decentralized control, to self-configure and self-organize.

However, distributed systems [36] providing communication capabilities between its nodes already exist in our societies and are rapidly being developed. Examples include the Internet, telecommunication networks, ad-hoc networks using 802.11 technology, etc.

The progresses in multi-agent modeling will provide insights in several real-life problems involving several individuals distributed in the same environment. An advantage is that individuals will have their autonomy and their own set of rules, and will no more be represented as mean individuals but as several identified interacting entities.

Throughout nature, an enormous amount of processing is taking place at the level of the individual organisms (e.g.: ants, wasps, humans, etc). Such individuals are autonomous nodes acting according to some rules (from simple to very complex), with imperfect knowledge of their environment (local knowledge), having memory, and able to interact with other similar entities. Thus, the society composed of all these individuals can be seen as a complex super-organism. This is the way wireless networks should be organized in the future.

Fortunately, tremendous technological progresses have been achieved in hardware and software, resulting in devices having high processing capabilities and memory size. However, as shown in the next section, technology alone will not provide a better understanding of the behavior of complex adaptive distributed systems.

5.5.2 Challenges of Distributed Optimization

Even though the distributed topology sounds very attractive, still the gap between natural and man-made systems is huge, and we have a long way to go before we will (1) fully understand (conceptual challenges) the underlying principles of distributed systems in nature, and (2) be able to implement (technological challenges) small size devices acting as an autonomous network of cognitive nodes.

Current distributed systems are not autonomous, they still require some tuning and suffer from coordination overhead because control communication is required to provide data communication reliably. As an example, the Internet traffic optimization considering all the heterogeneous requirements coming from all types of applications, considering all types of connections, etc, is an highly complex optimization problem. As a result, the wireless research community widely agrees that future wireless networks will benefit from a deeper understanding of the fundamental properties of complex adaptive and distributed systems, as demonstrated by an increasing interest in distributed systems.

5.5.2.1 Cultural Change

The main challenge to overcome is to undergo a cultural change: going from a top-down approach (command and control) to a bottom-up approach. Human beings suffer from a "centralized mindset"; they would like to assign the coordination of activities to a central command. But the way social insects form highways and other amazing structures (bridges, chains, nests, etc) and can perform complex tasks (nest building, defense, cleaning, brood care, foraging, etc) is very different: they self-organize through direct and indirect interactions.

A common feature of these distributed systems is that organized behavior emerges from the interactions of many simple parts. Dissimilar systems (businesses, ant colonies, and brains) share fundamental commonalities: individual cells interact to form differentiated body parts, ants interact to form colonies, neurons interact to form intelligent systems, and people interact to form social networks. Properties shared by many complex systems are emergent behavior, self-organization, adaptation, the development of specialized parts, patterns of cooperation and competition within social groups, decentralized control, social networks, and the development of stable and globally optimal structure.

Today, the bottom-up approach is attracting much importance and will continue in the future, because, even if fundamentally different from the centralized approach, it is much relevant to many real-life situations where several distributed entities interact using a limited knowledge of the environment. Examples include: Internet, network of employees in a large international company, etc. In fact, the challenges faced by current and future wired and wireless networks have already been overcome in large scale biological systems. Future wireless networks will benefit from adopting key biological principles and mechanisms.

The underlying modeling challenge of distributed systems is called the sub-optimization problem: optimizing the outcome for a subsystem will in general not optimize the outcome for the system as a whole. This intrinsic difficulty may degenerate into the "tragedy of the commons" [5]: corresponding to the exhaustion of shared resources because of competition between the subsystems. The central question for resource sharing is: "which rules should be locally applied by all interacting agents, in order to globally obtain the expected behavior?".

When trying to globally optimize the outcome for a system consisting of distinct subsystems (e.g. maximizing the amount of prey hunted for a pack of wolves, or minimizing the total punishment for the system consisting of the two prisoners in the Prisoners' Dilemma game, etc), you might try to do this by optimizing the result for each of the subsystems separately. This is called "sub-optimization". The principle of sub-optimization states that sub-optimization in general does not lead to global optimization. Indeed, the optimization for each of the wolves separately is to let the others do the hunting, and then come to eat from their captures. Yet, if all wolves would act like that, no prey would ever be captured and all wolves would starve.

Interactions between the different components of a system lead to the emergence property, summarized in the following principle: "the whole is more than the sum of its parts". In other words, the system as a whole is far more capable than each component taken separately, and than all the components put together. This is a major difference with a centralized system, where interaction between components is not required, resulting in no more than just the aggregation of all of its parts. Another way of expressing this aspect of "non-linearity" in cooperation is to say that the different entities are engaged in a non zero-sum game, that is, the sum of resources that can be gained is not constant, and depends on the specific interactions between the entities.

The micro-macro link (or system-individual link) is an instance of a general problem that arises across multi-agent systems. The optimal strategy [37] for the collective system made of all the individuals of a group is to cooperate, but the optimal strategy for each individual as a subsystem is not to cooperate (each individual being selfish). In general, global optimization is different from sub(system)-optimization. Evolution usually tries to optimize first the subsystem level, we need additional mechanisms to explain global optimization at the level of the cooperating system. We are faced with the individual's behavior and the several different social roles it can have in the system, depending on the environment.

As a consequence, a system architect's perspective (**top-down = looking/controlling the system from the outside**) is: how to control a system so that it behaves in a given way? On the contrary, the multi-agent perspective (**bottom-up = controlling the system from the inside**), tries to answer the opposite question: how to control agents in order to reach a given system behavior?

Thanks to the level of abstraction that mathematical modeling provides, one can build conceptual bridges between systems that operate in very different environments, at very different time and spatial scales, under very different constraints. As such, it has been possible to build artificial ants to help in packet routing, based on natural ants properties.

5.5.2.2 Group Communication

The second challenge to overcome is specific to wireless communication and consists in trying to develop a theory of group communication with novel capabilities at the PHY layer and at the MAC layer. As an example, how to build artificial pheromones with aggregation property using wireless devices?

5.5.2.3 Dynamic Adaptation

The third challenge to overcome is to design wireless devices able to dynamically adapt exactly as social insects. With the support of increasing computing capabilities it will be possible to build complex wireless agents leading to self-adapting wireless systems.

In this thesis, we try to dynamically optimize the spectrum resource allocation, and thus we need to answer the following question: how to best rule in a distributed fashion a resource utilization in a dynamic population? The next section describes some of the known interesting and powerful properties of the Swarm Intelligence (SI) metaheuristic.

5.5.3 Swarm Intelligence

To solve the spectrum resource allocation problem described in section 5.3, we will use a biologically inspired metaheuristic called SI, distributed in essence. It belongs to the domain of artificial intelligence and ethology (especially the study of social insects). SI is an artificial intelligence technique based around the study of collective behavior in decentralized and self-organized systems. The expression "swarm intelligence" was introduced by Beni and Wang in 1989, in the context of cellular robotic systems. SI is defined as: "the emergent collective intelligence of groups of simple agents" (Bonabeau and al, 1999). Computational intelligence structures at the system level emerge, from unstructured starting conditions, using powerful interactions mechanisms, while the action of each individual appears random (ants) when overlooked at the system level. Although there is normally no centralized control structure dictating how individual agents should behave, local interactions between such agents often lead to the emergence of a global behavior. Examples of such systems can be found in nature, including ant colonies, bird flocking, animal herding, bacteria molding and fish schooling. Figure 5.4 provides the example of ants building a temporary bridge to cross an obstacle.

Figure 5.4: Ants Building a Temporary Bridge to Cross an Obstacle

As seen in figure 5.4, the gap to cross is longer than any single ant. Thus, for each ant to be able to cross the bridge they need to coordinate their actions: some ants decide to create a temporary bridge out of their own body, while the others walk on this ant-bridge. This is a great example of the power of group coordination: a group of cooperating individuals is capable of creating emerging structures through joined efforts, thus giving the group capabilities that any single individual does not possess. In the following, we uniquely concentrate on the properties of social insects. Some examples of social insects achievements when using collective intelligence include: group foraging of social insects, division of labor, cooperative transportation, nest-building of social insects, collective sorting and clustering. For social insects, errors and randomness are not considered as "bugs"; rather, they contribute very strongly to their success by enabling them to discover and explore, in addition to exploiting. Self-organization feeds itself upon errors to provide the colony with flexibility (the colony adapts to a changing environment) and robustness (even when one or more individuals fail, the group can still perform its tasks). With self-organization, the behavior of the group is often unpredictable, emerging from the collective interactions of all of the individuals. The simple rules by which individuals interact can generate complex group behavior. Indeed, the emergence of such collective behavior out of simple rules is one the great lessons of the SI.

SI offers an excellent way of controlling complex distributed systems by defining and ensuring simple local rules at the lowest component level. A known drawback common to all heuristics and metaheuristics is the necessary initial tuning of the parameters to best adapt their numerical values to the problem at hand. However, an interest of the SI is that several parameters can be varied, thus providing a direct mean of controlling the algorithm, with the aim, for example, to trade between speed of convergence and quality of the final solution.

SI systems heavily rely on interactions (communication) between agents to coordinate their local actions:

- Directly: peer-to-peer communication. E.g.: physical contact between ants.
- Indirectly: through environment modifications. E.g.: leaving pheromone trails on the ground. In addition, each ant can adapt the concentration of pheromones put on the ground depending on the message to communicate (e.g.: quantify of food found on the current path).

Social insects live in self-organized colonies of insects. "Self-organization is a set of dynamical mechanisms whereby structures appear at the global level of a system from interactions of its lower-level components" (Bonabeau and al, in Swarm Intelligence, 1999). Self-organized systems are characterized by emerging structures, multistability (coexistence of many stable states), and state transitions with a dramatic change of the system behavior.

The four basic principles of self-organization are:

- Positive feedback (amplification by aggregation).
- Negative feedback (for counter-balance and stabilization).
- Amplification of fluctuations (randomness, errors, random walks).
- Multiple interactions.

The action of indirect agent interactions through the environment modifications is called "stigmergy", and consists in a reinforcement learning method. It is an highly powerful and simple tool for collective dynamic coordination, and serves as an external memory of all past actions (with a variable forgetting capability) enabling the entire colony to learn and forget. Each individual passing by uses the environment memory to decide on its next moves, and also contributes to change the system memory by itself modifying the environment. The environment keeps a distributed memory of exploration. Solution forgetting is also taken into account due to pheromone evaporation over time. As such, social insects are able to adapt to very dynamic environments, because any not maintained solution will be forgotten.

Using the previously described organization, social insects are unintentionally but efficiently able to solve complex tasks and optimization problems. For example, an ant looking for food does not necessarily tries to minimize its travel distance from the nest to a food source and then back to the nest. Instead, each ant relies on existing pheromone trails on the ground (the aggregation of its contributions and the others' contributions) to guide its food search as well as to find its way back to the nest. As a result, together they dynamically manage to solve the shortest path problem, as illustrated in figure 5.5.

Depending on the species, each individual in a colony of social insects is not necessarily specialized for a given task, but can perform different tasks according to the needs. This is particularly true for wasps which are not physically differentiated. In addition, a given individual can have several social roles over time. For example, individuals can be:

- Repeater: a "dumb" agent.
- Imitator: social copying.
- Comparer: social and rational.
- Deliberator: acts to maximize its own benefit.

Several factors (individual preferences, social drives, options, rationality, irrationality or random reason, etc) can modify the social role of an agent over time, as such there is no pretermined role in all situations, it is rather depending on environment state variations. Changes in an agent social role allows it to dynamically balance exploitation of new solutions and exploration of currently best solutions.

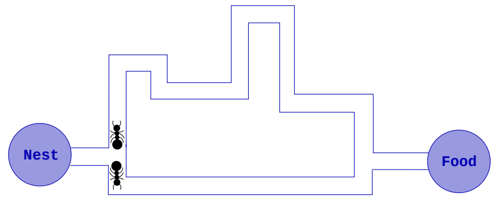
A colony of social insects has the following interesting properties while solving a problem:

- Scalability: a colony consists of a few tenth to several millions of individuals, and provides working examples of the swarm intelligence's efficiency under real-world dynamic conditions.
- Flexibility: the colony adapts to internal and external perturbations (environment changes).
- Robustness: tasks are completed even if some individuals fail or get lost or die. In addition, the amplification and solution improvement works without the knowledge of the amount of individuals in the colony.
- Decentralization: there is no central control(ler) in the colony to rule and supervise the colony actions.
- Self-organization: paths to solutions are emergent rather than predefined, and reform as environment changes, without any external help.

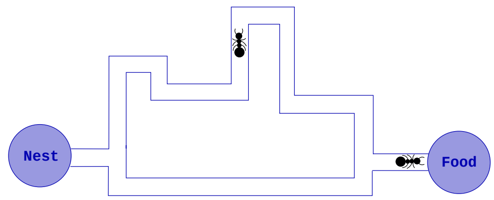
Note that no single social individual possesses all these characteristics, but they rather emerge from all the collective actions and interactions of all the individuals in the colony.

The SI metaheuristic is of particular interest for the wireless community because it possesses many of the required properties (see the list above) for next generations of wireless systems. A real strength of SI is that biology has proven working and reliable solutions to hard computational problems under real-world conditions.

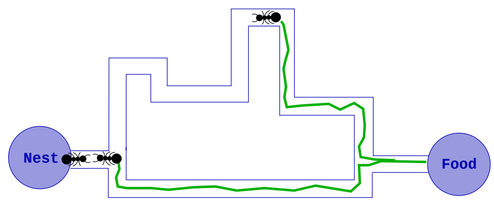
(1) Ants randomly choose a given direction



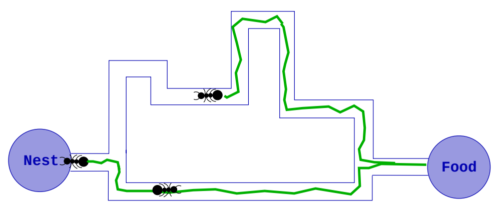
(2) The ants on the shortest path will reach the food source first



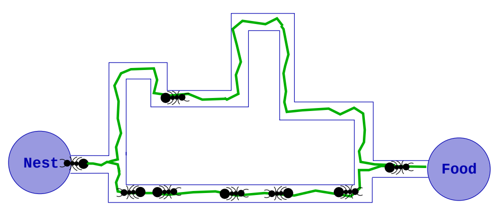
(3) The ants on the shortest path will be the first ones to reach the nest while carrying food taken from the food source



(4) Other ants in the nest will choose to follow the path with the highest pheromone concentration (the one at the bottom)



(5) More and more ants will choose the path at the bottom (the shortest path), resulting in an increased aggregated concentration of pheromones on this path compared to the other one



(6) As a result of this iterative process, all the ants finally use the path at the bottom. They have been able to select the shortest path, without any external control and help

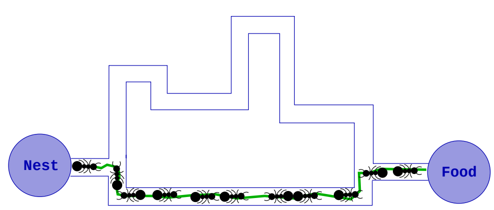


Figure 5.5: Ant Shortest Path Finding Example

For a wireless system designer willing to use the SI metaheuristic, the modeling challenge consists in:

1. Identifying analogies between distributed problem solving in a wireless context and social insects.
2. Creating mathematical models of artificial social insects (simplification, additional properties).
3. Building computer models of an individual with its rules and interaction capabilities.
4. Simulating multi-agents scenario and then tuning/refining the model parameters.
5. Ensuring that no interesting property is removed during the modeling process. Thus, artificial social agents use the same properties as their natural counterparts, augmented with some additional capabilities specific to the problem at hand.

One great result from SI is that even with a population of identical and simple agents (limited capabilities leading to cheap devices), it is still possible to create a very complex and dynamic behavior at the system level. Such a super-organism composed of social artificial agents would have a higher adaptation capability compared to a centralized approach where each of these agents are taught what to do. This structure is very promising for applications to future "intelligent" wireless communication systems. Indeed, cognitive radios are capable of sensing their environment, adapt to its variations, react and learn. This "intelligence" opens the way to more flexible regulatory rules, new ways of considering frequency reuse and avoiding interference.

In this thesis, we contribute to this global research effort using the SI metaheuristic. We model and use a multi-agent approach based on several key principles and mechanisms taken from social insects, to design dynamic spectrum allocation algorithms for wireless systems.

We found of particular interest the wasps capabilities of adaptive task allocation (dynamic division of labor) with a dynamic specialization of each individual to perform a given task, according to the changes in the task demand using pheromones. Our innovative approach is to apply such concepts in a context of dynamic spectrum allocation at the user level.

5.6 Conclusion

In this chapter, we have introduced and discussed the optimization problem we want to study. Out of the many other possible methods to solve this NP-complete optimization problem, the SI metaheuristic will be used. We have described its interesting properties as well as the previous results obtained with the SI as well as the current initiatives around it.

The next chapter presents more in details the spectrum context for the problem at hand as well as the algorithms developed.

Chapter 6

OFDM WLAN Systems

6.1 List of Notations for this Chapter

Table E.3 presents the notations used in this chapter.

Note that these notations are further explained in the remaining of this chapter.

6.2 Introduction

This chapter introduces the background, advantages and challenges of using an OFDM PHY layer. OFDM allows to have a parallel transmission of close but non-interfering radio channels, and the possibility to simultaneously receive from several channels at the same time, without having several Radio Frequency (RF) chains. We believe it is a great source of gain especially in a multiple users context [38] [39], [40]. Then, we present studies of the IEEE 802.11a/h standards as examples of WLAN OFDM-based radio communication systems.

6.3 OFDM Background

6.3.1 Principles

Even though discovered almost 40 years ago, the OFDM technique is nowadays becoming widely adopted by new generations of wireless technologies. In fact, it took several years to develop commercial products coming at a reasonable price. The history of OFDM, a multi-carrier modulation technique, started in 1966 with R.W. Chang, who first developed and patented this idea [41]. In 1971, Weinstein and Ebert proposed to use the Discrete Fourier Transform (DFT) for modulation and demodulation in a multi-carrier mode.

Notation	Description
γ_i	Ratio of successfully transmitted packets from STAtion (STA) i
Δf	Frequency offset between the interferer and the victim receiver
d_0	Some reference distance for free space
k	The Boltzmann constant = 1.381 x 10-23 W/Hz/K
n	Path loss exponent
t	An instant in time
B	Receiver bandwidth
$C_{Blocking}$	Received interfering level from 1 interferer due to blocking
$C_{Interf\ Total}$	Total interfering level reaching a receiver and coming from a single interferer
$C_{Spurious}$	Received interfering level from 1 interferer due to spurious emissions
F_T	Jain's fairness index to estimate the fairness of a topology
G_j^{Rx}	STA j antenna gain (Rx)
G_i^{Tx}	STA i antenna gain (Tx)
G_T	System aggregated goodput
I_j	Sum of all the interfering signals at STA j
K	Number of Inverse Fast Fourier Transform (IFFT) bins
N	Number of contending users in the WLAN cell
$N_{thermal}$	Thermal noise level
PL_{ij}	Propagation loss between STA i and STA j
$PL(d)$	Path loss over a distance d
$PL(d_0)$	Free space loss (on a distance d_0)
PR	Receiver protection ratio in mode 1
$PR(Mode_i)$	Receiver protection ratio in mode i
P_i^{Tx}	STA i packet Tx power
R_{ij}	Signal level received at STA j from STA i
T	OFDM sampling rate
T_{Kelvin}	Absolute temperature [Kelvin]
S	IEEE 802.11 MAC protocol efficiency
S_1	Sensitivity level of mode 1
S_j	Sum of all the signals levels received at STA j
S_t	Sensitivity level of mode 1 plus 20 dB

Table 6.1: List of Notations for this Chapter

OFDM is a transmission technique using frequency-division multiplexing (multi-carrier technique) where the information is transmitted on each SC. OFDM is particularly suited for indoor operations where Non Line Of Sight (NLOS) conditions are most common, due to its ability to mitigate the effects of multi-path (e.g.: 802.11a system), and for outdoor operations (e.g.: DVB-T, 802.16e). In practice, some of the carriers are used for channel estimation (pilots) and also extra bits are added for error detection and correction, thus reducing the number of coded bits, but providing a good performance of the technique, even under difficult channel conditions. An OFDM carrier signal is the sum of a number of orthogonal sub-carriers, with baseband data on each sub-carrier being independently modulated commonly using some type of quadrature amplitude modulation (QAM) or phase-shift keying (PSK). This composite baseband signal is typically used to modulate a main RF carrier.

In OFDM, the subcarrier pulse used for transmission is chosen to be rectangular. This has the advantage that the task of pulse forming and modulation can be performed by a simple Inverse Discrete Fourier Transform (IDFT) which can be implemented very efficiently. Accordingly, in the receiver we only need a Fast Fourier Transform (FFT) to reverse this operation. According to the theorems of the Fourier transform, the rectangular pulse shape will lead to a $\frac{sin(x)}{x}$ type of spectrum of the subcarriers. OFDM provides a selection of modulation constellations for each carrier. For example, the 802.11a standard provides for individual carrier modulation up to 64 Quadrature Amplitude Modulation (QAM).

In OFDM, the relationship between the carrier frequency and the symbol rate is:

- Each frequency carrier is separated by a multiple of $\frac{1}{KT}$ (Hz), K being the number of IFFT bins.
- The symbol rate for each frequency carrier is $\frac{1}{KT}$ (symbols/sec).

$\frac{sin(x)}{x}$ shape to each carrier's spectrum implies that the nulls of the $\frac{sin(x)}{x}$ (for each carrier) are at integer multiples of $\frac{1}{KT}$. Each frequency carrier is located at the nulls for all the other carriers. This means that none of the carriers will interfere with each other during transmission, even though their spectrum overlap. This ability to space carriers so closely (spectrum packing) together is very bandwidth efficient. An OFDM signal is created in the frequency domain and transformed in the time domain via a DFT or a FFT.

Two major problems are usually encountered in wireless systems:

- Inter-Carrier Interference (ICI): guard bands are required to isolate the carriers, consisting in a waste of spectrum.
- Inter-Symbol Interference (ISI): high data rates need short symbol periods, otherwise there is inter-symbol interference.

OFDM modulation resolves both ICI and ISI. Thus, OFDM is an important candidate modulation technique in a broadband and multi-path environment. Also, OFDM makes maximum use of available bandwidth by providing a good packing of the frequencies.

The principle of OFDM is to use overlapping orthogonal SCs, as shown in figure 6.1. For this orthogonality to be preserved, the following properties must stand:

- The receiver and the transmitter must be perfectly synchronized.
- The analog components, part of transmitter and receiver, must be of very high quality.
- There should be no multipath channel.

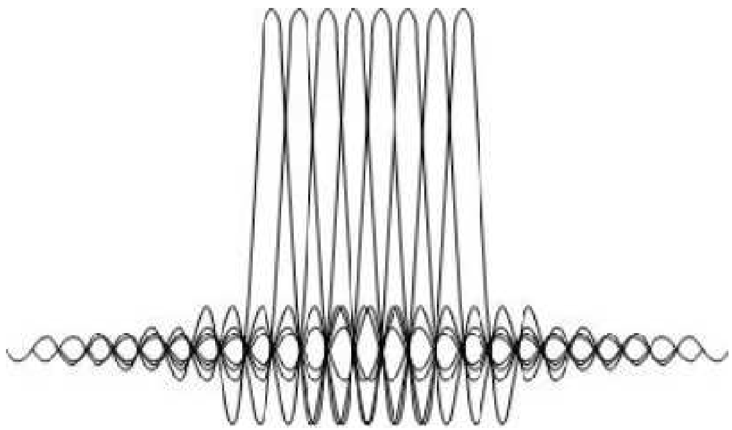


Figure 6.1: Spectrum of a Multicarrier Signal

Fortunately, there's an easy solution to combat the multipath channel problem: the use of a time guard interval per OFDM symbol (refer to figure 6.2). The OFDM symbols are artificially extended by periodically repeating the "tail" of the symbol and preceding the symbol with it. At the receiver, the time guard interval is removed. As long as the length of this guard interval is longer than the maximum channel delay, all reflections of previous symbols are removed and the orthogonality is preserved. However, note that some bandwidth efficiency is lost with the addition of the guard period (symbol period is increased and symbol rate is decreased).

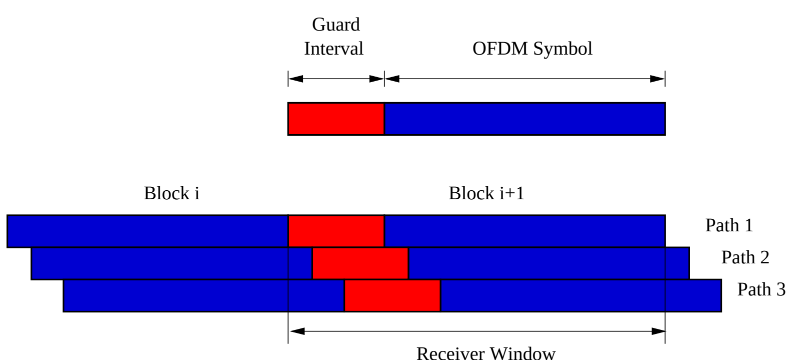


Figure 6.2: Guard Interval

Multipath distortion can also cause inter-symbol interference, which occurs when one signal overlaps with an adjacent signal. OFDM exhibits lower multi-path distortion (delay spread) compared to other techniques, since the high-speed composite's sub-signals are sent at lower data rates. Because of the lower data rate transmissions, multi-path-based delays are not nearly as significant as they would be with a single-channel high-rate system. For example, a narrowband signal sent at a high rate over a single channel will likely experience greater negative effects from delay spread because the transmitted symbols are closer together. OFDM signals typically have a time guard of 800 ns, however, which provides good performance on channels having delays preads up to 250 ns. This is good enough for all but the harshest environments. Delay spread due to multi-path propagation is generally less than 50 ns in homes, 100 ns in offices, and 300 ns in industrial environments.

6.3.2 OFDM Advantages

In addition to offering high data rates for wireless systems, the OFDM technology has several key properties:

- OFDM makes an efficient use of the spectrum resource by allowing subcarriers to overlap [42], [43], [44].
- By dividing the channel (and splitting the information transmission) into several narrow-band flat fading subchannels, OFDM systems are more resistant to frequency selective fading than single carrier systems [45].
- OFDM can remove Inter Symbol Interference (ISI) and Inter Carrier Interference (ICI) by introducing a time guard interval [46].
- Using adequate channel coding and interleaving, one can recover symbols lost due to the frequency selectivity of the channel [47].
- Channel equalization becomes simpler than adaptive channel equalization techniques used in single carrier systems [42].
- DFT modulation and demodulation reduce the complexity of OFDM [44] and allow the simultaneous reception of several channels.
- OFDM is less sensitive to sample timing offsets than single carrier systems are [43].
- OFDM is robust against co-channel interference and impulsive noise [48].
- No intercarrier guard bands: good packing and slicing of the frequencies.
- Resistance against multipath interference using guard interval and cyclic prefix (particularly in wireless communications).
- Ease of filtering out noise (if a particular range of frequencies suffers from interference, the carriers within that range can be disabled or made to run slower).

Usually, the OFDM technique is used in combination with error-correcting codes, adaptive equalization and reconfigurable modulation, leading to Coded Orthogonal Frequency Division Multiplexing (COFDM). Usually, some of the sub-carriers to carry pilot signals, which are used for frequency synchronization. Although highly complex, COFDM has high performance under even very challenging channel conditions. It is a complex technology to implement, but it is now widely used in digital telecommunications systems to make it easier to encode and decode such signals. The system has found use in broadcasting as well as certain types of computer networking technology.

In wide area broadcasting, receivers can benefit from receiving signals from several spatially dispersed transmitters simultaneously, thus actually creating a continuous coverage over a wide area. This is very beneficial in many countries, as it permits the operation of national Single Frequency Network (SFN)s, and avoids the replication of program content on different carrier frequencies which is necessary with FM or other forms of radio broadcasting. Such SFNs exploit the available spectrum more effectively than existing analogue radio networks.

6.3.3 OFDM Challenges

OFDM integrated circuits solution providers have to resolve a series of technical issues to reach a commercial product at a reasonable cost. One consequence of these tradeoffs could be a lower bit rate than expected for a given transmit power and cost. Or, it could result in more external components (to correct and solve the implementation issues), which would cause manufacturing issues, increased cost for a given data rate and range, as well as an increased power consumption due to power-angry external components (e.g.: surface acoustic wave (SAW) filters and crystal oscillators). Implementing the OFDM technology faces several main challenges such as:

- Linearity problems: non-linearity in the RF components could result in interference coming from intermodulation products and reducing the orthogonality of the SCs. Thus, in an OFDM modem design, linearity must be carefully controlled. For the RF components the linearity constraint is even more challenging at the receiver than at the transceiver due to interference from other systems to extract. This places extreme design challenges upon the architectural realization.

- The OFDM time signal has a noise like amplitude with a very large dynamic range, due to the signal being a sum of a large number of subcarriers. Clipping of the signal will introduce both in band distortion and out of band radiation [49]. Also, it is necessary to minimise intermodulation between the SCs, which would effectively raise the noise floor both in-channel and out of channel. For this reason, circuitry must be very linear. Thus, one of the most difficult engineering concerns in the RF portion of traditional OFDM modems is handling very large Peak-to-Average Power Ratio (PAPR)s.

- OFDM is more sensitive to carrier frequency offset and drift than single carrier systems are. Indeed, carrier frequency offset disturbs subcarrier orthogonality and results in inter-channel interference severely degrading the system performance. For high order modulation schemes such as those used for Digital Video Broadcast (DVB), a frequency offset of a small fraction of the subcarrier symbol rate leads to an intolerable degradation. Therefore, frequency synchronization is one of the most critical tasks performed by an OFDM receiver [50]. OFDM requires a perfect frequency synchronization in order to keep the subcarrier orthogonality, otherwise performance degradation occurs.

Even though many technical challenges exist to build commercial working systems, the OFDM technique is very promising and is planned to be used by several standards, as shown in the next section.

6.3.4 OFDM Applications Perspectives for Wireless Systems

OFDM-based products span everything from wireless LAN to mobile access, to digital television. This clearly indicates that several technologies have adopted the OFDM as a basis for their standard:

- ETSI Digital Audio Broadcast (DAB) standard (1995): was the first OFDM based standard. It provides a wireless CD-quality sound transmission.
- ADSL (also called Digital Multi-Tone (DMT)): as COFDM does not interfere easily with other signals is the main reason it is frequently used in applications such as ADSL modems in which existing copper wires are used to achieve high-speed data connections.
- UWB: may also use OFDM as multiband OFDM (MB-OFDM). This UWB specification is advocated by the WiMedia Alliance and is one of the competing UWB radio interfaces.
- Flash-OFDM: it is an OFDM-based standard. It has been developed and is marketed by Flarion. Flash-OFDM has generated interest as a packet-switched cellular bearer, on which area it would compete with GSM and 3G networks.
- OFDM is a candidate for IEEE 802.11n standard for next generation wireless LAN.
- IEEE 802.16-2004 standard for wireless MAN (WiMAX).
- OFDM is a candidate for 4G mobile communications systems.
- ETSI DVB-T (1997) and DVB-H (2004) standards: DVB systems employing COFDM are being standardized by ETSI.
- Wireless LAN standards: IEEE 802.11a [2] (1999) and IEEE 802.11g (2002).

6.3.5 Conclusions

OFDM is successfully used in many applications where high data rates are involved and ruggedness against multipath fading is required. Many other standards in the future will adopt the OFDM as a necessary component (PHY layer). However, in addition to the technological challenges to be solved, interoperability on a global basis between OFDM-based systems is still an issue. Even when using the same multiplex method, interference immune spectrum co-existence and/or inter-system handover capabilities are not necessarily achievable, sometimes due to incompatible MAC layers. The consequence is likely to limit the commercial successes of OFDM-based products as a global solution. Anyhow, OFDM will continue to experience a significant success in the radiocommunication industry. Also, a tradeoff needs to be found between high quality components (resulting in high performance) and cheap devices. The choice depends on the standard involved as well as the provided services.

6.4 Context of the IEEE 802.11 Standards Studies

Several WLAN standards are becoming widely adopted as shown by the tremendous increase in the number of deployed WLAN networks in the cities (hot-spots), in the public places, in the companies and in the homes. The WLAN market is forecasted to continue its increase despite the limitations of the currently available WLAN standards. The constantly launched new WLAN standards reflects an effort of improvement and adaptation to new requirements and services.

The following sections present the IEEE 802.11 standards and especially the IEEE 802.11a/h ones with their capabilities and limitations.

6.5 IEEE 802.11 Presentation

6.5.1 Context

Many WLAN standards are being proposed, some of them and are expected to provide high data bit rate within limited coverage area. Such systems are being deployed in locations with a high user density or high traffic volume and a limited user mobility. For example, WLANs can be used in environments such as home, public, corporate and industrial. In the home environment a WLAN can be used to interconnect electronic equipment, e.g., TVs, computers, surveillance equipment, and sensors. The public environment can support mobile access to backbone networks. Areas most likely to be covered in the public environment are airports, hotels, exhibition halls, shopping centers, and squares. In the corporate environment a WLAN can be used for an infrastructure replacement or to combine an existing wired Local Area Network (LAN) with a WLAN to create more flexibility and allow mobility. In an industrial environment a WLAN can offer connections where the environment is difficult and does not allow wiring. Maintenance and surveillance information can be displayed on handheld terminals.

Due to the complexity of the wireless medium (variable, hard to predict), the performances are sometimes difficult to predict. Also, security problems must be solved, since the medium cannot be physically secured, by means of encryption, which in turn means that security is achieved with an increased complexity. To achieve the mobility, stations must be battery powered and consequently power consumption is a vital issue in a WLAN.

WLAN systems' deployments could raise coexistence problems with other systems especially with the well-known next generation cellular systems, i.e. 3G systems. But, WLAN networks could potentially appear as a complement to other systems in terms of mobility, data rates and offered services. For example, in the future, a third generation cellular system can provide coverage and in hot spot areas WLANs provide increased capacity. In this study only WLAN systems are considered.

Several versions of the IEEE 802.11 standard exist as detailed in the next section.

6.5.2 Background of the IEEE 802.11 Standards

Since the ratification of the IEEE 802.11 standard intended to develop MAC layer and PHY layer standards for wireless connectivity of fixed, portable, and mobile stations within a local area in 1997, various task groups within the IEEE have undertaken to change and enhance portions of the original 802.11 standard, leading to new versions of the standard, each one with its specificity. The following section presents some of these major extensions.

6.5.3 IEEE 802.11 Extensions

The following paragraphs present some of the main extensions of the 802.11 family.

6.5.3.1 802.11a - Higher Data Rates at 5.4 GHz

The 802.11a standard was introduced to allow for higher transmission rates at a higher frequency. The 54 Mbps is a significant improvement to the 11 Mpbs available with the 802.11b. In addition, the use of a separate frequency is perceived as a clean environment for radio communications. However some military operations take place in the same band, thus leading to potential problems. The higher achievable data rates are due in part to the use of an OFDM PHY layer (single-user OFDM), as recognized by its useful properties. However, the range at which 54 Mbps is available from an AP is lower than that of the 11 Mbps from an 802.11b system, so implementation is more expensive as more AP will need to be deployed.

6.5.3.2 802.11b - Current Standard at 2.4 GHz

The 802.11 standard was introduced to increase the original 1 and 2 Mbps data rates to the existing 5.5 and 11 Mbps, but still maintaining connectivity for legacy 802.11 installations. The 802.11b standard was first introduced in September 1999 and became the so-called Wifi© technology. 802.11b was introduced as an extension of the wired Ethernet standards, applying the same principles to wireless communication. The system can work either with or without an access point (ad-hoc operation mode). The 802.11b is backward compatible with to earlier specifications, known as 802.11, allowing speeds of 1, 2, 5.5 and 11 Mbps on the same transmitters. Multiple 802.11b access points can operate in the same overlapping area over different channels, which are subdivisions for the 2.4 GHz band available in each country.

6.5.3.3 802.11c - Bridge Operation

The 802.11c is the standard that provides information, which ensures correct bridge operations take place. The 802.11c standard is incorporated within access points and bridges to allow for bridging to take place.

6.5.3.4 802.11d - Global Harmonization

Originally, the EU, USA and Japan had regulatory bodies in place to make the rules for the use of such equipment. This body was introduced to facilitate the adoption worldwide, and has an ongoing license to define the PHY requirements that satisfy conditions within other countries.

6.5.3.5 802.11e - QoS

The 802.11 standard was introduced to provide full Quality of Service (QoS) features and multimedia support to the 802.11a and 802.11b standards. 802.11e is essential to wireless networks where voice, video and audio are to be delivered. Broadband providers view Qos - and multimedia - capable home networks as an essential element to offering video, audio on demand, voice over Internet Protocol (IP) and high speed Internet access.

Voice, video and multimedia applications are time sensitive, and QoS Baseline proposed an improved way of handling such traffic. It accommodates time scheduled and polled communications during null periods where no data is being sent, which is known as Contention Free Period (CFP)s. In addition, it offers improvements to the efficiency of polling, and enhancements to channel robustness.

Improved channel access during CFP, and the ability to retain polling for backward compatibility, result in more efficient polling. This ability to schedule transmissions and chain a sequence of polls in a single command is also included. These mechanisms provide for maximum efficiency for high-bandwidth streams, power-management friendly implementations, and polled-style access for variable bit rate and bursty streams.

6.5.3.6 802.11f - Access Points Interoperability

The original standard did not specify how communications should take place between access points, thus requiring an additional standardization effort. Indeed, many problems occur with different access points from differing manufacturers, which may not work together when supporting the roaming function. This task force is responsible for enabling the exchange of information between access points to allow for the roaming function.

6.5.3.7 802.11g - Higher Data Rates at 2.4 GHz

The 802.11g Task Group has been given the task of developing a higher speed extension to the existing 802.11b PHY, but maintaining transmission in the 2.4 GHz frequency band. Essentially it consists in an enhanced 802.11b, reaching speeds of up to 54 Mbps. The fundamental difference lies in the use of OFDM instead of the current Direct Sequence Spread Spectrum (DSSS). Current 802.11b users are still able to access an 802.11g network, but with reduced data throughput. In addition, if an 802.11g client accesses an 802.11b access point, the g system reduces its speed to 11 Mbps.

The reduction in speed is due to the inclusion of RTS and CTS protocols, which generates overhead and hence lower the throughput available for all b and g users. RTS and CTS is used as a handshaking system, whereby the client sends an RTS signal and will not send data until it receives an CTS message. This removes the possibility of collisions and hence retransmissions of data, due to b systems being unable to 'hear' the OFDM transmissions of the g network components.

6.5.3.8 802.11h - Spectrum Managed 802.11a

In the EU, the 5 GHz band is split into 2 sections, Band A and Band B as part of this frequency allocation is used by the military, namely for satellite communications and radars. Therefore additional restrictions on the use of this band are required and the 802.11h standard aims to allow for widespread use within the EU. Therefore channel selection and power control has to be incorporated into the 802.11a standard. Dynamic Frequency Selection (DFS) ensures that the portion of the frequency is not accidentally selected in the system's properties, and hence causing potential problems with military operations. Transmission Power Control (TPC) is required as transmissions within Band A are at a lower level than Band B, and hence a mechanism is needed to adjust the power in tandem DFS.

802.11h implements both the DFS and TPC functions resulting in a change of both the MAC and PHY layers compared to the legacy standard.

6.5.3.9 802.11i - Enhanced Security at MAC Level

With the current issues with Wired Equivalent Privacy (WEP), the 802.11i standard defines enhancements to the MAC layer. 802.11 uses a relatively weak and static encryption key without any form of key distribution management. At present, determined hackers can access and decipher WEP-encrypted data on any LAN. 802.11i introduces stronger encryption techniques such as Advanced Encryption Standard (AES) (at the cost of reduced speed performance).

6.5.3.10 Other Extensions of the 802.11 Standard

The IEEE 802.11n task group is working on getting very high performance by improving both the MAC and the PHY layers. Also, the recently launched IEEE 802.11y [51] effort tries to achieve coexitence with other IEEE 802.11 systems.

6.6 Known Problems and Limitations

Refer to Appendix C for details about the IEEE 802.11a model used. Some of the main known limitations of WLAN systems are qualitatively presented in the following sections.

6.6.1 Co-existence of Transmit Modes

The medium use is shared in time between all the active cell users. Each one getting the channel for a duration determined by the data payload size as well as the data transmit mode. In a cell composed of Mobile Terminal (MT)s randomly distributed there is a variety of transmit modes from 1 (slowest but most robust mode) to 8 (fastest but less robust mode), depending on their distance to the AP. However, the co-existence between all the various modes has a tendency to "slow down" the higher modes. Thus, in the presence of slow modes, the high modes cannot fully benefit from the speed offered by high transmit modes.

6.6.2 Collisions and Hidden Nodes

In Distributed Coordination Function (DCF) mode, there is no central point of coordination in the cell, instead the coordination is distributed between all the STAs causing collisions in the channel access. A collision is said to occur when either two nodes that are not hidden transmit simultaneously or when two hidden nodes transmit overlapping frames in time, resulting in no packet being correctly decoded.

Each STA must regularly listen to the medium to determine its state (idle, busy). Due to each STA having limited transmission range, the medium state perception is location-dependent leading to hidden nodes. A hidden node is within the range of the intended destination but out of range of the sender. The result is an increase in the number of collisions, significant performance degradation and unfairness in accessing the medium. Figure 6.3 presents an example of hidden nodes.

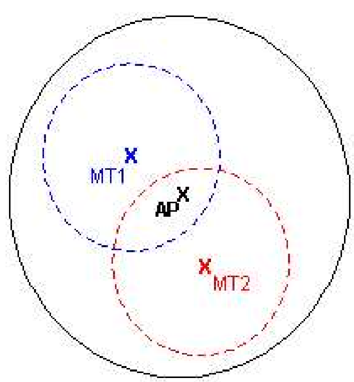


Figure 6.3: Hidden Nodes

In figure 6.3, both MT 1 and 2 are in range of the AP but are hidden (out of range) from each other, potentially creating collisions when trying to access the shared medium.

6.6.3 Exposed Node Problem

The exposed node problem does not occur in UL when there is a radio access infrastructure equipment, but rather it is specific to ad-hoc networks or hybrid networks (ad-hoc networks with a radio access infrastructure equipment). A node is said to be exposed whenever refraining from transmitting as a result of sensing an on-going transmission, while its intended destination is out of range (hidden) from the sensed transmission. Accordingly, the exposed node problem underuses a given spectrum resource which could in fact be successfully reused. Figure 6.4 presents an example of an exposed node.

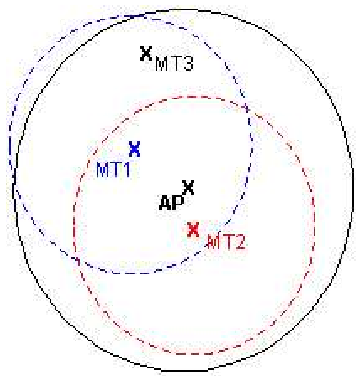


Figure 6.4: Exposed Node

In figure 6.4, MT 2 is sending an UL message to the AP. MT 1 would like to send a message to MT 3 but will refrain because it senses MT 2's on-going transmission. MT 1 is considered as exposed because as MT 3 is out of range from MT 2, MT 1 could have simultaneously sent its message successfully to MT 3.

6.6.4 Capture Effect

Capture effect is said to occur whenever a receiver can clearly receive one transmission out of two simultaneous transmissions, both within its receiving range, resulting in an unfair sharing of bandwidth. Figure 6.5 presents an example of a capture effect.

In figure 6.5, both MT 1 and 2 are in range of the AP and of each other. However, MT 2 is much closer to the AP than MT 1, thus any frame from MT 2 is received at the AP with a much stronger signal than MT 1. As such, for any simultaneous frame transmission of both MT 1 and MT 2, no collision occurs. Instead, MT 2's frame is always successfully decoded by the AP, only considering the combined signal from MT 1 as background noise. Thus, a capture effect is always more beneficial to system throughput than collisions.

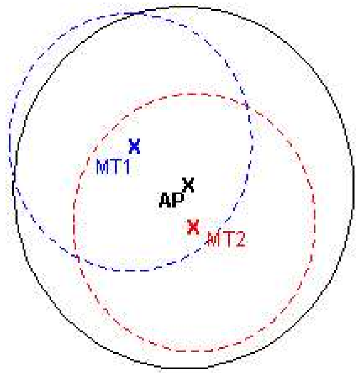


Figure 6.5: Capture Effect

6.6.5 Medium State Threshold

The problem described in this section is specific to the IEEE 802.11a standard. The "channel state" section of [2] specifies that:

"The start of a valid OFDM transmission at a receive level equal to or greater than the minimum 6 Mbit/s sensitivity (-82 dBm) shall cause CCA to indicate busy with a probability > 90% within 4 μs. If the preamble portion was missed, the receiver shall hold the carrier sense (CS) signal busy for any signal 20 dB above the minimum 6 Mbit/s sensitivity (-62 dBm)."

Let us consider an instant in time t. Note that all the radio signals introduced in the following are expressed in mW. Let us define as R_{ij} the signal received at STA j from STA i. R_{ij} is always considered an util signal not an interfering signal. Let us define as S_j the sum of all the signals received at STA j:

$$S_j = R_{ij} + I_j + N_{thermal} \tag{6.1}$$

where I_j and $N_{thermal}$ are respectively the sum of all the interfering signals at STA j and the thermal noise. Accordingly, the following relation is always verified:

$$S_j > R_{ij} \tag{6.2}$$

Let us take as an example the reception of a frame in mode 1 (lowest mode). For STA j to correctly decode the transmitted frame from STA j it is necessary to have:

$$\frac{R_{ij}}{I_j + N_{thermal}} \geq PR \tag{6.3}$$

where PR is the protection ratio (expressed in linear unit) of the frame in mode 1.

Combining equations 6.1 and 6.3 we obtain:

$$\begin{aligned} \frac{R_{ij}}{S_j - R_{ij}} &\geq PR \\ R_{ij} &\geq (S_j - R_{ij}) \times PR \\ S_j &\leq \frac{1+PR}{PR} \times R_{ij} \end{aligned} \tag{6.4}$$

Figure 6.6 presents a graphical interpretation of the medium state given the values of R_{ij} and S_j at STA j for a Tx frame in mode 1.

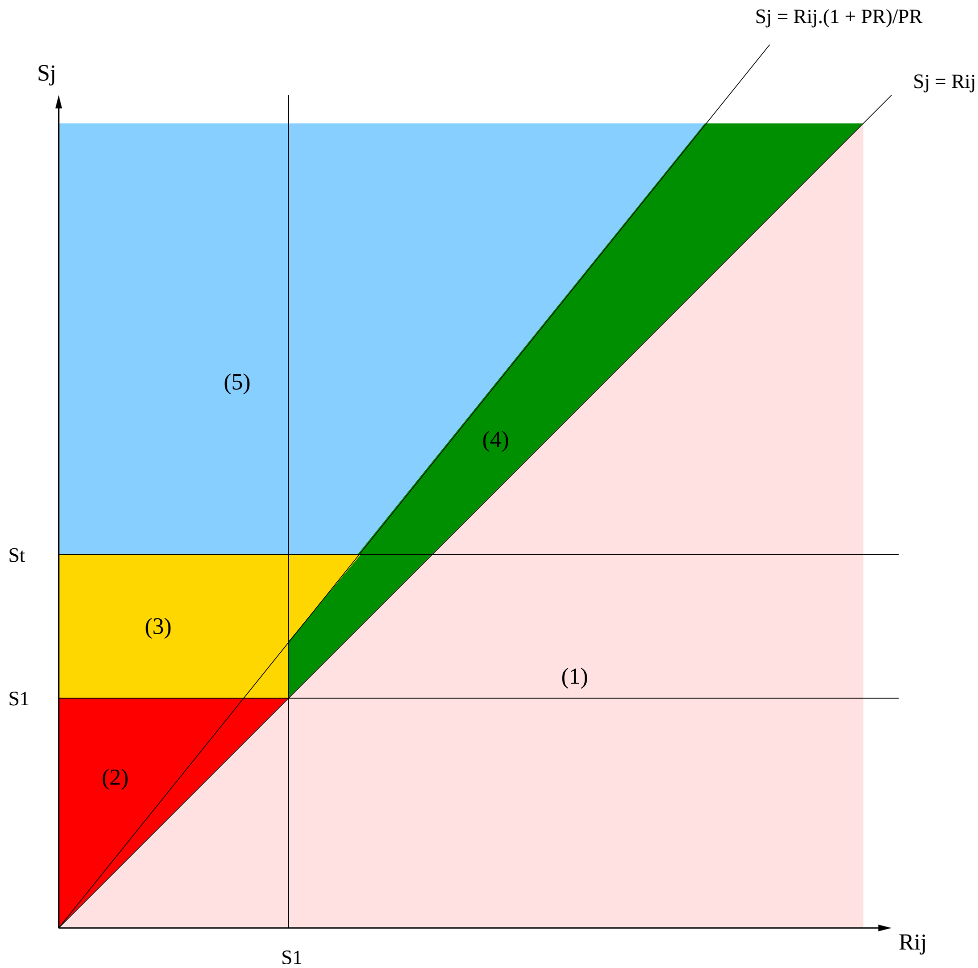


Figure 6.6: Medium and Frame State Areas

In figure 6.6, S_1 and S_t respectively stand for the sensitivity level of mode 1 and the sensitivity level of mode 1 plus 20 dB, in linear unit. Table 6.2 presents the various medium and frame states as interpreted by the STA according to the 802.11 MAC protocol.

As seen in figure 6.6 and table 6.2, the only area for correct reception is area (4). Another interesting remark is that due to threshold S_t, in some cases, a collision of several frames can be

	Area (1)	**Area (2)**	**Area (3)**	**Area (4)**	**Area (5)**
Medium State	Impossible	Idle	Idle	**Busy**	Busy
Frame State	Impossible	Nothing received	Wrong reception	**Reception Ok**	Wrong reception

Table 6.2: Medium and Frame States

interpreted as a creating other hidden nodes. This is contained in area (3). Such a S_t threshold may have been designed to protect the MAC protocol from surrounding interferences. However, an important consequence of such a threshold is that some users are sometimes desynchronized regarding the current medium state in the cell (as hidden nodes). Indeed, as a collision can be interpreted as a idle medium, a STA might be decrementing his Backoff Counter (BC) while others are waiting for a frame response.

6.7 Simulation Studied Parameters

The goal of all these simulations is to quantify by means of simulations the behavior of the 802.11a technology for a deployment of cells and users.

6.7.1 Simulation Metrics

The metrics studied in the simulations are:

- **Goodput:** the goodput is defined as the ratio of the delivered data payload to the total time necessary for transmission including all the protocol overheads (MAC/PHY overheads, backoff delay, interframe intervals, the control frames, the potential frame retransmission times and the other users sharing the same channel).

- **MAC protocol efficiency S:** in term of channel reservation and data transmission success (expressed in %) as:

$$S = \frac{\text{Number of intended Rx ACK frames}}{\text{Number of Tx RTS frames}}$$

  where an "intended Rx ACK frame" corresponds to a successful response to a "Tx RTS frame".

- **Collision rate**

- **Average access delay:** the delay is defined as the average time spent by a packet in the MAC queue, i.e., from the instant it is enqueued till its transmission is complete. Delay is a function of protocol and traffic characteristics. We study the delay to Tx a given amount of data size.

- **Delay distribution** for a variable load of the system.

6.7.2 Fairness Indexes

We refer to a "cell topology" as the combination of AP and users' relative geographical locations. In some simulation scenarios, the fairness of a cell topology was evaluated by looking at the metrics presented in section 6.7.1. The MAC protocol fairness is defined as the capability of the protocol not to exhibit preference to any single node when multiple nodes are using the channel.

To estimate the fairness of a topology Jain's fairness index F_T is computed and applied to several metrics, the standard traditional measure of network fairness [52]. We used a single window size (the entire simulation time) to allow every user enough time to converge to a constant regime. We calculated γ_i the ratio of successfully transmitted packets from STA i over the entire simulation time, for N contending users in the cell. We obtained the expression:

$$F_T = \frac{(\sum_{i=1}^{N} \gamma_i)^2}{N \cdot \sum_{i=1}^{N} {\gamma_i}^2}$$

Another interesting index applied to several metrics is the extreme dispersion of values by using the $\frac{Max}{Min}$ index, ratio of the cell user having the maximum value by the cell user having the minimum value.

6.8 Overall System Deployment Assumptions

6.8.1 Simulation Environment

WLANs will most likely be found in three different environments. The different environments are presented in table 6.3.

Environment	**Examples**	**Common factors**	**Environment specific constraints put on the WLAN system**
Corporate	Offices, hospitals, universities, and hotels	They are rather large, have one owner, and have a lot of internal walls	High attenuations due to many walls
Public (urban)	Exhibition halls, airports, and town squares	Large open areas where many operators could coexist	Coverage increase need due to large open areas
Home	Homes and small offices	Places with small areas and a lot of neighbors	Potentially high level of interference due to aggregation of several close neighbors

Table 6.3: WLAN Deployment Environments

Our simulations model only the corporate and the public environments, places most likely to deploy WLAN systems for business purposes. In order to study the performance and limitations of the 802.11a technology, several scenarios are investigated. A network of one or several APs is deployed in the same geographical region, with a various number of users attached to the APs.

6.8.2 Network Infrastructure Deployment

In the 802.11a simulations the deployed network will always be composed of an infrastructure, no ad-hoc mode will be envisaged. Therefore, according to the envisaged scenario, several APs will be deployed as described in section 6.10.

All the simulations take place in a single building (maximum size: 75 m x 75 m). Some scenario only consider a part of it as the simulation area. The simulations with a network of several co-located cells (each cell being independent/unsynchronized with its neighbors) study the potential inter-cell interference inside the building. The simulations present the results for a 802.11a network. The cell topology (users' distribution and AP relative location) greatly influences system performance.

6.9 User Deployment

In the simulations, all users are deployed indoor, i.e. inside a single building (maximum size: 75 m x 75 m) as described in section 6.10. The indoor walls and obstacles are not modeled but are included in the propagation model as a density of obstructions to the radio signal with a corresponding attenuation. In all the simulations, the users remain fixed at their initial random location in the building. Depending on the scenario, several users' distributions are studied.

6.9.1 System Load

All users arrive at the beginning of the simulation, until the end of the simulation. Thus the number of users attached to a cell is fixed for a given simulation. In other words, no user is arriving in a cell or leaving a cell while the simulation is running.

6.9.2 Service Profile

Several services are considered in a WLAN system. The first is the data service, as it could be known for classical LANs: e-mails, web consulting or data exchanges, etc. The other service is speech, which is enabled by Voice over IP (VoIP).

The two services are specified by their $\left(\frac{C}{I}\right)$ requirements and the data rates necessary as well as maximum acceptable delay. To test the system's performance under saturated load conditions, we implemented a simple traffic model. A user always has a fixed data payload to transmit of

1500 octets (or 2000 octets in some scenarios). Also, all the users in the network use the same data payload during all the simulations. In other words the users' distribution in the network reflects the load distribution.

6.9.3 QoS Control

A communication system must offer the service a sufficient throughput, packets are supposed to be delivered correctly and within a maximum delay to the receiver. If these requirements are fulfilled the system is said to offer QoS. The offered levels of QoS depend on several parameters such as the user service and the mobility profiles.

When several users in the same geographical area simultaneously contend for the same radio resource, conflicts may arise. The WLAN technology under study is intended to support several services with different delivery time requirements such as voice and data delivery. Thus, mapping these services with several priority levels is a potential solution to ensure the user's QoS. The 802.11e technology allows such a possibility. However, no service differentiation is offered in the 802.11a, as the QoS is not controlled.

6.9.4 Propagation Model at 5 GHz (indoor only)

The indoor propagation channel differs from the traditional outdoor mobile radio channel in two aspects: the distance covered are much smaller, and the variability of the environment is much greater for a much smaller range of transmitter/receiver separation distance.

Several radio propagation models can be found in the literature ranging from empirical to analytical or a combination of the two. Not all of them are applicable at the same frequency and are sometimes too site-specific.

The chosen propagation model [53] is based on a power law. It is a compromise between complexity of the model and accuracy of the results in several configurations. This model assumes that the average Received Signal Strength (RSS) decreases with distance raised to some exponent, which value depends on the environment (taking into account internal wall penetration losses, floor attenuation factors, etc).

The loss is assumed to be a power function of distance, thus the total logarithmic loss is given by the following expression:

$$PL(d) = PL(d_0) + 10 \cdot n \cdot log_{10}\left(\frac{d}{d_0}\right)$$

where d_0 is some reference distance and n is the exponent. The theoretical free space loss is given by this expression with $n = 2$ and: $PL(d_0)$ = Free Space Loss (d_0). In the following $d_0 = 1$ m.

No shadowing effect is included. Due to the cell size, the air propagation time ($\ll 1\mu$s) is neglected. Only class A scenarios, referring to corporate environments, have been modeled. The implemented path loss exponent value (NLOS conditions and a single floor building) is $n = 3.6$.

This propagation model is applicable to indoor communicating entities (AP or MT) located in the same building. It will be used to calculate the RSS (useful or interfering signal) between: AP $\leftrightarrow$ AP, AP $\leftrightarrow$ MT and MT $\leftrightarrow$ MT.

6.10 Simulations Scenarios

The scenarios studied in the simulations are presented in the following sections.

6.10.1 Scenario 1

Overview: In this scenario a single isolated cell is considered with a single static user attached to it. No power control mechanism is implemented. Each simulation models a different user to AP distance.

Layout: Figure 6.7 presents the scenario layout.

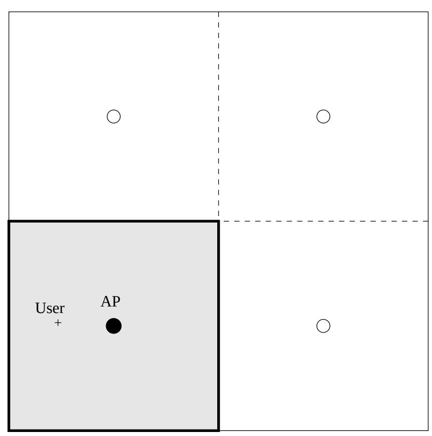


Figure 6.7: Scenario Layout

The shaded area is the simulated area for the current scenario. The AP is located at the center of this area. The user to AP distance is increased by 1 m every new simulation from 1 m to 25 m.

Introduction: The deterministic aspect of the channel modeled (no shadowing effect included) allows to carefully focus on the protocol performance for various user to AP distance.

Objective: The objective of this scenario is to find by means of simulations (analytically verified) the maximum achievable goodput for a not shared channel. Indeed, the single user in the single isolated cell has an exclusive and maximum use of the radio resource.

Indexes: N/A

Metrics: The metrics studied is the user goodput. In this scenario, the user and cell goodput (only UL are equal, because of the exclusive use of the channel.

Proof of Evidence: Some analytical calculations are provided as well as a curve of the maximum achievable goodput versus distance to the AP.

Conclusion: The found maximum achievable goodput (effective throughput seen by the user on top of the MAC layer) is very reduced compared to the PHY rate. The reason is due to the MAC protocol messaging overhead.

6.10.2 Scenario 2

Overview: In this scenario a single isolated cell is loaded with a variable number of static users from 5 to 40 users by increment of 5 users. No power control mechanism is implemented. Each simulation models a different number of users in the cell, but within a simulation the number of users remains constant.

Layout: Figure 6.8 presents the scenario layout.

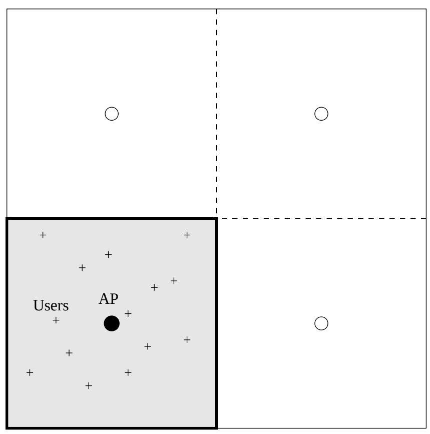


Figure 6.8: Scenario Layout

The shaded area is the simulated area for the current scenario. The AP is located at the center of this area. The user are distributed within the shaded area according to (a) a uniform distribution, and (b) a micro-cluster distribution, see Figure 6.9.

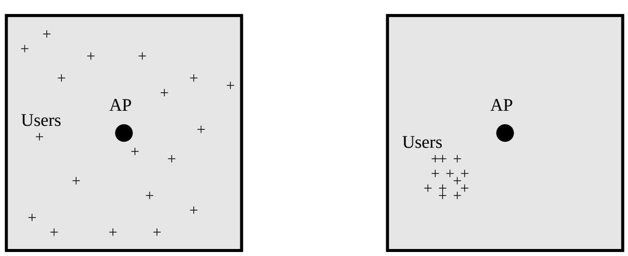


Figure 6.9: (a) Uniform Distribution (b) Micro-Cluster Distribution

Introduction: Uniform and micro-cluster users' distributions will be separately investigated and then compared.

Objective: The objective of this scenario is to show by means of simulations the impact of the cell topology (users distribution and relative AP location) on the MAC protocol fairness and performance in the case of a single isolated cell.

Indexes: The indexes used to measure the fairness of various metrics are: Jain's index and the $\frac{Max}{Min}$ ratio.

Metrics: Various metrics have been considered (mean values over time). The metric studied in these simulations are: goodput, aggregated goodput, BC distribution, Nb Tx RTS, Nb Rx CTS, Nb Rx ACK/Nb Tx RTS, Tx time.

Proof of Evidence: Several curves will illustrate the simulation results.

Conclusion: There is a real impact of the cell topology (users distributions and AP relative location) on the cell performance creating fairness issues, in favor of the very close users to the AP. The higher the cell load in term of number of attached users, the greater the unfairness.

6.10.3 Scenario 3

Overview: In this scenario two cells (two APs) are considered, with a fixed number of attached users. The APs' geographical and/or frequency channel separation is investigated. No power control mechanism is implemented.

Layout: Figure 6.10 presents the scenario layout (example where both APs are centered).

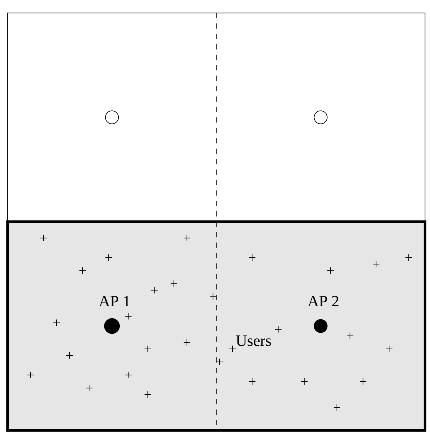


Figure 6.10: Scenario Layout

The shaded area is the simulated area for the current scenario. In this scenario two cells (two APs) are considered. Each AP (AP 1 and AP 2) are located at the center of each half of this area. The number of users (static) is constant for all the simulations and fixed to 40 users. All the users are distributed in the shaded area. It is to be noted that the vertical dashed line does not represent a physical separation in the building but rather indicates the half part of it. For each simulation the APs' geographical and/or frequency channel separation is investigated.

Figure 6.11 indicates the various AP locations simulated.

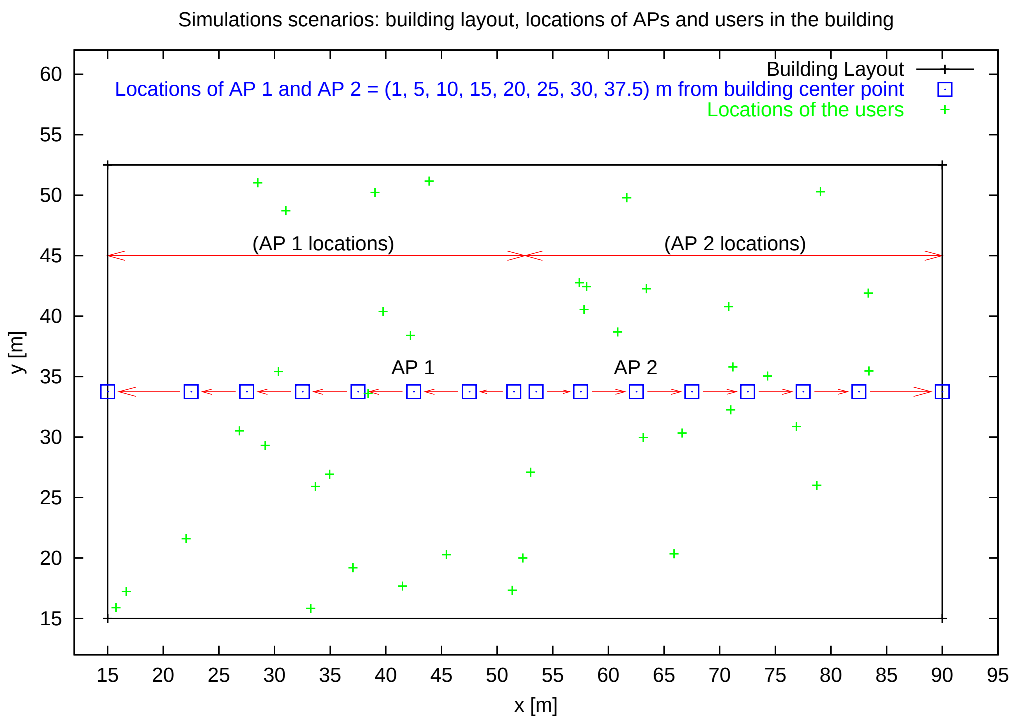


Figure 6.11: Various AP Locations

Introduction: For each relative AP geographical separation, all possible frequency channel offsets will be simulated.

Objective: The objective of this scenario is to illustrate by means of simulations the impact of inter-cell interference on the performance of a 802.11a network of multiple cells.

Indexes: N/A

Metrics: Various metrics have been considered (mean values over time) both at the cell and network levels. The metrics studied in these simulations are: cell load (number of users attached per cell), Nb Rx ACK/Nb Tx RTS, aggregated goodput, user, goodput.

Proof of Evidence: Several curves will illustrate the simulation results as presented in the simulation results section.

Conclusion: The simulations showed that in some configurations the inter-cell interference can greatly reduce cell and network performance. The reasons are an imperfect channel isolation and unsynchronized cells. Accordingly, the recommendation is to deploy APs not too close geographically otherwise even the greatest possible channel separation may not be able to sufficiently isolate each cell and avoid inter-cell interference.

6.10.4 Scenario 4

Overview: In this scenario four cells (four APs) are considered, with a fixed number of attached users. The APs' frequency channel offset is investigated. No power control mechanism is implemented.

Layout: Figure 6.12 presents the scenario layout.

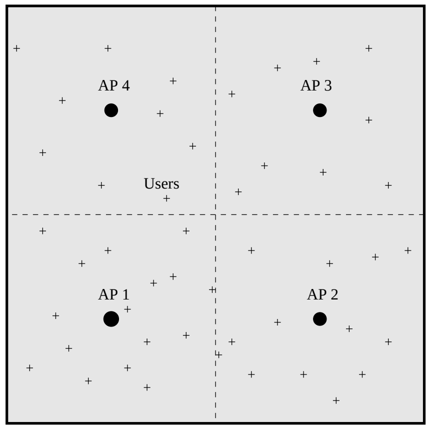


Figure 6.12: Scenario Layout

The shaded area is the simulated area for the current scenario. In this scenario four cells (four APs) are considered. All APs are centered in each fourth of the entire building. The number of users (static) is constant for all the simulations and fixed to 40 users. All the users (48 in total)are uniformly distributed in the shaded area in order to have approximately the same number of users attached to each AP (approximately 12). It is to be noted that the vertical and horizontal dashed lines does not represent a physical separation in the building but rather indicates the half parts of it. For each simulation the APs frequency channel offsets are investigated.

Introduction: APs are relatively far apart, but inter-cell interference can still arise and reduce cell and network performance

Objective: The objective of this scenario is to illustrate by means of simulations the impact of inter-cell interference on the performance of a 802.11a network of multiple cells, for various combinations of channel offsets.

Indexes: N/A

Metrics: Various metrics have been considered (mean values over time). The metric studied in these simulations are: goodput and MAC protocol efficiency.

Proof of Evidence: Several curves will illustrate the simulation results as presented in the simulation results section.

Conclusion: The simulations results indicate that inter-cell interference can greatly reduce cell and network performance. The reasons are an imperfect channel isolation and unsynchronized cells.

6.11 Signal Level Calculations

6.11.1 Packet Transmission

Any packet transmitted by a STA spends a given time in the air equal to its transmission time plus the propagation delay (negligible). The intended receiver as well as the other STAs will potentially receive some radio power (RSS) generated by the packet transmission. The amount of power received by a station depends on various parameters such as:

- The time window offset between the transmitter and the receiver.
- The geographical distance between the transmitter and the receiver.
- The frequency offset between the transmitter and the receiver channels.
- The propagation losses.
- The Rx and Tx masks.
- The antenna gains.

The consequences of a packet transmission inside and outside the current cell (Basic Service Set (BSS)) are:

- For the intended receiver the received power is a useful signal.
- For the other STAs in the same BSS this transmission gives information about the channel state in the BSS.
- For all the STAs in other BSSs and receiving some power it is considered as interference.

At any given time several packets might be present in the air from the same cell or from various cells. The aggregation of these packets at each listening receiver can create interference issues. To monitor the aggregation (duration, power, distance, etc) of all the packets transmitted in the air from all over the network (including all the BSSs of the environment) we will store them all in a single list. They will be sorted by start time. When the current time is greater than the packet end time, then this packet is removed from the list.

Interference is caused only by transmitting stations (MTs or APs) when they send a packet in the air. The Carrier Sensing Multiple Access / Collision Avoidance (CSMA/CA) protocol minimizes the radio spectrum used and thus the throughput. Indeed, at any given instant in a BSS, many stations are waiting whereas only one is transmitting. The idea of transmitting at different data rates (LA algorithm in 802.11a) offer a better chance to succeed in the transmission: the higher the PHY rate, the shorter the transmission time in one transmission attempt. Thus, it also reduces the delay for all the waiting stations. Except that the control frames and the fixed wait times (MAC protocol overhead) introduces fixed delays.

Due to the limited STAs Tx ranges, interference is always limited to a certain area around the Tx.

6.11.2 Signal Level Calculation

The MAC protocol modeled in the simulations is the DCF (CSMA/CA) mechanism. Each station must sense the medium (tuned on the BSS channel) before transmitting. This mechanism is referred to as carrier sensing. To decide the channel state two carrier sensing mechanisms must take place: PHY and MAC. The PHY layer returns the channel state by the Clear Channel Assessment (CCA) response (RSS). The MAC layer carrier sensing (virtual) only takes place when using the RTS/CTS frames, not with the basic access. In the case of RTS/CTS frames transmissions, the Network Allocation Vector (NAV) information indicates the duration of the next intended transmission for a STA. This information can be found in the MAC frames header (RTS, CTS, ACK and DATA frames).

Both the Tx and the Rx use the same channel. The level of signal R_{ij} received at STA j from STA i (same cell) is expressed as:

$$R_{ij} = \frac{P_i^{Tx} G_i^{Tx} G_j^{Rx}}{PL_{ij}}$$

where:

- P_i^{Tx} is STA i packet Tx power,
- G_i^{Tx} is STA i antenna gain (Tx),
- G_j^{Rx} is STA j antenna gain (Rx),
- PL_{ij} is the propagation loss between STA i and STA j.

6.11.3 Interference Model

Interferences represent some unexpected signals combined at the Rx with the wanted signal and necessitating to separate them in order to successfully decode the intended message. Interferences may come from various Txs as presented in the following sections.

6.11.3.1 Thermal Noise level

Thermal noise is always present at the Rx, independently of the system load. The equivalent noise power in dBm at the input of a receiver is given by the well-known formula:

$$N_{thermal} = kT_{Kelvin}B$$

where:

- k is the Boltzmann constant = 1.381 x 10-23 W/Hz/K,
- T_{Kelvin} the absolute temperature (at room temperature) = 290 K,
- B the receiver bandwidth = 18 MHz.

Thus, the noise floor level (transmission mode independent) for both the APs and MTs is: $N_{thermal}$ = -101 dBm. The thermal noise floor level is much lower than the minimum Rx sensitivity (-82 dBm), thus it does not generate interference by itself but is sometimes combined with other sources of interference.

6.11.3.2 Intra-cell Interference

In a 802.11a network, only one frequency channel is used per cell for both UL and DL. Thus, all active STAs attached to a cell simultaneously contend for the medium. Collisions could occur if several packets were simultaneously transmitted. If two or more packets are simultaneously Tx using the same frequency channel, the intra-cell interference is called collision. Collisions result in none or one packet to be correctly received.

6.11.3.3 Inter-cell Interference

Inter-cell interference occurs when a signal is received from a Tx attached to another cell (using another frequency channel). In a 802.11a network composed of several cells, all cells are independent from each other and not coordinated.

According to the standard, a total of 8 channels are available for indoor operations. For important deployments of APs in the same area, a frequency planning is necessary so to avoid inter-cell interference. Such a frequency planning can be either done by hand or dynamically (DFS). We choose not to implement the DFS mechanism, thus the frequency planning will be done manually and depending on the considered scenario. The isolation between channels is not perfect and is specified by the Tx and Rx masks.

Unwanted Emissions at the Receiver [1]

Unwanted emissions consist of spurious emissions and out-of-band emissions:

[1] Draft revision of recommendation ITU-R SM.329-7, *Spurious emissions*

- **Spurious emissions** is the emission on a frequency, or frequencies, which are outside the necessary bandwidth and the level of which may be reduced without effecting the corresponding transmission of information. Spurious emissions include harmonic emissions, parasitic emissions, intermodulation products and frequency conversion products but exclude out-of-band emissions.

- **Out-of-band emissions** is emission on a frequency or frequencies immediately outside the necessary bandwidth which results from the modulation process, but excluding spurious emissions. For a given class of emissions, the **necessary bandwidth** is the width of the frequency band just sufficient to ensure the transmission of information at the rate and with the quality required under specified conditions.

The level of unwanted emissions falling within the victim's receiver bandwidth is determined using the interferer transmit mask, interferer/victim frequency separation, antenna gains and propagation loss.

For all configurations of receiver versus interferer channels let us calculate the spurious level captured in the receiver band (assumed to be 18 MHz).

Co-channel Interferer: Both the receiver and the interferer are using the same channel. Thus, the co-channel spurious level at the receiver frequency is 0dB below the interferer transmit power.

Adjacent Channel Interferer: The interferer is transmitting on a channel separated by only 20 MHz from the current receiver channel. As the transmit mask is not perfect, the receiver is receiving a certain level of interference in its receiving bandwidth. The resulting spurious level captured by the receiver is the integration of the transmitter mask for the following offset interval: [11 MHz - 29 MHz]. The adjacent channel spurious level at the receiver frequency is finally 24.75 dB below the interferer transmit power.

Alternate Adjacent Channel Interferer and All Other Channels: The interferer is transmitting on a channel separated by at least 40 MHz from the current receiver channel. The alternate adjacent channel and other channels spurious level at the receiver frequency is finally 40 dB below the interferer transmit power.

All these values are presented in the following sections and added to the receiver attenuation level.

Blocking [2]

Blocking is a measure of the capability of the receiver to receive a modulated wanted input signal in the presence of an unwanted input signal on frequencies other than those of the spurious responses or the adjacent channels, without these unwanted input signals causing a degradation of the performance of the receiver beyond a specified limit.

[2] Document I-ETS 300 113: 1992

The receiver blocking power i.e. the power captured from the on-channel transmissions of the interferer due to selectivity imperfections of the victim's receiver, is determined using the interferer's transmit power, victim receiver blocking performance, interferer/victim receiver frequency separation, antenna gains and propagation loss.

To determine the blocking level we will use the receiver blocking mask presented in section Receiver characteristics (§17.310.1-17.310.4 of [2]) and interpolate the corresponding value $y \in [y_1, y_2]$ at the needed frequency $x \in [x_1, x_2]$ with the following formula:

$$y = y_1 + \frac{y_2 - y_1}{x_2 - x_1} \times (x - x_1)$$

where $(x1, y1)$ and $(x2, y2)$ are the line segment limits.

The receiver attenuation value depends on the mode used, and is defined as:

$$Attenuation(Mode_i) = 3\ dB + PR(Mode_i) + Blocking(\Delta f)$$

where:

- $PR(Mode_i)$ is the receiver protection ratio,
- Δf is the frequency offset between the interferer and the victim receiver.

The protection ratio is defined as the ratio between the noise floor $N_{thermal}$ and the receiver sensitivity level of the considered Tx mode (given in section Receiver characteristics (§17.310.1-17.310.4 of [2]). Thus the attenuation is given by:

$$Attenuation(Mode_i) = 3\ dB + Sensitivity(Mode_i) - N + Blocking(\Delta f)$$

The blocking (considered a positive value) level is 27 dB for adjacent channel ($\Delta f = 20$ MHz) is 27 dB, and 45 dB for all other channel offsets. Table 6.4 shows the attenuation levels for each mode and for every combination of channel offsets.

Mode	**Data rate [Mbps]**	**Minimum sensitivity [dBm]**	**Protection ratio [dB]**	**Adjacent attenuation [dB]**	**Alternate Adjacent channel attenuation [dB]**
1	6	-82	19	49	67
2	9	-81	20	50	68
3	12	-79	22	52	70
4	18	-77	24	54	72
5	24	-74	27	57	75
6	36	-70	31	61	79
7	48	-66	35	65	83
8	54	-65	36	66	84

Table 6.4: Receiver Attenuation Capabilities

The real received interfering level due to blocking is:

$$C_{Blocking} = \frac{P^{Tx}_{Interf} G^{Tx}_{Interf} G^{Rx}_{Victim}}{PL \times Attenuation_{Blocking}}$$

Total Interference Calculation

The total interfering level reaching a receiver and coming from a single interferer is given by:

$$\begin{aligned} C_{Interf\ Total} &= C_{Spurious} + C_{Blocking} + N_{thermal} \\ &= \frac{P^{Tx}_{Interf} G^{Tx}_{Interf} G^{Rx}_{Victim}}{PL \times Attenuation_{Total}} + N_{thermal} \\ &= \frac{P^{Tx}_{Interf} G^{Tx}_{Interf} G^{Rx}_{Victim}}{PL \times Attenuation_{Spurious}} + \frac{P^{Tx}_{Interf} G^{Tx}_{Interf} G^{Rx}_{Victim}}{PL \times Attenuation_{Blocking}} + N_{thermal} \quad (6.5) \\ & \quad (6.6) \end{aligned}$$

Thus:

$$Attenuation_{Total} = \frac{1}{\frac{1}{Attenuation_{Spurious}} + \frac{1}{Attenuation_{Blocking}}}$$

Table 6.5 gives the $Attenuation_{Total}$ for interfering levels at a receiver for each mode and for every combination of channel offsets. These values are relative to the interferer transmitting power. They do not take into account the path loss, antenna gains and interferer transmitting power, noise floor.

Data rate [Mbps]	**Spurious Level† [dB]**			**Attenuation Level [dB]**			**Total Interfering Level [dB]**		
	Co-channel	**Adjacent channel**	**Alternate Adjacent channel**	**Co-channel**	**Adjacent channel**	**Alternate Adjacent channel**	**Co-channel**	**Adjacent channel**	**Alternate Adjacent channel**
6	0	24.75	40	N/A	49	67	0	24.73	40
9	0	24.75	40	N/A	50	68	0	24.74	40
12	0	24.75	40	N/A	52	70	0	24.74	40
18	0	24.75	40	N/A	54	72	0	24.74	40
24	0	24.75	40	N/A	57	75	0	24.75	40
36	0	24.75	40	N/A	61	79	0	24.75	40
48	0	24.75	40	N/A	65	83	0	24.75	40
54	0	24.75	40	N/A	66	84	0	24.75	40

†Calculated for a Tx bandwidth of 18 MHz.

Table 6.5: Interfering Level at a Receiver

All the above values are considered as attenuations of the interferer transmit power. Thus, the above interfering levels (in dB) will be subtracted from the interferer transmit power (in dBm). It is to be noted that the influence of spurious emissions is much greater than the blocking.

Signal Aggregation Problem

However, the above calculations only accounts for a single interferer. Thus, the total interfering power I_i received at receiver i is a sum over all transmitting interferers.

However, the aggregation phenomenon is very limited in 802.11 systems compared to other systems.

6.12 Simulation Results

6.12.1 Maximum Achievable Goodput in an Isolated Cell

This section corresponds to scenario 1 as described in section 6.10.1. The user transmit mode is determined by the transmit power and the distance to AP. The maximum achievable goodput for a single user in a single isolated cell is shown in figure 6.13. It was obtained with the maximum transmit power (23 dBm).

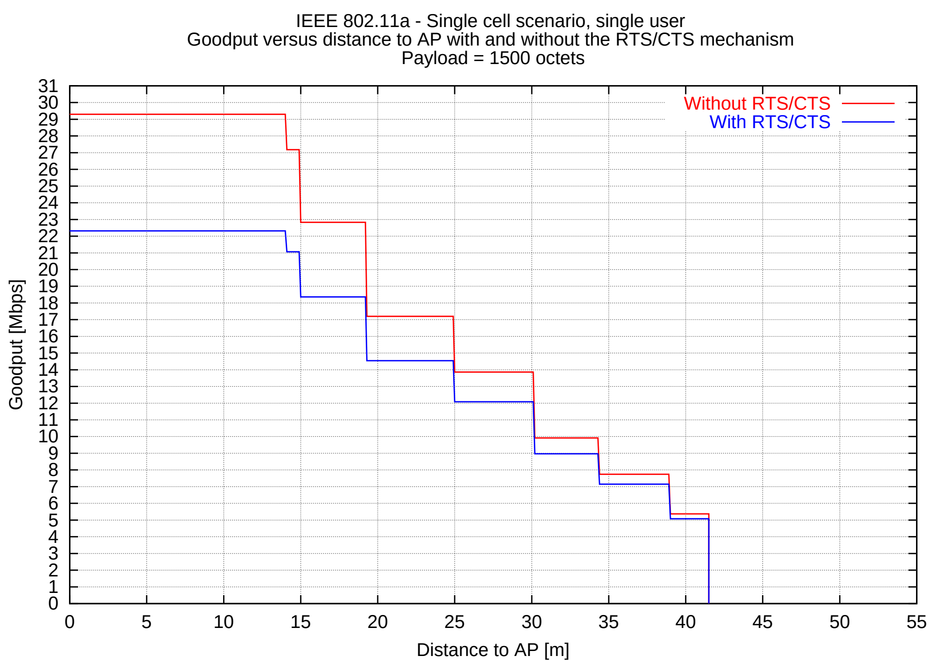


Figure 6.13: Maximum Achievable Goodput for a Single User in a Single Isolated Cell

Each "step" corresponds to a transmit mode: from mode 8 to mode 1. When a cell contains only a single user, no collision nor capture effect can occur. Accordingly, using the RTS/CTS mechanism is an overhead as shown in figure 6.13. Another remark is that the maximum achievable goodput is different from the PHY rate and is the effective performance offered at the top of the MAC layer for a user. The simulated values are superimposed with the theoretical values.

6.12.2 Impact of User Distribution

This section corresponds to scenario 2 as described in section 6.10.2.

6.12.2.1 Introduction

The 802.11 MAC protocol suffers known problems such as hidden nodes and capture effects, leading to degraded performance. The aim of this scenario is to investigate the impact of the users' distribution (uniform and micro-cluster) on the protocol fairness and performance. We show my means of simulations that the MAC protocol is highly sensitive to the cell topology, leading to a spatially unfair system in terms of channel access, reservation and transmission success. As a consequence, cell users experience very different performance from each other.

6.12.2.2 Related Work

In the WLAN literature no analytical model of a single 802.11a cell with several users can be found and incorporating the effects of: cell topology, cell load and distribution, users' transmit modes, users' data payload size, MAC protocol, radio parameters. However, many simplified models have been developed. Bianchi [54], assuming that all users always see the channel in the same state, found that the system throughput expression was almost independent from the cell load when using the RTS/CTS mechanism. In [55], a LA algorithm is detailed for a single user, using an expression of the effective goodput. In [56], Chhaya and Gupta have studied the protocol behavior under realistic radio conditions such as hidden nodes, capture effects, collisions, etc... They derived a simplified throughput model resulting in a spatially unfair protocol in favor of users close to the AP. Many other articles [57], [52] have pointed out the 802.11 MAC protocol unfairness and tried to measure it with some metrics, sometimes resulting in an enhanced MAC protocol. This scenario deals with the infrastructure mode, where there is always an AP in the cell.

6.12.2.3 General Presentation

We refer to "cell topology" as the combination of AP and users' relative geographical locations. We modeled a fully deterministic channel (no shadowing effect), so the RSS is fully determined by the distance and the transmission power. Thus, the only random parameters are: users' distribution and random BC values. As shown in section C.2.3, the MAC protocol suffers collisions in accessing the shared medium due to:

- Distributed MAC scheduling (causing users to transmit simultaneously),
- Hidden nodes.

Using the MAC protocol in DCF mode, several contending users cannot avoid collisions even without any hidden nodes or capture effects. A cell can only achieve maximum performance with a single user. However, the system is still fair if all the contending users equally suffer from the collisions. We simulate a scenario where the users are all grouped in a micro-cluster, to avoid hidden nodes.

Then, we investigate the effect of hidden nodes, capture effects and radio propagation on the medium state perception and thus on the spatial fairness of the MAC protocol. The constraints for two MTs to be in the same protocol state are:

1. Reach the AP with almost the same received signal level,
2. Receive almost the same average signal level from all the other users (excluding the AP).

As there is no power control in 802.11a, all the STAs use the same transmit power. Constraint 1 means that all users must be at the almost same distance from the AP, and constraint 2 means that all users must have almost the same average inter-distance with each other. Two different users' distributions are tested: uniform distribution with various number of users, and micro-cluster distribution with a fixed number of users but a moving micro-cluster creating hidden nodes and capture effects.

6.12.2.4 Micro-Cluster Distribution

We simulate a total real time of 100s. Several scenarios are investigated with a constant number of 6 users, to show the impact of hidden nodes, capture effects and collisions on the MAC protocol fairness. In a micro-cluster, several users are closely grouped in a 2m radius circle, all using the same transmit mode. For each series of 4 scenarios (1 to 4), the micro-cluster centre is respectively located at 0m, 7.5m, 17.5m and 22.5m from the AP.

- *Scenarios G1 to G4*: All 6 users are in the micro-cluster,
- *Scenarios IC1 to IC4*: There is always 1 user at 5m from the AP, and a micro-cluster of 5 users, diametrically opposed,
- *Scenarios IF1 to IF4*: There is always 1 user at 25m from the AP, and a micro-cluster of 5 users, diametrically opposed.

Scenarios G1 to G4 show fair situations, no matter the distance between the micro-cluster and the AP, because all users are always in the same state, and there is no hidden node/capture effect. In scenarios IC1 and IC2, there is no hidden node/capture effect, giving also a fair situation. However, in scenarios IC3 and IC4 we have a capture effect but no hidden node: any RTS attempt from a user in the micro-cluster is always dominated by a simultaneous RTS attempt from the user close to the AP (5m) because of its much stronger RSS at the AP. Thus, he achieves almost 100% of reservation success leading to an unfair situation (RTS packets fairness index = 0.87). Once the channel is reserved, data is nearly transmitted with total fairness (0.97). In scenarios IF1 and IF2, the capture effect is dominated by the micro-cluster's users leading to an RTS packet fairness index of 0.98. Then the capture effect becomes a hidden node problem in

Scenarios	User Nb Tx RTS		User S		User DATA success		G_T [Mbps]
	$\frac{Max}{Min}$	F_T	$\frac{Max}{Min}$	F_T	$\frac{Max}{Min}$	F_T	
G1	1.15	1.0	1.08	1.0	1.18	1.0	21.82
G2	1.15	1.0	1.04	1.0	1.18	1.0	21.82
G3	1.11	1.0	1.02	1.0	1.12	1.0	17.98
G4	1.13	1.0	1.03	1.0	1.16	1.0	14.26
IC1	1.15	1.0	1.08	1.0	1.18	1.0	22.40
IC2	1.23	1.0	1.04	1.0	1.27	0.99	21.57
IC3	2.33	0.87	1.49	0.97	3.46	0.71	20.82
IC4	2.30	0.87	1.49	0.97	3.4	0.72	17.67
IF1	1.45	0.98	1.06	1.0	1.38	0.98	19.37
IF2	1.46	0.98	1.07	1.0	1.38	0.99	19.33
IF3	1.20	1.0	1.80	0.97	1.68	0.98	15.99
IF4	1.21	1.0	1.72	0.97	1.55	0.98	13.19

Table 6.6: Micro-Cluster Distribution Fairness and Performance Metrics

scenarios IF3 and IF4 because the micro-cluster's centre and the isolated user are too far apart (42.5m and 47.5m respectively). The result is an unfair situation as shown in table 6.6.

As a conclusion, the MAC protocol is fair when all users are always "synchronized" on the same channel and protocol state and face the same number of collisions. This is achieved when there is no hidden node nor capture effect. Such situations are found for distributions of users locally concentrated, such as users in the same small meeting room. But, as soon as hidden nodes or capture effects appear, the MAC protocol turns out to be unfair and in favor of the very close users to the AP (whether isolated or grouped in micro-cluster).

6.12.2.5 Uniform Distribution

We simulate a total real time of 200s. We vary the cell load from 5 up to 40 users. The users are uniformly distributed in the building. Thus, when the number of users increases, the proportions of users in each mode remain almost constant. Table 6.7 shows the metrics used.

Distribution of BC Values

Prior to data transmission, each STA draws a BC value to determine its transmission start time. The protocol retry philosophy is "the more a STA fails, the more it waits, whatever its mode and data packet size". Figure 6.14 shows the distribution of BC values versus the distance to the AP. These values are obtained for a cell load of 40 users.

In figure 6.14, various "steps" are found along the BC value axis, corresponding to the different values of Contention Window (CW). First, the number of occurrences of BC values rapidly decrease for increasing BC values. The most significant "steps" being actually contained in the first two CWs. In other words, a success usually follows a collision. Second, the function giving the successful transmit opportunities versus the distance to the AP is not a simple one. Even though the closest users from the AP are the most favored ones, the farthest are not

Nb of users	**User Nb Tx RTS**		**User S**		**User DATA success**		G_T **[Mbps]**
	$\frac{Max}{Min}$	F_T	$\frac{Max}{Min}$	F_T	$\frac{Max}{Min}$	F_T	
5	1.36	0.99	1.11	1.0	1.37	0.99	18.02
10	2.24	0.96	1.49	0.99	3.34	0.92	18.51
15	2.14	0.97	1.88	0.98	4.04	0.90	18.48
20	2.71	0.92	2.00	0.97	4.66	0.77	18.56
25	2.38	0.95	1.91	0.97	3.74	0.84	18.38
30	2.82	0.93	2.03	0.97	4.79	0.80	17.67
35	3.23	0.92	2.10	0.97	5.92	0.77	17.44
40	3.21	0.91	2.19	0.97	6.01	0.77	17.49

Table 6.7: Uniform Distribution Fairness and Performance Metrics

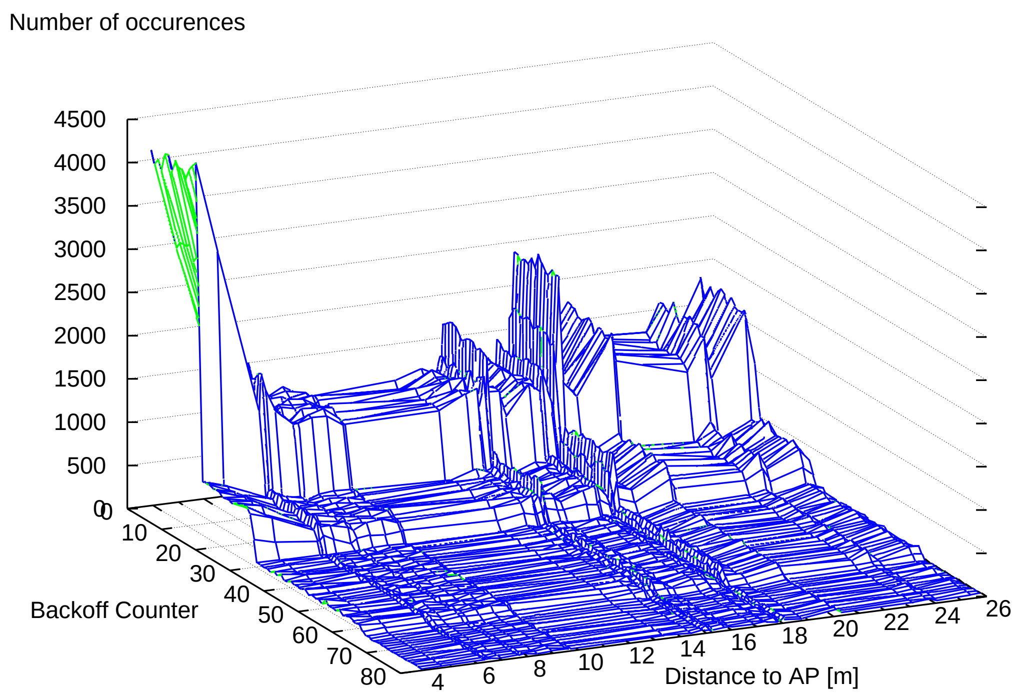


Figure 6.14: Number of Times Each BC Value was Drawn Versus Distance to AP

necessarily the least favored as seen in figure 6.14. This is the impact of the cell topology. If all users were equal regarding success and failure, they would have drawn the same number of times each BC value regardless of their location in the cell. However, this is not verified by figure 6.14, thus showing the spatial unfairness of the MAC protocol, even with a fairly large cell load.

Opportunity of Channel Access, and Success in Channel Reservation and Data Transmission

The profile of the number of RTS frames per user versus the distance to the AP follows the BC distribution (figure 6.14). MTs having few channel access opportunities are suffering a high percentage of collisions and a high waiting time between two transmissions, leading to poor performance. In table 6.7, when the number of contending users increases, the number of RTS $\frac{Max}{Min}$ values, over all users, also increases and reaches up to 3.2 and a fairness index of 0.92, for 40 contending users. This indicates the spatial unfairness of the MAC protocol in the channel access.

Figure E.5 shows S versus cell load and the distance to the AP.

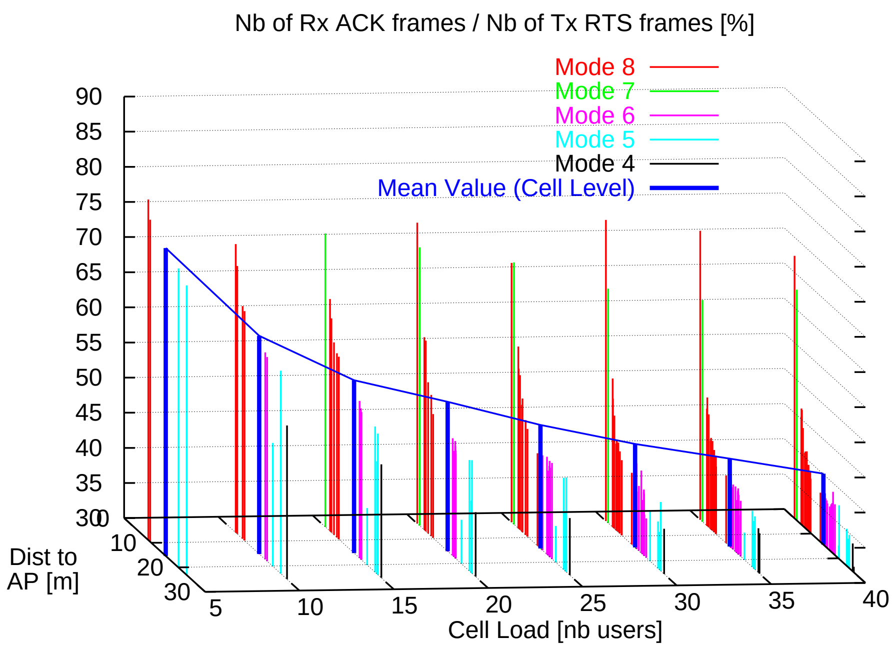


Figure 6.15: S (%) Versus Cell Load and Distance to AP

The spatial unfairness of the protocol is obvious. The assumption that all users have the same S conditions regardless of their location is not verified. The importance of using the RTS/CTS mechanism in presence of many users to minimize the effect of the hidden nodes is highlighted. Indeed, the average S value per user decreases for an increasing number of contending users, due to an increasing number of hidden nodes and thus of collisions. In figure E.5, the users having the best S values are the ones located close to the AP. As no power control is used in 802.11a, users close to the AP exchange stronger radio signals with the AP than the ones far from it. Thus, users at the cell border are disadvantaged regarding the channel reservation efficiency compared to the users at the centre. As cell load grows, some users become more and more disadvantaged compared to others (up to 2 times), leading to an increasingly unfair protocol. In a larger building (more hidden nodes), the users far from the AP (mode 1) would have been even more disadvantaged.

Goodput

Figure 6.16 corresponds to a combination of users operating in various transmit modes.

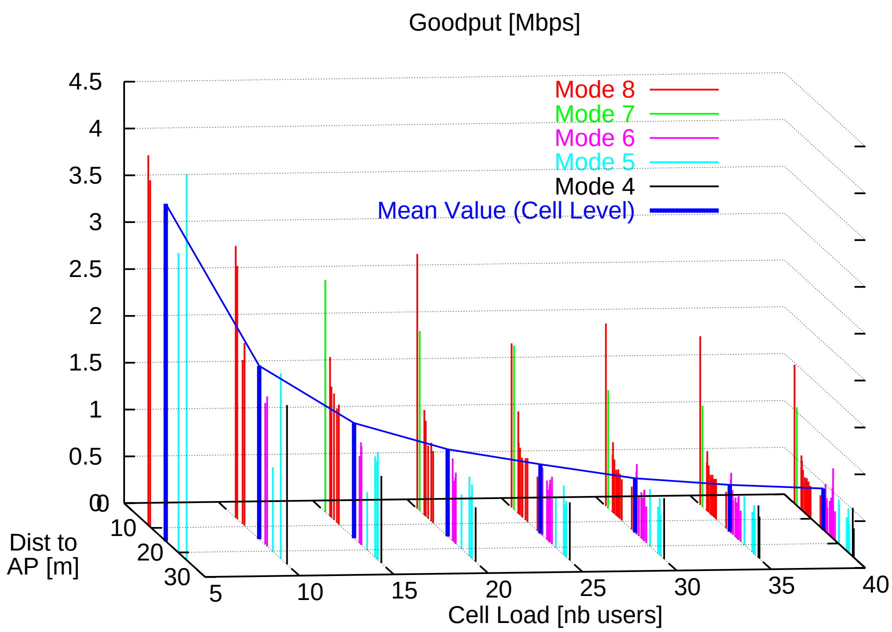


Figure 6.16: Goodput Versus Cell Load and Distance to AP

First, the average goodput per user decreases for an increasing cell load, in favor of the MTs close to the AP. Indeed, they already have a better opportunity of channel access, medium reservation success and data transmission success, and they also have a better mode. Thus, for 40 contending users, the most favored users send 6 times more successful data frames than the least favored ones. Second, even under heavy load conditions, the average cell goodput (considering an average mode for all users) is almost inversely proportional to the cell load. Table 6.7 shows that the aggregated goodput is almost constant whatever the number of users. Thus, the user average time to transmit a given data payload almost linearly increases with the number of contending users, creating a greater delay, due to the shared radio channel.

6.12.2.6 Conclusion

Our simulation results exhibit a spatially unfair 802.11 MAC protocol and the effects of hidden nodes and capture effects. Even users transmitting the same amount of data payload faced different opportunities to contend for the channel, different level of success in channel reservation and data transmission, depending on the cell topology. Cell topology is thus a key parameter to consider, for example, when triggering inter-cell handovers.

6.12.3 Inter-cell Interference in 802.11a (no power control)

This section corresponds to scenario 3 and 4 respectively described in sections 6.10.3 and 6.10.4.

6.12.3.1 Introduction

WLAN systems such as IEEE 802.11a bring end-user to the high-speed data rates within limited coverage areas. Deploying additional access points might extend the coverage as well as create inter-cell interference, thus reducing the network performance, as studied in this article through various scenarios of overlapping cells.

WLANs are being more and more deployed in homes, public places and offices. They offer high data rates to end-users but are limited in coverage area, thus necessitating AP densification to provide all users with sufficient quality of service. Inter-system and intra-system interferences are significant issues in WLANs. The reasons are a small number of available channels for 802.11a [1], [2] systems, as well as insufficient isolation between them. Accordingly, there is the need for appropriate deployment strategies taking into account geographical and channel separation between overlapping cells.

6.12.3.2 Related Work

For a given area to cover, finding the optimum AP locations [58] is a complex task to achieve, depending on various objectives: cover the greater area with the smaller number of APs, or minimize capture effects to maximize fairness, or minimize inter-cell interference, or maximize overall network goodput. In each case, the final network configuration might be different. The AP placement optimization algorithms developed are either complex solutions (optimum solutions) or simple ones (sub-optimum solutions obtained quickly and at a reasonable cost). The parameters usually considered by the planning algorithms are the propagation model and the geographical separation between APs. In fact, all the following parameters must be considered: channel separation, geographical distance between APs and the cells' topologies (users' distribution and AP relative position). The aim of this scenario is to show by means of simulations that the impact of inter-cell interference can be detrimental to network performance.

6.12.3.3 General Presentation

We first investigate the impact of a single AP placement (no inter-cell interference) on the coverage and the network performance. Then we study the case with two APs. Adding another cell can increase the offered load as well as increase the level of inter-cell interference. In the case of overlapping cells, all the following inter-STA interference cases have to be considered: inter-AP, inter-MT and MT-AP.

6.12.3.4 Coverage

Our simulations showed that a single AP can cover a maximum of 95% of the entire building (38 users) when located near the building centre (1 m offset). Then, as the AP offset increases, the maximum number of users potentially attached to its cell almost linearly decreases until 47.5% (19 users) for 37.5 m of AP offset. For a user to be attached to a given cell, he chooses the AP signal received with the maximum RSS. As our propagation model is deterministic, it is always the geographically closest AP. Thus, each AP coverage area is determined by the inter-AP distance.

For two APs, cell 1 (left) and cell 2 (right) are approximately equally loaded (half the total number of users) whatever the AP offset. The result is a maximum network coverage of 100% (from 10 m to 30 m AP offset) and a minimum of 92.5% at 37.5 m AP offset. The cell load will impact the performance as presented in the following sections.

6.12.3.5 S Parameter

Figure 6.17 shows the average S values per user (cell level) without (1 AP) and with (2 APs) inter-cell interference.

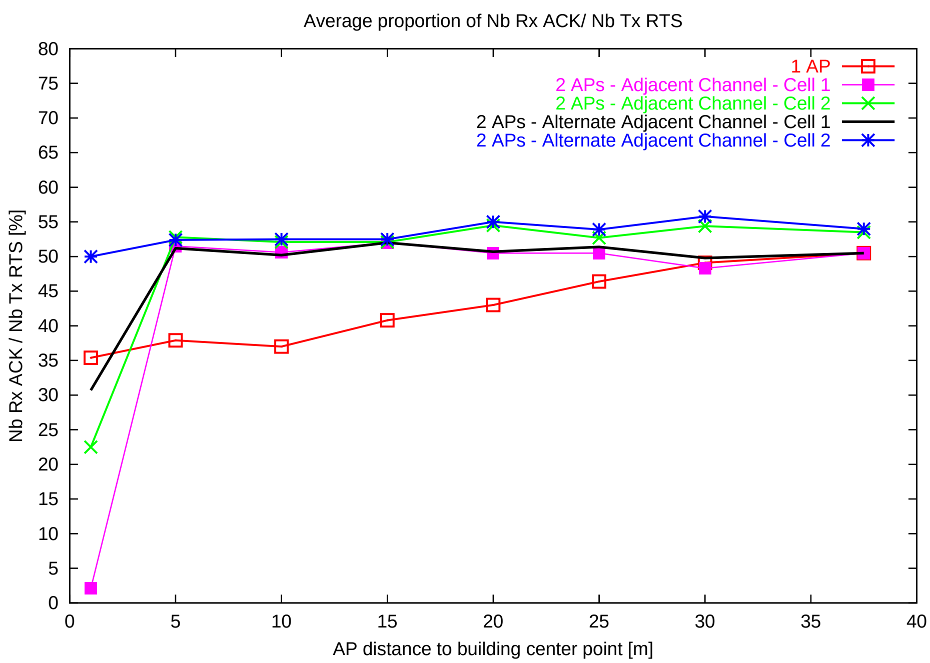


Figure 6.17: S Parameter

Due to the known spatial unfairness of the MAC protocol, the average S values per user at the cell level do not reflect the differences experienced at the user level. Some AP locations will favor some users compared to others. For a single AP, the greater the AP offset the smaller the number of attached users, and so does the number of collisions. The result is an increase of S value per user going from 35% to 50%.

For two APs, the larger the AP offset, the smaller the overlap between both cells, and so does the inter-cell interference. For each cell and for a given AP offset (from 5 m to 37.5 m), increasing the inter-cell frequency offset (Alternate Adjacent Channel (AACh) rather than Adjacent Channel (ACh)) only slightly improves average cell S value. These values are almost constant for various AP offsets (from 5 m to 37.5 m) for both cells 1 and 2, achieving respectively 50% and 55% of average S value per user (cell level) (cell 1 being slightly more loaded than cell 2).

However, for 1 m AP offset both cells are almost fully overlapping (2 m of inter-APs distance), thus maximizing the inter-cell interference. The result is a sharp reduction of S for any channel offset, achieving worse efficiency than with a single AP, even though the load per cell is only half of the single AP case. Cells 1 and 2 are equally loaded, however, cell 1 achieves 2% (ACh) and 30% (AACh) of average S, whereas, cell 2 achieves 22% (ACh) and 50% (AACh) of average S. The reason for such a difference is due to each users' distribution.

The channels are insufficiently isolated and overlapping cells are not synchronized. Accordingly, inter-cell interference can prevent the correct Rx of a frame UL and/or DL, for example a CTS frame, thus reducing the performance of a cell. The result can be a collision with a DATA frame, wasting a lot of time for the entire cell even though only a single user missed a single CTS frame. This problem is recurrent in some configurations. Another consequence of inter-cell interference is a wrong estimation of the medium state making a user freeze his BC decrement, thus increasing his waiting time.

For a high level of inter-cell interference (APs too close), the channel offset is not sufficient to cope for it, leading to worse efficiency than with a single AP. The decrease in the efficiency of the channel reservation and data transmission will be reflected in the goodput performance presented hereafter.

6.12.3.6 Average Goodput per User

The higher the number of contending users in a 802.11a cell, the lower the average goodput per user. Accordingly, maximizing the AP coverage (high number of users attached) is opposed to maximizing performance at the user level (more users having higher transmit modes) as discussed in section 6.12.3.4 for a single AP.

To combine both, maximum coverage and maximum performance, another AP was added. However, inter-cell interference appeared. All average goodput values per user for 1 AP and 2 APs converge to the almost same value for the maximum inter-AP distance, all cells having almost the same number of users attached. Increasing the frequency separation between APs is always beneficial for the performance, unless APs are geographically too close (2 m inter-AP distance). In this last case, cells and network performance are worst than with a single AP. In the "5 m AP offset and 2 APs" case, cells 1 and 2 have equal load. However, in the ACh case cell 2 performance is twice the one of cell 1, whereas in the AACh case cell 2 performance is almost the same as cell 1. For a given load and a given inter-AP distance, a greater channel isolation leads to better performance for both cells. But for greater values of AP offsets cell 2 always achieves better goodput than cell 1 because of a lighter load and of each cell's distribution of

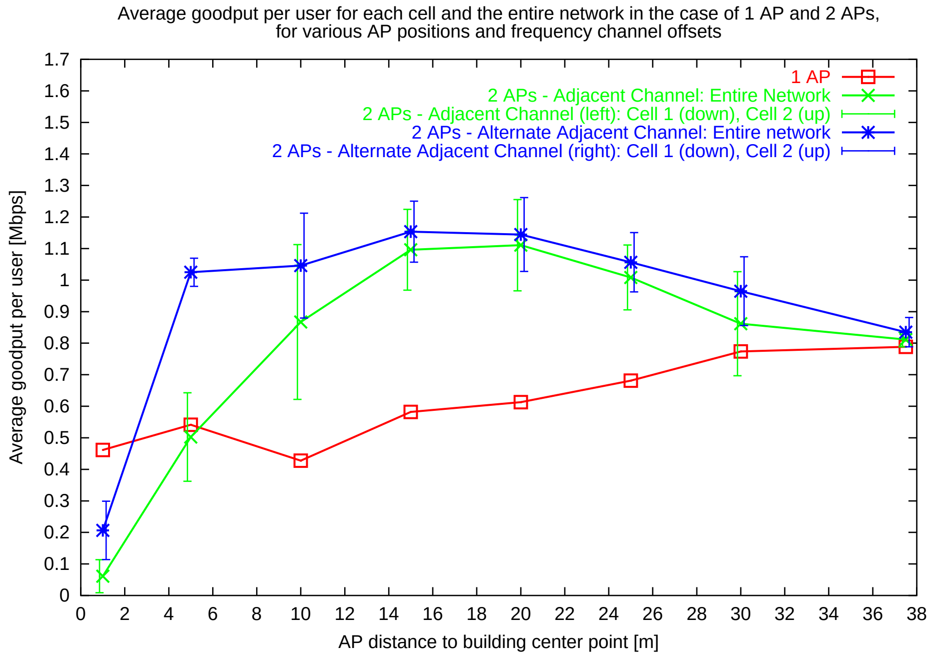


Figure 6.18: Average Goodput per User

users.

6.12.3.7 Aggregated Goodput

Figure 6.19 shows the aggregated goodput values for a single AP and two APs (cell and network levels).

For a single AP, the aggregated goodput (total available bandwidth) is almost constant whatever the AP offset and cell load. However, if two APs are placed too close (AP offset of 1m), whatever their channel offset, the network aggregated goodput is lower than with only a single AP, for the same total number of attached users. Channel offset is unable to cope for inter-cell interference. Then, when each AP offset is 5 m, both cells support in total more users than a single AP and achieve the same aggregated goodput (ACh case), and twice of it (AACh case). The combination of both channel separation and AP geographical distance is necessary to effectively separate overlapping cells. In any case, the higher the channel separation, the higher the performance. Then, as the inter-AP distance increases, the aggregated goodput for two APs network is always at least twice the values for the single AP network.

6.12.3.8 Conclusions and Future Work

Our simulation results highlight the importance of considering inter-AP geographical and channel separation when deploying overlapping cells, to effectively increase the goodput per area. Not taking it into account can become very detrimental to the entire network performance.

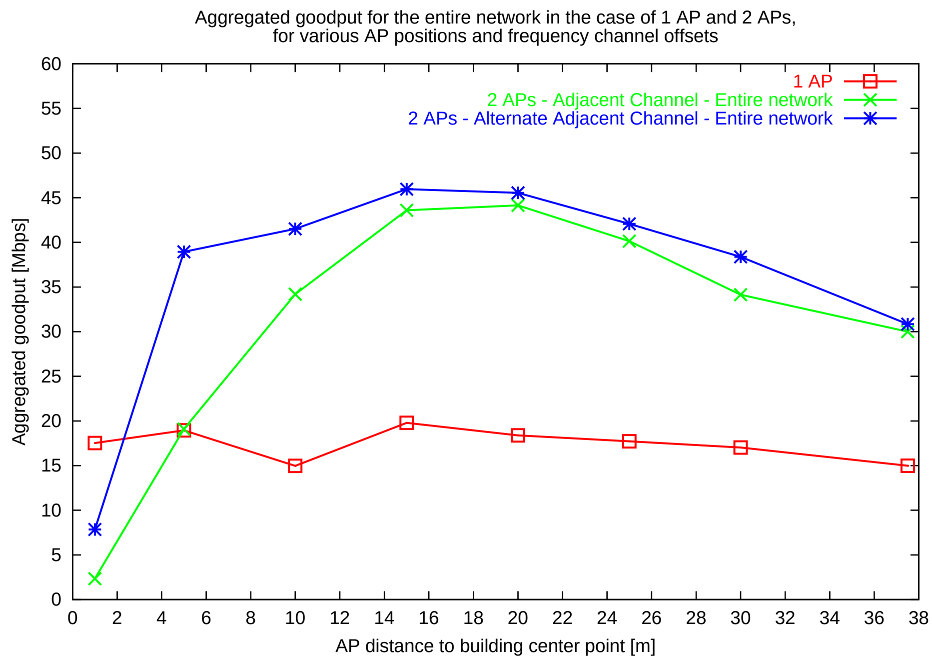


Figure 6.19: Aggregated Goodput

Also, cell topology (users' distribution and AP relative positions) plays an important role in generating inter-cell interference. Load balancing algorithms could help improving performance, for example, triggering handover for users (costly in terms of generated interference) to another cell.

The impact of combining multiple services, as well as inhomogeneous load distributions on the inter-cell interference should be studied. Adding a power control algorithm and a more realistic channel model (shadowing effect) could bring as well additional interesting results.

6.12.4 IEEE 802.11a Study Conclusions

Due to the shared medium as well as the overhead introduced by messaging, the performance achieved by a single user is really lower than the claimed PHY rate (which cannot be achieved anyway). In addition, other weaknesses of the MAC protocol are present when one or several cells are deployed in the same area. Such issues are spatially unfair channel access (due to capture effects and hidden nodes), inter-cell interference (due to insufficient channel isolation and no cell synchronization). Some improvements of the MAC protocol are needed to effectively get the best out of the 802.11a technology.

6.12.5 Power Control Algorithm

To reduce the inter-cell interference effects and bring more fairness into the MAC protocol, we have implemented a power control algorithm in the 802.11a model. Including a power control algorithm will be included in the 802.11h standard currently

6.12.5.1 Scenario 5

This scenario consists of several users in a single isolated cell, exactly as in scenario 2 (see section 6.10.2) augmented by a power control algorithm. The only implemented user distribution is uniform. The scenario objective is to study if the implemented power control algorithm can improve single cell performance and reduce fairness issues.

The implemented power control algorithm is fairly simple: each STA adapts its transmission power to reach the AP (reporting the RSS from the user) plus a confidence margin of 3 dB. This adapted power takes into account the frame mode for the destination (the AP) to be able to successfully decode the frame (sufficient $\frac{C}{I}$ margin). Accordingly it is based on the RSS and does not take into account any interference level at the receiver.

Figure 6.20 presents a comparison with and without power control when 40 users are attached to a cell.

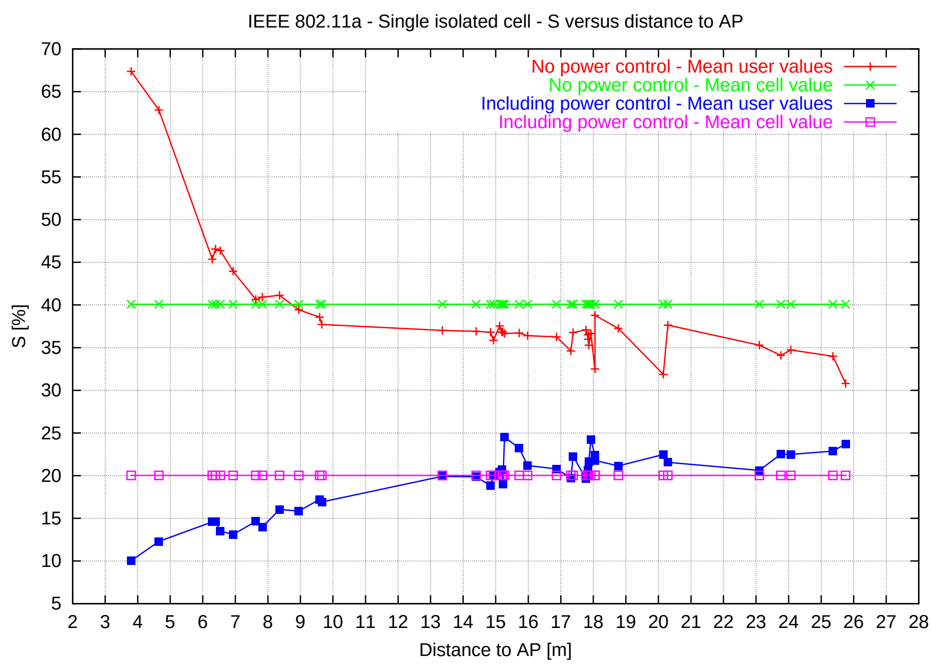


Figure 6.20: Single Cell - S Versus Distance to AP

Reducing the users' transmit range (due to transmit power decrease) increases the number of hidden nodes, thus increasing the number of collisions. In the case of an isolated cell, the results for the power control case demonstrate degraded performance at the user and cell level as shown in figure 1. Let us define S the measure of the combined efficiency of both channel access and data transmission phases expressed in percentage. S is studied at cell and user levels and reflects the data transmission success. As seen in figure 6.20, S is reduced by 2 times for a cell with 40 attached users, when using the power control. At the user level, the difference reaches up to 6.5 times for the closest users to the AP. When no power control is used, the users close to the AP are favored, leading to fairness management issues. For the same locations (uniform distribution), same load, same data payload, adding our power control algorithm degrades the performance, especially for the users close to the AP.

An advantage of the power control is that it reduces the capture effects as all the users would provide the AP with the same uplink RSS. It also reduces the transmitted power level outside the cell, thus potentially reducing inter-cell interference.

As a conclusion, there is an importance of appropriately adapting the power control algorithm with the protocol used. Importance of the cell topology and real impact of the users distributions on the cell performance creating fairness issues, in favor of far users from the AP.

6.12.5.2 Scenario 6

This scenario consists of several users in several cells, exactly as in scenario 4 (see section 6.10.4) augmented by a power control algorithm. The only implemented user distribution is uniform. The scenario objective is to study if the implemented power control algorithm can improve overall network performance. The implemented power control algorithm is the same as in section 6.12.5.1.

In a network of multiple 802.11a neighboring cells, inter-cell interference can limit the maximum performance in some configurations. Indeed, a lack of synchronization between several neighboring cells (each cell operating independently) combined with an imperfect channel isolation can potentially lead to inter-cell interference. It can result in un-decoded packets at the receiver due to the combination of the wanted signal with the interfering signals from other cells. Thus, reducing the level of radio signals generated outside a cell (thanks to power control for instance) is beneficial for surrounding cells and in turn can potentially improve the overall network performance. We showed that for a network of four neighboring cells, the considered power control algorithm is unable to improve cell performance (looking at goodput and S parameters) by reducing inter-cell interference (tested for various channel offsets).

As a conclusion, our studies highlight the importance of considering the 802.11 MAC protocol specificities with an appropriate power control algorithm to effectively improve user and cell performance. Potentially, power control mechanisms taking into account users' distribution and access point relative location shall be considered.

6.13 Conclusion

In this chapter, we have presented the OFDM technique, its advantages and challenges. It is a very promising technique, as shown by its adoption by several key emerging standards. However, IEEE 802.11a/h standards are using OFDM PHY layer but the allocated spectrum band is decomposed into several channels, each one corresponding to a single cell. In each cell, at any time, only a single user is allowed to acquire the channel for DATA transmission.

Our work concentrates on the use of OFDMA as a framework to reach a more dynamic spectrum allocation between a set of users. We believe that the swarm intelligence bottom-up approach implemented for UL SC allocation is very promising, as shown in the next chapter 7.

Chapter 7

Collective UL OFDMA Solution For Performance Improvement

7.1 List of Notations for this Chapter

Table E.4 presents the notations used in this chapter.

Note that these notations are further explained in the remaining of this chapter.

7.2 Introduction

The current tendency is to use the allocated spectrum bands more dynamically. One day, the entire spectrum bands might be considered as a huge common spectrum pool (each system operating in a part of the spectrum better suited to its capabilities) and might use the spectrum only when needed, with a dynamic spectrum right-of-use limited to the time-space area of use. New standards are being launched and able to make a more flexible use of the radio resource. Examples include the IEEE 802.11y standard (July 2005) for its capability to coexist with other IEEE 802.11 systems [51].

We strongly believe that future technological progresses will reinforce the relevance of using an OFDM PHY layer for wireless communications, especially in an OFDMA context. Indeed, OFDM enables to opportunistically occupy some identified unused spectrum bands ("spectrum holes") for operations by another system than the intended ones. Accordingly, cognitive radios could take advantage of the good spectrum granularity offered by OFDM to use spectrum more efficiently [29] [30].

In this chapter we start with an OFDMA description. OFDMA allocation can be either static or dynamic and organized (collaborative). Then, we present several studies showing the interest of an UL distributed collaborative spectrum reservation mechanism compared to the selfish one

Notation	Description
α, β	Parameters controlling $f(x)$ function's shape
$\alpha_{ij}(t_1, t_2)$	Service time (for DATA transmission) of node i on SC j between t_1 and t_2.
α_{Fair}, β_{Fair}	Parameters controlling f_{ij} function's shape (used for time-based fairness)
α_s, β_s	Parameters controlling p_{ij} function's shape
ϵ	Parameter used to find out x_l such that $f(x_l) = \varphi - \epsilon$ and x_u such that $f(x_u) = -\varphi + \epsilon$
γ_{ij}	Signal to Noise Ratio (SNR) value of node i on SC j
γ_{max}, γ_{min}	Respectively max. and min. achievable SNR values in the system
θ_{ij}	Threshold value in our algorithm
θ_{init}	Initial threshold value after reset
θ_{max}, θ_{min}	Respectively maximum and minimum possible threshold value
ξ	Parameter controlling the level of time-based fairness in the system
ϕ_i	Weight of flow i
φ	Parameter controlling the maximum $\Delta\theta_{ij}$ variation per negotiation time slot
$\Delta\theta_{ij}$	Threshold variation in our algorithm
Δc	Range of capacity values for a scenario
Δx	Range of x values for a scenario
a, b	Parameters to control the transformation of c_{ij} into T_{ij}
a_{ij}	Binary value used to indicate whether or not node i has selected SC j
c_{ij}	Shannon capacity achievable by node i on SC j
c_{max}, c_{min}	Respectively max. and min. achievable capacity values in the system
$c_{max\ j}$, $c_{min\ j}$	Respectively max. and min. received c_{ij} values at the central node for SC j
c_l, c_u	Respectively lowest and highest achievable capacity value for a scenario
$f(i, j, k, x)$	Function controlling the amount of time-based fairness in the system
$f(x)$	Our algorithm's thresholding function (modified Fermi-Dirac distribution)
i, j	Respectively index for the nodes and for the SCs
n_i	Number of allocated SCs to node i
n_{max}, n_{min}	Respectively max. and min. simultaneous number of SCs used per active node
p_{ij}	Probability that an agent i performs its task j
q	Quantification step
$q(c_{ij})$	Quantified c_{ij} value
x_{ij}	Result of the transformation of c_{ij} values on an x scale
x_l, x_u	Respectively transformed c_l and c_u values
D	Index of the number of transmitted DATA frames during a simulation
D_{ij}	Number of times node i used SC j
E	Number of elements used to quantify c_{ij} values
F_j	Jain's fairness index on SC j
K	Total number of DATA frames transmitted per simulation
$K_i(p)$	Credit counter of flow i at time p
$L_i(p)$	Head-of-line packet length for flow i at time p
M	Number of nodes
N	Number of available SCs to allocate
T_D	Duration of the data phase
T_F	Duration of a frame including negotiation and data phases
T_{ij}	UL burst message duration
T_N	Duration of the SCs negotiation phases
T_{DL}, T_{UL}	Respectively DL and UL negotiation time slots duration
$T_{timeout}$	Timeout duration for the negotiation phase
$U_i(p)$	Estimated cost for node i to transmit the p^{th} packet
V	Value to compare several nodes in the centralized time-based fairness algorithm

Table 7.1: List of Notations for this Chapter

implemented in the IEEE 802.11 standard. We develop an intelligent and adaptive mechanism based on the swarm intelligence meta-heuristic (refer to chapter 5) in which all the contending nodes collaborate to achieve the best system UL sum capacity.

7.3 OFDMA Context

OFDMA, also referred to as multi-user-OFDM (multi-user extension of OFDM), is being considered as a modulation and multiple access method. Indeed, in current OFDM systems, only a single node can transmit on all of the subcarriers at any given time, and time division or frequency division multiple access is employed to support multiple nodes. However, a major drawback with this static multiple access scheme is that it does not exploit the different channel conditions seen by each node, to modify accordingly the SC allocation. OFDMA, on the other hand, allows multiple nodes to transmit simultaneously on different subcarriers per OFDM symbol. Figure 7.1 shows an example of an UL SC allocation between 3 nodes in a cell. The variations of the amplitudes per SC make each one more suitable for some nodes compared to others. The nodes share the system SCs to get the most out of this allocation. On figure 7.1, the colored spectrum parts indicate the best SC amplitude for each node, while the white blocks correspond to idle SCs. As seen, the final reached allocation between nodes is interleaved, resulting in not necessarily contiguous SCs. Also, the opportunistic aspect is illustrated by the fact that each terminal tends to use the best SCs for its use.

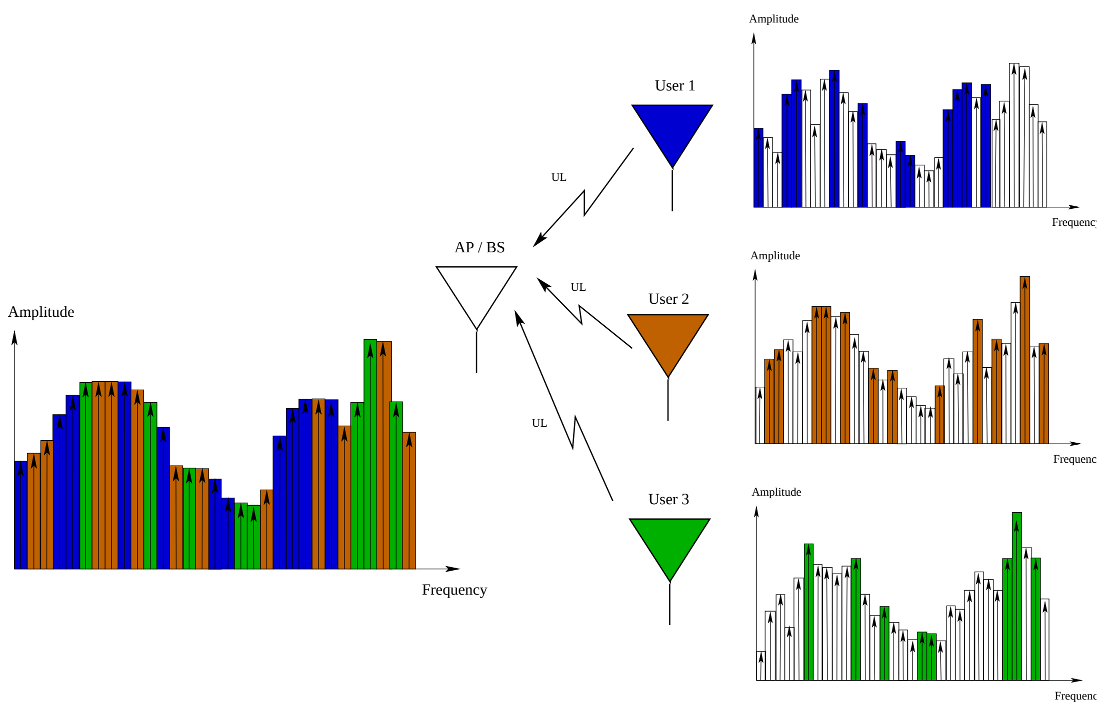


Figure 7.1: UL OFDMA Example with 3 Nodes

The minimum granularity for a radio resource will be the use of an OFDM SC for a given period of time for data transmission. An OFDM-based system divides a high-speed serial information signal into multiple lower-speed sub-signals transmitted simultaneously in parallel using different frequencies. OFDMA provides a very good slicing of the spectrum, making it more efficient for each node to precisely identify his most relevant set of SCs: use only the best required SCs for one's use.

Also, OFDMA offers a better opportunity for a secondary spectrum usage than with spread spectrum since it enables to occupy non contiguous spectrum holes using OFDM. In other words, it provides a better dynamic reuse [39] of the spectrum resource between multiple nodes, in a cognitive radio context.

7.4 Collective UL OFDMA

Our approach studies the mechanisms leading to an UL OFDMA arrangement of the nodes on the system's SCs. We assume that for the data transmission, the final SC allocation is orthogonal: a SC is assigned to no more than one node simultaneously. However, several nodes from the same cell might be transmitting simultaneously but using different SCs. Using the swarm intelligence meta-heuristic, we provide a distributed algorithm ran by each node and able to reach collectively the final best allocation for the system.

No inter-SC interference exists if SCs are kept orthogonal. Accordingly, to benefit from the good spectrum packing provided by the OFDM technique, we must maintain the inter-node synchronization. As shown in figure 7.1, we consider only the UL SC allocation problem. In this case, all the UL messages intended for the radio access infrastructure equipment, shall be simultaneously received at this destination. Note that, propagation delay corrections can be introduced, thus leading to not simultaneously transmitted UL messages. However, when using small cell sizes, as it is the case for WLAN systems, we can consider the propagation delay as negligible. In such a case, the UL messages have to be simultaneously sent in order to be simultaneously received at the radio access infrastructure equipment.

After receiving a broadcast message from the radio access infrastructure equipment, each node is able to estimate its SINR level on each cell's SC. Then, all the contending nodes perform a parallel UL reservation on the intended SCs. Several nodes might have requested the same SC. An iterative process based on a collective intelligence algorithm helps in finally reaching an orthogonal SC allocation between all the nodes in UL. Note that using our approach, the intelligence and the SC allocation control is located in the user's device and not anymore in the radio access infrastructure equipment. Of much importance, note that our dynamic allocation algorithm supports input data variations during the optimization phase, leading to a real-time optimization algorithm.

7.5 Swarm Intelligence Modeling Challenge within an OFDMA Context

7.5.1 Introduction

The algorithm presented hereafter is implemented at the MAC layer, but a technique for parallel reservation using the OFDM PHY layer is also modeled, leading to a cross-layer optimization algorithm for intelligent and dynamic spectrum use.

The following sections present the results from the modeling challenge, to go from a world of social insects to a wireless world. We highlight the interest of a bottom-up approach in a multi-user (OFDMA) spectrum allocation context, while taking care of reducing the coordination overhead. There is a benefit in using a coordinated distributed approach as shown in chapter 5.

First note that, in the following studies, "users" are called "nodes". Also, let us define as "central node" the infrastructure equipment (AP), and as "nodes" the user equipments (MT). Finally, we refer to "sum capacity" as the sum of the capacity per allocated SC over all the system SCs.

Licensed and unlicensed systems are covered by our scenarios. It includes spectrum pooling (an unlicensed spectrum band is shared by all the cells using a WLAN system) and spectrum sharing (primary and secondary users of a spectrum band).

We are in a context where cognitive radios can cooperatively negotiate for spectrum access with others. Our proposed algorithm, achieves a dynamic and flexible UL OFDMA SC allocation by exploiting multi-user diversity. Nodes cooperate using a distributed approach, to maximize the cell sum capacity. As multi-user diversity is taken into account in the spectrum allocation, the system sum capacity increases with the number of nodes [59].

Also of great interest, our algorithm can be controlled by few global parameters, to provide a network operator with an easier tuning of its network QoS. Thus, it reduces the burden and cost associated with spectrum network planning, maintenance and upgrade.

Our method is particularly interesting when there are many nodes in the cell (without the need for the algorithm to know the number of contending nodes), as well as if the final allocation can be reached faster than a centralized approach for the same resulting throughput. In addition, we do not require (as it is the case in a centralized approach) any prior TDMA-like frame to tell each node when to transmit his reservation request. The centralized approach does not offer the same dynamicity in SC reallocation as our approach.

7.5.2 Optimization Problem

The considered optimization problem is the same as presented in section 5.3.2, but more specifically developed in an OFDMA context.

The optimization problem considered can be described as follows (index i refers to the nodes and index j to the SCs). Given N available SCs to allocate among M nodes, each node i having a SNR value γ_{ij} on SC j, must find the best UL allocation of nodes on SCs, to maximize the cell sum capacity, subject to some constraints described hereafter. Let us define c_{ij} as the Shannon capacity achievable by node i on SC j (for a unit bandwidth) which is: $c_{ij} = log_2(1+\gamma_{ij})$. Let us define a_{ij} as the value used to indicate whether or not node i has selected SC j, with $a_{ij} = \{0; 1\}$. We are interested in the case of continuous coding and modulation, independent between SCs. If we define n_i as the number of allocated SCs to node i, an "allocated node" has $n_i \neq 0$.

Using all these notations, the optimization problem can be formulated in mathematical terms as follows. The objective of the optimization problem is to:

$$\max \sum_i \sum_j a_{ij} c_{ij}$$
$$\text{s.t.:} \begin{cases} (1) & \{n_{min}, n_{max}\} \in [0; N] \\ (2) & \forall i,\ n_i = 0 \Leftrightarrow \sum_j a_{ij} = 0 \\ (3) & \forall i,\ n_i \neq 0 \Leftrightarrow \\ & n_{min} \leq \sum_j a_{ij} \leq n_{max} \\ (4) & \forall j,\ \sum_i a_{ij} \leq 1 \end{cases} \tag{7.1}$$

In other words, the objective is to find the final system allocation vector maximizing the system sum capacity, assuming all the constraints are respected.

Figure 7.2 presents two simple examples of this optimization problem's resolution. Several nodes need to be allocated SCs given input capacity values c_{ij}. In each table of figure 7.2, the green elements represent the optimal solution for the problem.

Example 1

Table of cij values and final BEST allocation for : $M = 4$, $N = 2$, $n_{min} = 1$, $n_{max} = 1$		SCs (N)	
		1	2
Users (M)	1	1	4
	2	9	3
	3	2	7
	4	3	6

Example 2

Table of cij values and final BEST allocation for : $M = 4$, $N = 3$, $n_{min} = 1$, $n_{max} = 2$		SCs (N)		
		1	2	3
Users (M)	1	1	2	4
	2	9	3	3
	3	2	10	7
	4	3	5	6

Figure 7.2: Example of Optimization Problem Resolution

In a spectrum sharing or pooling context, a cell spectrum resource is a variable quantity over time. Regularly, the central node updates the amount and position of the cell acquired SCs. This operation is based on the knowledge, at the central node as well as at the nodes, of the channel statistics for all the SCs in the band. To exchange spectrum resource (augmentation or reduction), each central node can negotiate with other central nodes. Then, the central node broadcasts within the cell the positions of the new set of SCs to use. N is the spectrum size acquired/negotiated by the central node. The values of n_{min} and n_{max} are controlled by the central node and can be tuned according to the entire available spectrum resource, channel statistics, etc.

We assume there exists link adaptation capabilities per SC such that nodes with good channel conditions will have a higher capacity than those with worst channel conditions. The n_i constraints (refer to equation E.5) are requirements on the spectrum budget per allocated node with an objective to maintain a certain level of sum capacity per allocated node: n_{min} ensures a minimum sum capacity per allocated node, whereas n_{max} limits the over dimensioning of the necessary spectrum resource to run a given service in the best channel conditions. As such, each allocated node is ensured to have at least n_{min} SCs corresponding to a minimum sum capacity to run a given service. However, the algorithm does not try to equalize the capacity between the nodes.

As the considered optimization problem is NP-complete, there exists no algorithm to find the optimal solution in all cases in a polynomial time. Accordingly, to find a "good" sub-optimal solution we use an heuristic. In designing the algorithm, the compromise between the quality of the found solution (distance to the optimal solution) and the speed of convergence (time to obtain it) is taken into account.

7.5.3 Distributed Optimization

This work is not a theoretical study on distributed optimization but rather it presents a flexible algorithm with an emphasis put on the realistic aspects of its implementation, the quality of the final allocation and the speed of convergence.

The envisaged scenarios consist in a single cell with a central node and several geographically distributed nodes in the cell. The nodes have full responsibility in choosing their set of SCs to use for transmission, whereas the central node does not influence the nodes' allocation. In that sense, the algorithm is distributed and the power of decision is spread between the nodes. The optimization is achieved by cooperation between geographically separated nodes. Such a distributed approach is of particular interest when the number of attached nodes becomes large.

We assume a low mobility of the nodes, a known channel at the receiver, a synchronized UL reception of the messages and a symmetric channel with not too rapid variations. This is reasonable for WLAN systems with small cell sizes.

While designing our algorithm, we have limited all losses due to coordination overhead: only the necessary information is transmitted to the central node. A SC reallocation is achieved (if necessary) for every data packet, but the frequency of SC reallocation can be varied by the operator, depending on the rate of change of: the channel, the availability of new SCs, the cell traffic. The optimization algorithm consists in an opportunistic spectrum access at the node level.

The time is decomposed into frames of variable duration $T_F = T_N + T_D$. Each frame comprises a negotiation part (variable duration T_N) and a data part (fixed duration T_D). The negotiation part consists in searching for the optimized allocation, whereas the data part consists in the transmission of data frames. We assume that SNR values do not change during T_F. Figure E.6 presents the timeframe context of the optimization algorithm and shows an example of SC negotiation in a cell composed of 3 nodes.

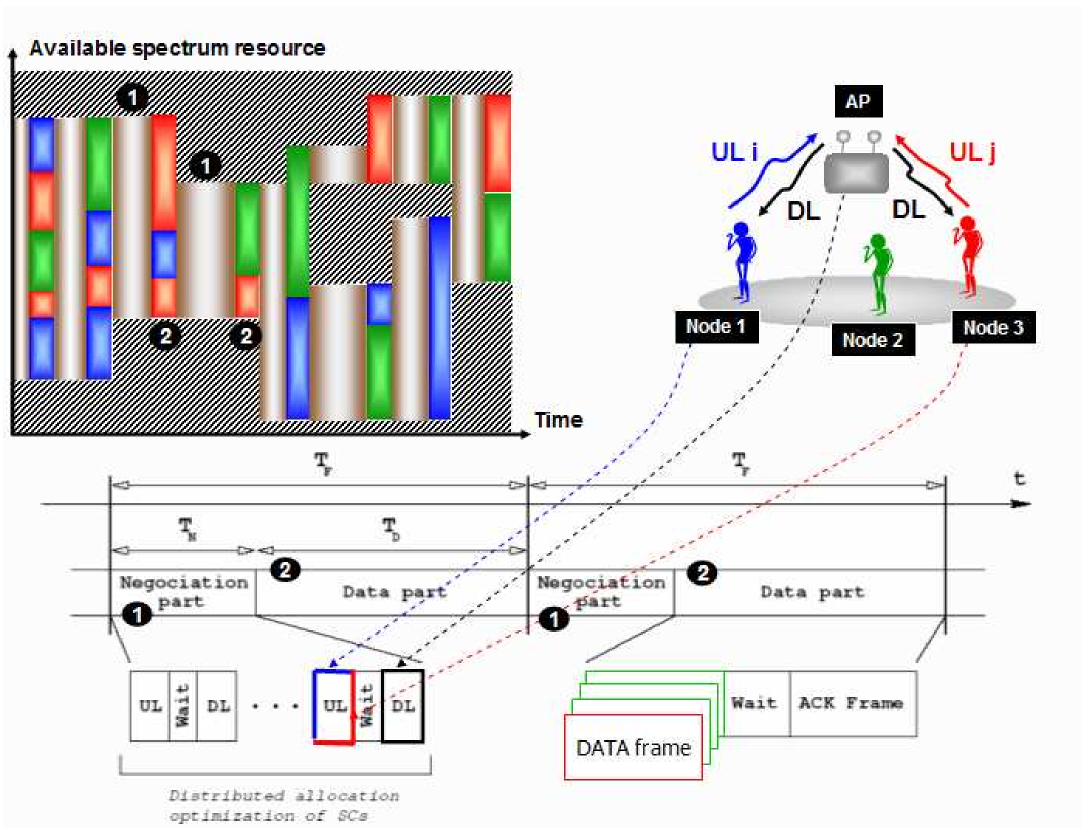


Figure 7.3: Description of the Optimization Algorithm Parts

As seen on figure E.6, the negotiation part consists itself in a succession phases:

- UL time slot: UL messages are simultaneously sent by the contending nodes and intended for the central node. Each node contending for a given SC transmits to the central node an UL burst message on that SC.
- DL time slot: Then, the central node broadcasts in the cell a DL feedback message (using a dedicated channel common to all SCs) solely containing the minimum and maximum capacity value found on each SC.

These two phases are repeated during the negotiation part until the allocation is finished. Then, the data part takes place, during which a data frame can be transmitted by the nodes using their acquired (ON status) SCs.

Distributed optimization methods possess many advantages such as: scalability, robustness, adaptability to varying conditions. Also, distributed agents can be equipped with learning capabilities such as Reinforcement Learning (RL) (see section E.8.3.4). Each agent runs the algorithm separately and by interacting with other agents, they all contribute to create a globally optimal solution, without the need for a central controller.

Solving the resource allocation problem with a distributed approach for the optimization has the advantage of parallelizing the optimization task among all the nodes. It reduces the complexity per node. Our algorithm is scalable with the number of nodes and able, without any prior knowledge of the number of nodes, to dynamically adapt to all the following scenarios: $N = M$, $N < M$ and $N > M$, as well as to any n_{min} and n_{max} values.

The next section develops the wireless communication mechanism used to achieve an efficient distributed control.

7.5.4 Simulation Propagation Model

This section describes the simulation model used for all the simulations concerning our algorithm. We use a simple propagation model with the main objective to concentrate on showing the interesting properties of the algorithm rather than simulating realistic channel models.

For each node, for each SC, SNR values are randomly drawn from a uniform distribution. The total range of SNR values is determined as follows:

- Minimum SNR value (γ_{min}): obtained as the worst possible conditions of reception for the messages. In case several transmission modes exist (each one being characterized by its sensitivity level, coding rate, modulation), γ_{min} is calculated per SC with the minimum signal level enabling to correctly receive messages using the minimum Tx mode.
- Maximum SNR value (γ_{max}): obtained as the best possible conditions of reception for the messages, i.e. the minimum coupling loss. γ_{max} is the SNR value corresponding to the

closest possible distance between a node and the central node (thus, in Line Of Sight (LOS) conditions). γ_{max} is calculated per SC.

Note that γ_{min} and γ_{max} are parameters characterizing the system, they are not deduced by the nodes.

Then c_{ij} is deduced according to Shannon capacity as follows:

$$c_{ij} = log_2(1 + \gamma_{ij}) \tag{7.2}$$

Nodes then try to optimize their total UL capacity starting from the c_{ij} values.

Every time a channel change occurs, all the nodes update all their SNR values for all their SCs. No correlation between the nodes is modeled. Also, for a given node, his SCs have no correlation between them.

Depending on the studied scenarions, the channel can be either:

- Not changing during several negotiation phases
- Not changing during a negotiation phase but after a DATA Tx phase
- Changing during negotiation phases and not during the DATA Tx phase

Note that we mimic the fading per SC by using the following model: when SNR values change, they change according to a uniform distribution (per SC).

When no live results are provided at the end of a simulation, the results are averaged over several loops of simulation conducted under the same conditions. In such a case, the standard deviation of the results vary depending on the scenario.

7.5.5 Description and Modeling of the Heuristic

The objective of this part is to described how the distributed optimal allocation of SCs is reached among the nodes.

At the end of the optimization phase, there is only a single specialized node per SC, but a node can be specialized in several SCs. During the optimization process there is both an inter-node negotiation (to respect constraint 4) and an intra-node negotiation (constraints 2 and 3). Indeed, each node would want to use all the n_{max} SCs but by means of local interactions it must be optimally shared among all the nodes to globally maximize the cell UL sum capacity.

A node contains (1) a constraint controller and (2) a set of N cooperating agents (one by SC). The constraint controller ensures that the constraints in number of active SCs per node are always ensured. On the contrary, the main objective of an agent is to stay active (ON) at the end of the optimization process. Agents interact internally (with other SC-agents within the same node) and externally (with agents representing the same SC index but from other nodes). The rules of interactions of agents between them, are build to reach a global optimum by using local interactions, as in auto-organization. While they interact, agents exchange information about capacity values (see section 7.5.5.3).

Figure 7.4 depicts the context of the internal and external interactions.

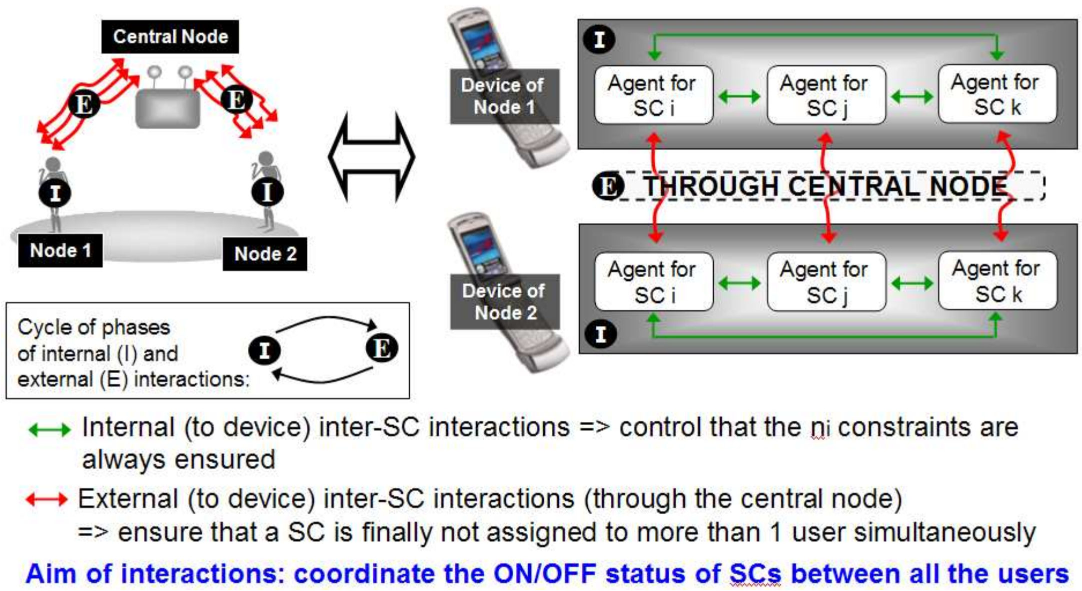


Figure 7.4: Internal and External Interactions

To become stronger and have better chances to stay active, agents organize in groups (within each node) and try to find the best other members of their group, for them to be ON at the end. Groups are an aggregation of each agent force: the group capacity is the sum of the capacities from all agents engaged in this group. The weakest agents are left apart and groups are reconstructed until the end of the optimization process. Groups are organized and re-organized dynamically and autonomously to reach the best final configuration for the system. Thus, groups have a variable size over time.

We assume that a node is able to simultaneously perform n_{max} tasks out of a total of N tasks. An agent is able to perform only a single task, and it can either perform it (ON status for that SC) or not.

In our model, an agent's threshold value changes over time and its value depends on (1) its previous history, (2) its ability to perform its task better than the agents from the same node (intra-node contests) and (3) its ability to perform its task better than the agents from other nodes for that same task (inter-node contests for the same SC).

The amplitude of change of an agent's activity threshold is variable and is a function of the agent's force compared to others agents for this task. The threshold value modifies the probability of an agent to perform its task in the future, such that: a low threshold leads to a high probability to perform its task, whereas a high threshold means a low probability to perform its task.

In optimization, there is an important trade-off between the exploitation of the current solution and the exploration of new solutions. The chosen heuristic, owing to its variable threshold model, adapts well to a dynamic environment and leads to faster convergence (but not premature) than other heuristics. It is also very efficient in avoiding local optima. It uses a simple RL algorithm described hereafter. In addition, our model is able to adapt to a large set of constraints. As a result, the same algorithm solves the problem of SC allocation for a range of SCs per node from 1 up to N. The particular case where a single node can capture all N available SCs can be interpreted as a new allocation method for a single channel system.

The next sections will describe more in details the internal and external operations taking place between nodes.

7.5.5.1 General Presentation

To solve the considered optimization problem we implemented a meta-heuristic inspired from a model of division of labor used by social insects such as ants, wasps, etc. This heuristic comes from the domain of biology (studies of social insects) then modeled by the domain of artificial intelligence, improved and successfully applied to solve real world optimization problems, using a set of cooperating agents. This meta-heuristic was proven to be very efficient to optimize real life problems of dynamic and distributed tasks allocation in the industry [60], [61].

This method was found to give better results than a market-based approach for the dynamic and distributed task allocation problem [62]. Clearly the obtained results on well-known optimization problems are very good, and the underlying theory is still in its infancy (theory of cooperating agents). All the agents use the same finite set of simple rules to interact with each other. The result of this cooperation is an auto-organization at the system level. A modeling challenge is to find the appropriate set of local rules that will globally create the desired behavior.

We managed to successfully adapt and use this heuristic in our problem with good results, thus showing its great potential for optimization and control of cooperating agents.

The method is inspired from social insects where the ability of an agent to perform a task depends on its ability for this task, as well as the number and ability of other agents willing to perform this task. Thus, agents "specialize" over time in the tasks where they are best fitted for, by means of a variable threshold, as well as by inter-agent conflicts (dominance contests).

This model of variable threshold brings robustness to the system. Indeed, all the individuals together are able to appropriately adapt and find an optimum solution to a problem even under very dynamic environments. An improved version [63] uses a moving threshold with time of use. Agents receive stimuli from tasks in order to perform a task. They effectively perform the task according to a given probability, function of their threshold regarding this type of task.

The more an agent becomes specialized in a given type of task, the greater the probability to engage in doing it for future stimuli of this type. Each agent maintains and updates over time a threshold per type of job to be done.

Figure 7.5 summarizes the similarities and differences in the analogies between the social insects and the wireless agents we have modeled.

Aspects	Biological World	Wireless World
Colony	Group of all social insects in a swarm	Swarm of all agents in a cell
Individual	Each ant, wasp, etc of a colony	A wireless agent (1 per SC), thus N per user's device
Task	Action to perform for the colony (foraging, construction, etc)	For an agent: spectrum allocation to perform for the system (and the user) - ON = contending for its SC, - OFF = not contending for its SC
Stimulus	Parameter k. Generated to engage others in an action: high needs leads to high stimuli	N/A
Threshold	Parameter Θ: - Fixed for ants - Variable for wasps Indicates the level of specialization of the individual for a task	Parameter Θ: variable. It incorporates the history of past actions and varies as a function of the agent fitness compared to other agents of other users for that SC
Proba transition function	Proba to engage in this task: $P = k^2 / (k^2 + \Theta^2)$	Proba of this SC to be ON: $P = 1 / (1 + \alpha e^{\beta\Theta})$

Figure 7.5: Social Insects vs Wireless Agents

The next sections describe more in details our model of the wireless agents. In particular, the next sections will explain how and where the variable threshold is used, as well as how we managed not to use the stimulus part. The following sections concentrate on the steps and operations of the negotiation phase, both in the nodes and in the central node:

- Operations internal to the nodes (intra-node), performed within each node's terminal, and starting after the reception of a DL message from the central node: introduced in section 7.5.5.2, and described more in details in section 7.5.5.6.
- Operations external to the nodes (inter-node), whose results are analyzed within the central node's equipment, and starting after the reception of all the transmitted UL burst messages from all the contending nodes: introduced in section 7.5.5.3, and described more in details in section 7.5.5.5.

7.5.5.2 Algorithm Negotiation Steps: Intra-node Operations

This section describes the intra-node operations. These operations refer to the actions that are repeatedly realized within each node during the negotiation phase.

The optimization starts with each node setting to ON its n_{max} SCs with the best capacity values and then, the algorithm tries to improve this solution. Then, several steps are gone through (see below) until all the optimization constraints are enforced, and the found solution is considered as the final one.

At each iteration of the process, the stations are randomly transmitting on one or several SCs. The probability to transmit or not on a SC is a decreasing function of a parameter θ_{ij}.

That parameter is defined for each agent j of each station i, and is updated at each iteration, as a function of the feedback of the central node.

Several nodes may simultaneously transmit on the same SC. Each one has a particular capacity for this SC. The central node is then able to determine on each SC j the greatest capacity $c_{max\ j}$ among all the contending nodes that have transmitted on that SC, thanks to a duration-coding technique (refer to section 7.5.5.3). The maximum capacity values are then broadcasted by the central node (feedback process).

Each agent makes the difference between c_{ij} and $c_{max\ j}$ and computes a variation step $\Delta\theta_{ij}$. The function that defines $\Delta\theta_{ij}$ is rather complex and is detailled in section 7.5.5.6. Each station then computes a new value of θ_{ij} (which is $\Delta\theta_{ij} + \theta_{ij}$) and iterates the transmission process.

Algorithm within each node i:

- **Step 1:** For each agent j, compute the initial values of θ_{ij}.
- **Step 2:**
  - a) Determine the probability of each agent j to be active (ON) for the next UL contention Time Slot (TS) according to :

$$p_{ij} = \frac{1}{1 + \alpha_s e^{\beta_s \theta_{ij}}} \tag{7.3}$$

    where α_s and β_s are two parameters controlling p_{ij} function's shape.
  - b) Determine the activity status (ON or OFF) of each agent j: if p_{ij} is greater than or equal to a randomly drawn value between 0 and 1 then the agent is active (ON), otherwise it is inactive (OFF).
- **Step 3:** Check if the maximum number of active agents is lower than n_{max} (constraint controller) and inactivate (by choosing the n_{max} agents with the highest p_{ij} to be ON) enough active agents until there is less active agents than the maximum n_{max}.
- **Step 4:** Transmit an UL burst message to the central node using only the active SCs for that node i.
- **Step 5:** Wait for the feedback from the central node, consisting in the $c_{max\ j}$ and $c_{min\ j}$ capacity values for all j.
- **Step 6:** For each j, compute $\Delta\theta_{ij}$ (as a function of $c_{max\ j}$ and c_{ij}) the θ_{ij} threshold variation as a result of the newly received $c_{max\ j}$ value per agent j. The computation is rather complex and will be further described in section 7.5.5.6.
- **Step 7:** For each j, update the threshold value θ_{ij} accordingly:

$$\theta_{ij} \leftarrow \theta_{ij} + \Delta\theta_{ij} \tag{7.4}$$

- **Goto Step 2 ...** while the negotiation phase is not completed. If the ending criteria is matched, then stop the negotiation algorithm (i.e. do not go to step 2 but exit the algorithm). The ending criteria is matched whenever in step 5, the central node signals in its

DL message that all the constraints of the optimization problem are met (for the contraints refer to section 7.5.2). In particular, the central node controls that there is at maximum a single transmission (single contender) per SC, by checking that the burst message can be properly decoded. Also, it controls that the number of idle SCs is minimum, otherwise the negotiation continues because more SCs can be acquired by the nodes. The nodes (using their constraint controller) are responsible for verifying that they respect the n_i constraints. The end of the negotiation phase should correspond to an optimum allocation for the system.

The next section explains the principles and the practical implementation of the inter-node communication mechanism, and the conflicts resolution.

7.5.5.3 Algorithm Negotiation Steps: Inter-node Operations

In a distributed system, the following questions have to be answered to exchange information between entities:

1. Question (1): What information to include in the messages?
2. Question (2): How to communicate (inter-node communication mechanism)?

To answer question (1), a simple but efficient duration-coding technique was used: each c_{ij} value is coded in UL as the duration T_{ij} of a burst message (transmitted in UL) such that: $T_{ij} = \frac{a}{c_{ij}} + b$. Thus, the greater c_{ij}, the shorter T_{ij}.

To answer question (2), all contending nodes simultaneously (same transmission start time) use an UL transmission scheme on the SC. Even though figure 7.6 presents an example with a single SC, 3 nodes, $a = 1$, $b = 0$, and fictitious numerical values, the explanations suffer no loss of generality.

As shown in figure 7.6, 3 nodes contend for SC j, each with a different signal level at the receiver and c_{ij} value (refer to the table located at the top of the figure). The left part of the figure indicates the time-coding of c_{ij} into T_{ij}, and the RSS. The right part of the figure shows the resulting signal aggregation at the receiver, with the detection of the minimum and maximum values.

All UL signals aggregate independently per SC. Then, the central node uses the aggregated signals to extract the minimum and maximum coded c_{ij} values. In practice, the central node easily finds the maximum value by detecting the first drop (simple gap detection rule) in the aggregated signals. The duration between the start of the UL reception and the first drop gives the corresponding transmitted value. By detecting the longest message duration, the central node finds the minimum value. As the optimization progresses, only the best values remain. A DL message is sent only after the end of the longest burst message. An advantage of our method is the simplicity of the coding and decoding, simultaneous and parallel exchange of information on the same radio resource, robustness against interference (if constant during the UL time slot) and collisions, and it does not require an orthogonal coding to separate nodes or capacity values.

Node index	cij	Tij	Rx signal strength (RSS) at central node from node i
1	40	1/40	3
2	10	1/10	0.5
3	20	1/20	2

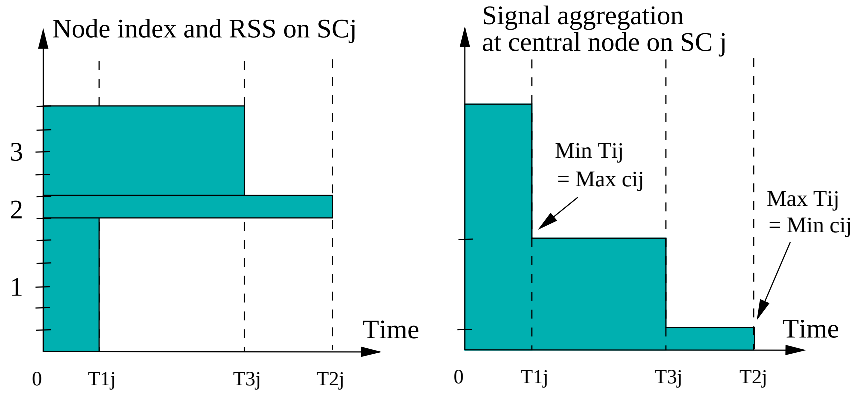


Figure 7.6: Capacity Duration-Coding and Aggregation Mechanism with 3 Nodes on SC j

To coordinate their allocation, the nodes do not communicate directly with each other, but only require to exchange information with the central node. The central node acts as a "filter" extracting the maximum and the minimum values out of all the received values, and sending it back to the nodes for information sharing.

WLAN systems with a distributed control (e.g.: IEEE 802.11 in DCF mode) suffer performance loss due to the hidden node problem. Out of range nodes are said to be hidden nodes. They have no possibility to communicate to avoid UL collisions on a shared channel. Our mechanism of inter-node exchange of information is built to encourage and manage message overlap (aggregation of signals) even between hidden nodes. The consequence is no more collision during the data transmission. Accordingly, the problem of hidden nodes is completely removed and there is no risk of performance degradation due to collisions.

7.5.5.4 A Basic Working Example

This section presents an example of how a SC allocation negotiation between nodes evolves over time. The following scenario consits of $M = 3$ nodes distributed in a cell. They try to cooperatively share $N = 3$ SCs with the following constraints: $n_{min} = 1$ and $n_{max} = 1$. Also, the θ thresholds have the following values: $\theta_{min} = -75$ and $\theta_{max} = 75$.

This example illustrates how the reinforcement process orients the negotiation towards the best solution for the system. p_{ij} illustrates the probabilistic choice achieved by each node on each of its SCs to decide on its status (ON or OFF) for the next negotiation step, based on θ_{ij} as well as on internally enforcing the n_i constraints.

Table 7.2 presents the input capacity values.

Nodes\SCs	**(1)**	**(2)**	**(3)**
(1)	4.54781	7.31514	20.6921
(2)	14.3226	15.9123	9.18369
(3)	5.52536	19.046	19.0553

Table 7.2: Input Capacity Values

Table 7.3 presents the evolution of the collaborative negotiation over time with the θ_{ij} values and the SCs status.

Nb Nego Steps	θ_{ij} **Table**					**SC Status Table**			
	Nodes\SCs	**(1)**	**(2)**	**(3)**		**Nodes\SCs**	**(1)**	**(2)**	**(3)**
	(1)	00.0000	00.0000	00.000		(1)	.	.	$\surd$
1	(2)	00.0000	00.0000	00.000	$\Rightarrow p_{ij} \Rightarrow$	(2)	.	$\surd$	.
	(3)	00.0000	00.0000	00.000		(3)	.	.	$\surd$
	Nodes\SCs	**(1)**	**(2)**	**(3)**		**Nodes\SCs**	**(1)**	**(2)**	**(3)**
	(1)	-0.0052	74.893	-28.234		(1)	.	.	$\surd$
2	(2)	0	0	0	$\Rightarrow p_{ij} \Rightarrow$	(2)	$\surd$	.	.
	(3)	-29.837	-66.865	74.999		(3)	$\surd$	.	.
	Nodes\SCs	**(1)**	**(2)**	**(3)**		**Nodes\SCs**	**(1)**	**(2)**	**(3)**
	(1)	-0.0052	74.893	-28.234		(1)	.	.	$\surd$
3	(2)	-4.0148	-74.992	75	$\Rightarrow p_{ij} \Rightarrow$	(2)	.	$\surd$	.
	(3)	45.163	-75	75		(3)	.	$\surd$	.
	Nodes\SCs	**(1)**	**(2)**	**(3)**		**Nodes\SCs**	**(1)**	**(2)**	**(3)**
	(1)	-0.0052	74.893	-28.234		(1)	.	.	$\surd$
4	(2)	-75	0.00836	75	$\Rightarrow p_{ij} \Rightarrow$	(2)	$\surd$	.	.
	(3)	15.327	-75	75		(3)	.	$\surd$	.

Table 7.3: Time Evolution of the SC Negotiation

Table 7.3 details for each negotiation steps the evolution of both tables of θ_{ij} values and SC status tables. In this example, 4 negotiation steps are necessary to achieve the final allocation. The θ_{ij} values are updated according to the relevance of the choice of a node activating a given SC for Tx.

According to equation 7.12, the higher (resp. the smaller) the θ_{ij} value, the smaller (resp. the higher) the probability of this choice. For each negotiation steps Table 7.3 is read from left to right. In the SC status table, a "." indicates an OFF SC whereas a "$\surd$" indicates an ON SC. The negotiation is continued until reaching a non conflicting allocation between the nodes. All the nodes collaborate to find the best allocation for the system, which is not necessarily the best allocation for the each node taken separately.

This example is interesting because it illustrates the algorithm's behavior on an example easy to verify by hand. In addition, it shows how the system learns over time with all the θ_{ij} values together consisting in the system's memory.

7.5.5.5 Algorithm Negotiation Steps: Inter-node Operations - Variable Quantification Step

We now show the practical implementation aspects of our inter-node communication mechanism. Before reaching the final SC allocation, several UL/DL time slots are usually necessary. The UL and DL negotiation time slots have a respective duration of T_{UL} and T_{DL}. To bound the UL burst messages duration, the c_{ij} values were quantified into a fixed number E of elements. However, the resulting precision is not sufficient to differentiate close values. Accordingly, to increase the precision, we used a variable quantification step per SC with a zoom window. The principle is explained in figure 7.7 (each triangle and cross represents the non-quantified c_{ij} value of node i on SC j).

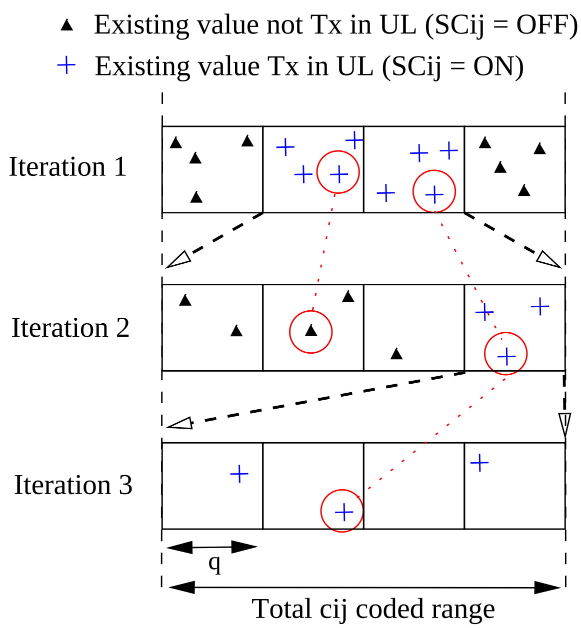


Figure 7.7: Variable Quantification on SC j

Figure 7.7 represents the evolution over time (iteration after iteration) of the range of expressed quantified c_{ij} values (i.e. $q(c_{ij})$ values) in which the final best values will be located. At each iteration, the central node:

1. Collects the UL received quantified extreme values (minimum and maximum) per SC. In figure 7.7, the minimum (resp. maximum) value corresponds at each iteration to the left (resp. right) corner of the more at the left (resp. right) square. A square representing the range of different c_{ij} values having the same quantified value $q(c_{ij})$ due to the quantification step.
2. Adjusts the extreme window limits around these extreme values.
3. Broadcasts the new window limits and quantification step q. Indeed, q is changed after each zoom/dezoom operation.

This variable quantification process uses a zoom window to ensure a shorter negotiation duration, and to concentrate the precision of the quantification only for the current expressed values.

If after a period of time $T_{timeout}$ no final convergence was obtained, the central node imposes to stop the optimization process and forces all the nodes to apply a quick solution. Thus, this technique bounds the negotiation duration. After the timeout was reached, an acceptable solution can be found in two UL/DL time slots.

7.5.5.6 Algorithm Negotiation Steps: Intra-node Operations - More in Details

This section describes more in details some of the intra-node operations presented in section 7.5.5.2.

Firstly, we present more in details how the step 6 (refer to section 7.5.5.2) of the algorithm is performed: the update of the $\Delta\theta_{ij}$ values using a function. We wanted to use a function which shape could be simply tuned using few parameters and which would be similar to a sigmoïd function. We have chosen to use the following function $f(x)$, which is a modified Fermi-Dirac distribution as described in equation 7.5:

$$f(x) = 2\varphi \cdot \left(\frac{1}{1 + \alpha e^{\beta x}} - \frac{1}{2} \right) \tag{7.5}$$

The function $f(x)$ is a decreasing function of x. The input x values can take any real value between $[-\infty; +\infty]$ and as a result, $f(x)$ would take any real values between $[-\varphi; +\varphi]$. Also, $f(x)$ is controlled by the following parameters:

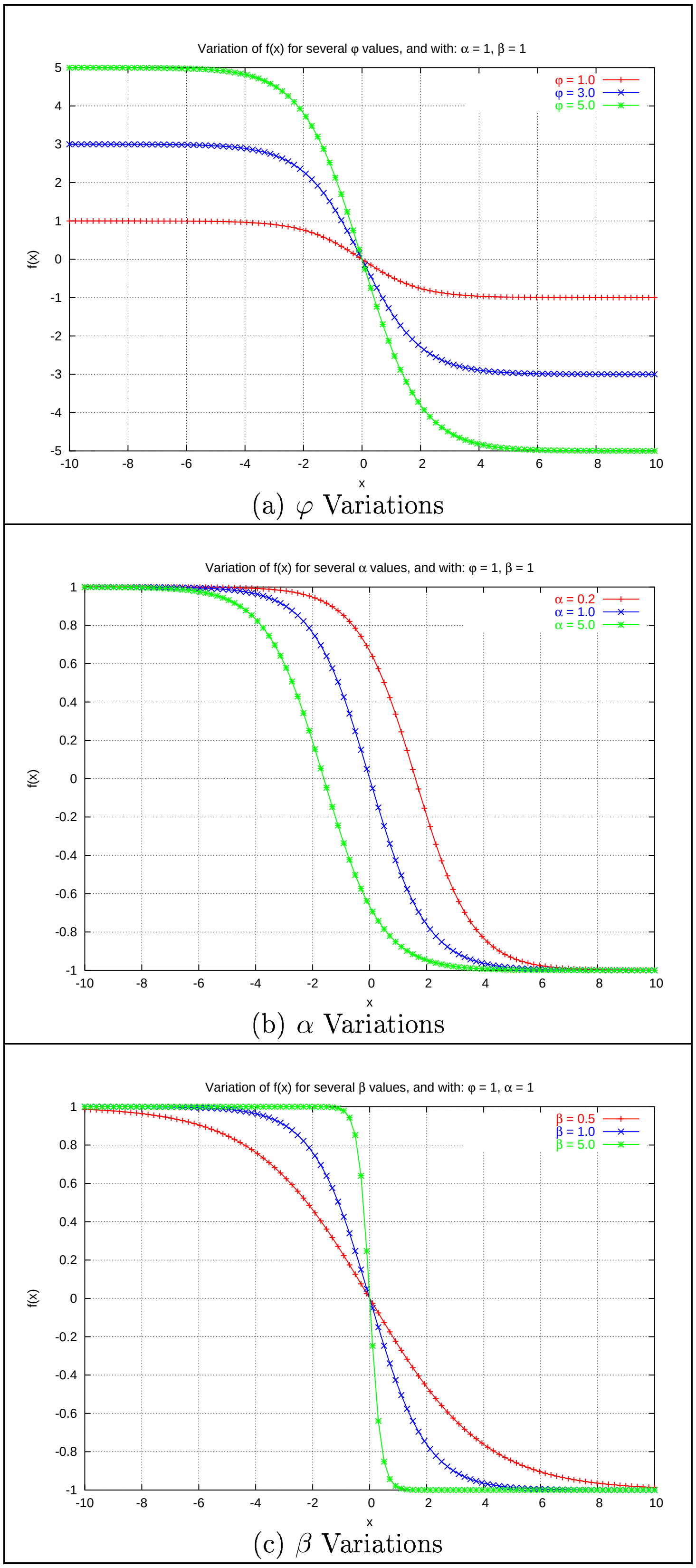
- φ parameter: controls the function amplitude following the y axis.
- α parameter: controls the function's offset following the x axis.
- β parameter: controls the slope of the function.

Figure 7.8 illustrates the aspect of $f(x)$ as a result of variations in φ, α and β.

Let c_l and c_u be respectively the maximum and the minimum achievable capacity values by the system. Note that c_l and c_u are derived respectively from γ_{min} and γ_{max}, using equation 7.2. Hence c_l and c_l are input parameters (dimensionning parameters) for the system. Thus, these values have to be known by the nodes and are not calculated by them.

We would like to limit the range of possible x values while being sure that for these extreme x values, $f(x)$ reached $-\varphi$ and $+\varphi$ with a precision controlled by the parameter ϵ. Thus, after setting a value for ϵ and $|\varphi|$, we then calculate the corresponding extreme x values (x_l and x_u), for the condition that $\alpha = 1$, corresponding to the centered function ($f(0) = 0$) as seen on figure 7.8. We define x_l (minimum achievable x value) such that: $f(x_l) = \varphi - \epsilon$. Similarly, we define x_u (maximum achievable x value) such that: $f(x_u) = -\varphi + \epsilon$. Using equation 7.5 we obtain the following values for x_l and x_u:

$$x_l = \frac{1}{\beta} ln \left(\frac{\epsilon}{2\varphi - \epsilon} \right) \tag{7.6}$$



(a) φ Variations

(b) α Variations

(c) β Variations

Figure 7.8: $f(x)$ Aspect when Separately Varying φ, α and β Parameters

$$x_u = \frac{1}{\beta} ln \left(\frac{2\varphi - \epsilon}{\epsilon} \right) \tag{7.7}$$

In DL and as a result of simultaneous UL Tx from contending nodes, the central node broadcasts all $c_{max\ j}$ values on all the cell SCs. These values correspond to the maximum quantified capacity values extracted among all UL burst messages on each SC j, as shown on figure 7.6. We have: $c_{max\ j} = \max\limits_i \{q(c_{ij})\}$, where c_{ij} and $q(c_{ij})$ are respectively non-quantified and quantified capacity values. Remember that only quantified values are transmitted in UL by the nodes.

To calculate $\Delta\theta_{ij}$ for the current UL/DL time slot, we will associate the x axis to the c axis (capacity axis). Thus, we perform an association between $c_{ij} \in [c_l, c_u]$ values and $x_{ij} \in [x_l, x_u]$ values using the simple following scaling transformation: $x_l \iff c_l$ and $x_u \iff c_u$. Thus, by defining $\Delta x = x_u - x_l$ and $\Delta c = c_u - c_l$, we deduce the following expression for x_{ij} as a function of c_{ij}:

$$x_{ij} = \frac{\Delta x}{\Delta c}(c_{ij} - c_l) + x_l \tag{7.8}$$

As the nodes exchange coded c_{ij} values (not x_{ij} values), we need to use the association presented in equation 7.8 to perform the $\Delta\theta_{ij}$ update. The φ and β parameter values will be the same for all the nodes. There will be one function $f(x)$ used per agent per node.

Whenever receiving $c_{max\ j}$, each node i will:

1. Convert the received $c_{max\ j}$ into an $x_{max\ j}$ value which would be usable as an input for $f(x)$.
2. Calculate a new α value for $f(x)$ such that: $f(x_{max\ j}) = 0$. This corresponds to centering the function for this agent at $x_{max\ j}$. In other words, at this value there will not have any change in the threshold value ($\Delta\theta_{ij} = 0$). Note that for a given node, there will be different α values per node (one per agent), corresponding to the different $c_{max\ j}$ returned per SC j. Finding the new α value is performed by solving $f(x_{max\ j}) = 0$. Using equation 7.5, this leads after some basic calculations to:

$$\alpha = e^{-\beta x_{max\ j}} \tag{7.9}$$

3. Convert its c_{ij} value into an x_{ij} value.
4. Compute the $\Delta\theta_{ij}$ value such that $\Delta\theta_{ij} = f(x_{ij})$. By incorporating the expression of α from equation 7.9 into the expression of $f(x)$ from equation 7.5, we obtain after some elementary calculations:

$$\Delta\theta_{ij} = 2\varphi \cdot \left(\frac{1}{1 + e^{\beta(x_{ij} - x_{max\ j})}} - \frac{1}{2} \right) \tag{7.10}$$

For node i, the operation of calculating the threshold update $\Delta\theta_{ij}$ corresponds to computing equation 7.10 for each agent j.

This is best explained through an example. Thus, figure 7.9 illustrate all these concepts. As an example, and without any loss of generality, we take the following numerical values: $\varphi = 5$, $\alpha = 1$, $\beta = 1$ and $\epsilon = 0.001$. Figure 7.9 shows the variation of the thresholding function with $c_{max\ j}$, and presents an example of $\Delta\theta_{ij}$ calculation. The top horizontal axis corresponds to the capacity axis, and the lower horizontal axis corresponds to the x axis. The following values will be computed before finding out the $\Delta\theta_{ij}$ value:

1. We decide to set for this scenario: $c_l = 4.54$ and $c_u = 15.9$ (refer to the top horizontal axis of figure 7.9).
2. We deduce: $x_l = -6.14$ and $x_u = 6.14$ (refer to the bottom horizontal axis of figure 7.9).
3. Accordingly, we have: $\Delta x = 12.28$ and $\Delta c = 11.36$.
4. In this example, we use a fixed quantification step $q = 1$.
5. In this example, the maximum quantified capacity value between all the nodes on SC j is: $c_{max\ j} = 8$.
6. Using equation 7.8 we obtain: $x_{max\ j} = -2.4$ (refer to the bottom horizontal axis of figure 7.9).
7. Thus, using equation 7.9, the new α value for that SC is: $\alpha = 36.59$ (refer to the red curve in figure 7.9 using the bottom horizontal axis).

Given the point of view of a node i, which has $c_{ij} = 7.5$, it will have to compute its $\Delta\theta_{ij}$ threshold variation using equation 7.10. This operation is shown on figure 7.9 by the arrows, and leads to $\Delta\theta_{ij} = +1.88$.

Using a threshold variation based on the maximum value is a very efficient and scalable method for discriminating a large number of values. As described in [64], there can be an important gain in using an optimized exchange of information when using multi-user diversity. Accordingly, we obtain:

$$\forall i,\ \text{on SC}j\text{:}\ \begin{cases} c_{ij} < c_{max\ j} & \Rightarrow \Delta\theta_{ij} > 0 \\ c_{ij} = c_{max\ j} & \Rightarrow \Delta\theta_{ij} = 0 \\ c_{ij} > c_{max\ j} & \Rightarrow \Delta\theta_{ij} < 0 \end{cases} \tag{7.11}$$

By properly adjusting the numerical values of φ and β we can realize a compromise between the speed of convergence and the quality of the final solution.

Secondly, we want to present more in details how the step 2 (refer to section 7.5.5.2) of the algorithm is performed: the update of each agent's status (ON or OFF) as a function of its threshold θ_{ij}. The probability p_{ij} that an agent i performs its task j is defined in equation 7.12:

$$p_{ij} = \frac{1}{1 + \alpha_s e^{\beta_s \theta_{ij}}} \tag{7.12}$$

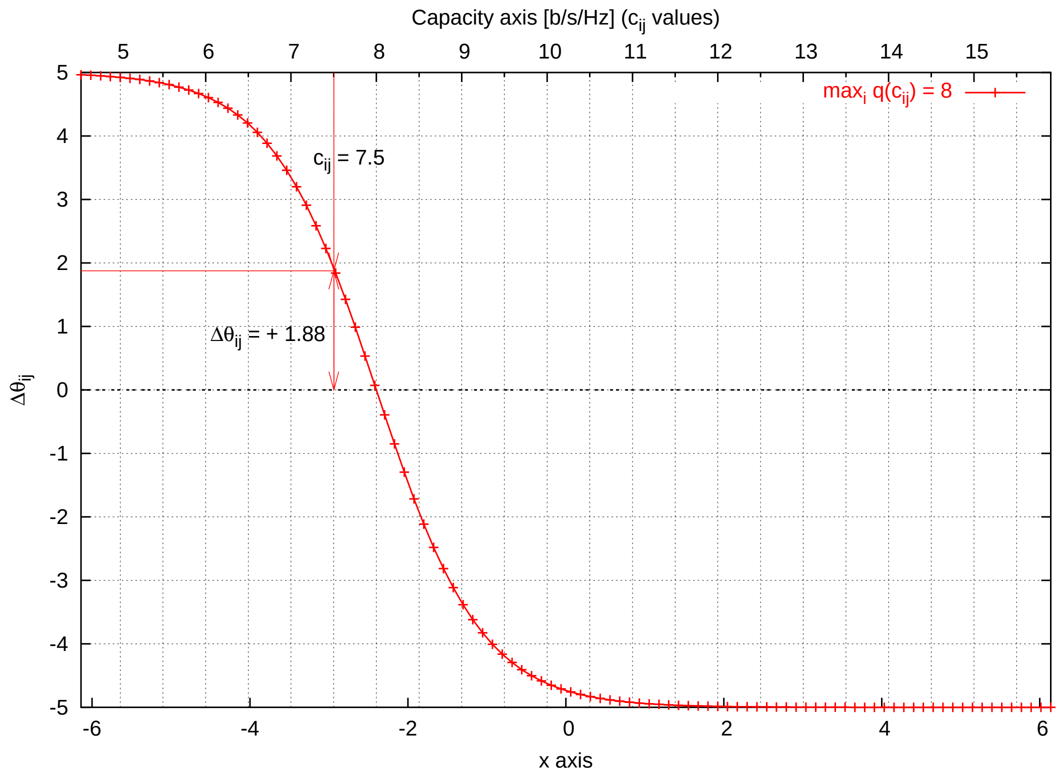


Figure 7.9: $\Delta\theta_{ij}$ Variation for Node i on SC j

The form of this status transition function was chosen for its interesting properties: by adjusting the parameters α_s and β_s it is possible to modify the smoothness of the decision. In the simulations, the following values remain: $\alpha_s = 0.5$, $\beta_s = 2.0$.

Note that:

$$\begin{cases} \theta_{ij} \rightarrow -\infty & \Longrightarrow p_{ij} = 1 \quad \Longleftrightarrow \text{high probability of agent } j \text{ to be active (ON)} \\ \theta_{ij} \rightarrow +\infty & \Longrightarrow p_{ij} = 0 \quad \Longleftrightarrow \text{high probability of agent } j \text{ to be inactive (OFF)} \end{cases} \tag{7.13}$$

In other words, the smaller θ_{ij}, the better the specialization of node i for SC j, thus the more node i will be contending for transmitting on SC j.

Thirdly, note that all the other steps of the algorithm (including the ending criteria) are the same as already presented in section 7.5.5.2.

As a conclusion, our algorithm contains several global and easy to control parameters, allowing to tune the quality of the solution as needed. The variable threshold model offers a great dynamicity in the optimization process, even in changing conditions, and is well adapted to a distributed implementation.

Considering the start values for θ_{ij} between two optimization phases, there is a compromise between letting the algorithm learn over time, versus the rate and amplitude of change of the input data (SNR values).

7.6 Impact of Varying the Algorithm's Control Parameters

The following sections show the role of each of the algorithm's parameters on the speed of convergence and quality of the final solution.

7.6.1 Slope of the Thresholding Function

The slope of the thresholding function (refer to equation 7.5) is controlled by the β parameter as shown in figure 7.10 with various β values.

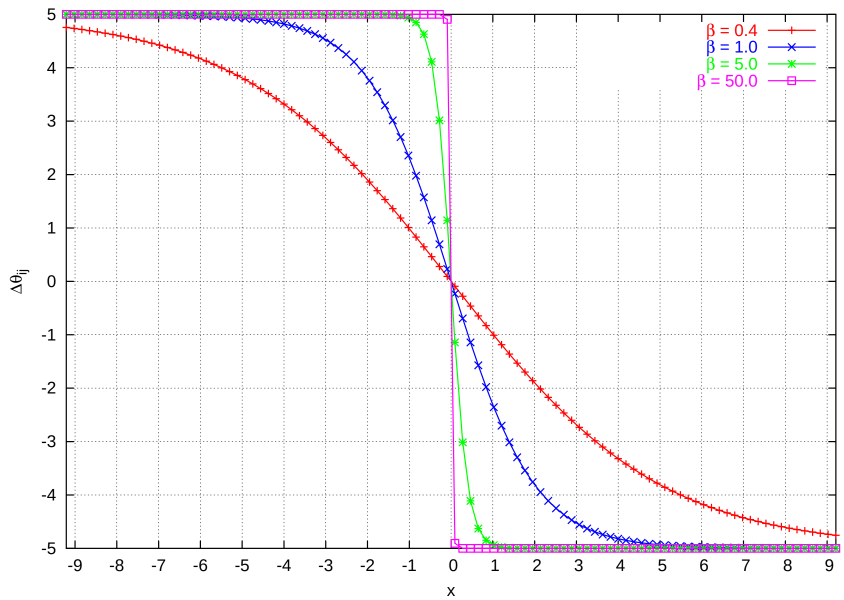


Figure 7.10: $\Delta\theta_{ij}$ Variation as a Function of β

The range of x values used is indicated in section E.8.3.4. Figures 7.11 and 7.12 show the impact of β variation respectively on the number of negotiation steps, and on the total sum capacity. In both cases the numerical values used are summarized in table 7.4.

Figure 7.11 shows that the higher β (i.e. the higher the thresholding function's slope), the smaller the convergence, (i.e. the higher the number of negotiation steps). In addition, the impact of β variation tends to vanish as M increases. In fact, in a given scenario all the nodes use the same thresholding function for all the SCs. Also, due to the adaptive zoom window mechanism, the thresholding effect (variable depending on β) is mostly concentrated within the window, whereas the values outside the window receive extreme $\Delta\theta_{ij}$ values (minimum or maximum) (almost independent from β).

Figure 7.12 shows that the impact of β variation is negligible on the total sum capacity. As a conclusion, depending on β, the final best solution might take longer to be found but will be the same regardless of the current β value.

The following sections present the impact of other parameters' variation on the algorithm's

Notation	Description	Values
ϵ	Parameter used to find out x_l such that $f(x_l) = \varphi - \epsilon$ and x_u such that $f(x_u) = -\varphi + \epsilon$	10^{-3}
θ_{init}	Initial threshold value after reset	0
θ_{min}	Minimum possible threshold value θ_{ij} (refer to equation **??**)	-25
θ_{max}	Maximum possible threshold value θ_{ij} (refer to equation **??**)	$+25$
φ	Parameter controlling the maximum $\Delta\theta_{ij}$ variation per negotiation time slot	5
n_{min}	Minimum simultaneous number of SCs used per active node	1
n_{max}	Maximum simultaneous number of SCs used per active node	1
M	Number of nodes	[10 : 10 : 100]
N	Number of available SCs to allocate	20
During the entire simulation, there is an SNR change and SC re-negotiation before each new DATA transmission. Also, a total of $K = 100$ DATA frames are transmitted for each new number of nodes		

Table 7.4: Numerical Values for the Simulation Parameters (refer to figures 7.11 and 7.12)

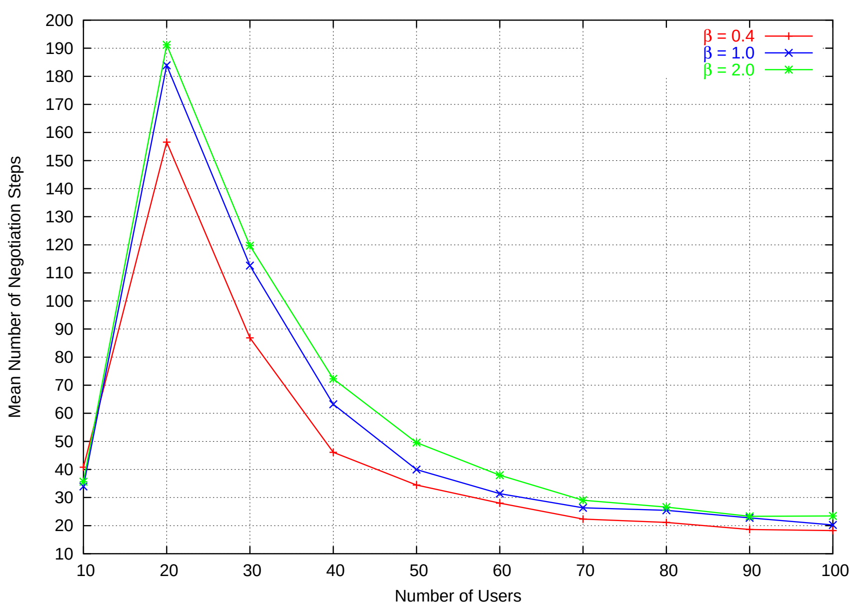


Figure 7.11: Number of Negotiation Steps as a Function of β

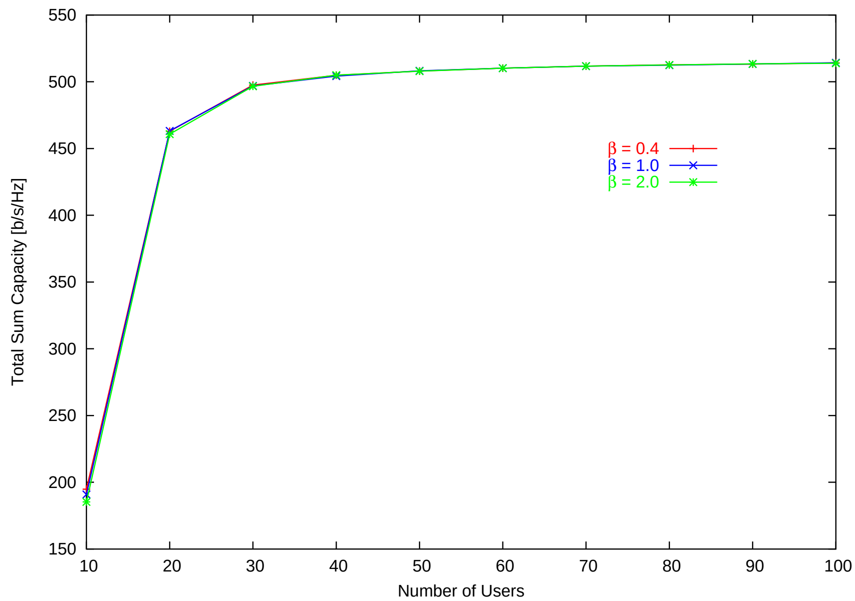


Figure 7.12: Total Sum Capacity as a Function of β

performance.

7.6.2 Controlling the System's Learning and Forgetting Capabilities during a Negotiation Phase

The system's (composed of all the nodes) memory during a negotiation phase is controlled by the following parameters that require to be optimized together: φ, θ_{min} and θ_{max}.

Two scenarios will be investigated as described in tables 7.5 and 7.6.

The numerical values used in the **first scenario** simulations are summarized in table 7.5. The results are presented in figures 7.13 and 7.14.

Figure 7.13 shows the mean and live values for the number of negotiation steps as a function of φ, for the first scenario. Note that for lisibility purposes the results for each value of number of nodes has a small offset in the horizontal axis. The standard deviation reduces as the number of nodes increase. Also, the best results (low number of negotiation steps and low dispersion of values) are obtained for $\varphi = 5$. Figure 7.13 shows that properly tuning the system's learning and forgetting capability has an important impact on the system's negotiation duration. Especially, when considering the system's negotiation duration, the difference between φ and the θ extreme limits (θ_{min} and θ_{max}) should be not too small (nodes become instable deciders with almost no memory of past experiences) and not too important (nodes become slow learners). φ constraints the maximum $\Delta\theta_{ij}$ achievable value (updated after each negotiation time slot), which then feeds θ_{ij}.

Figure 7.14 shows the mean and live values for the total sum capacity as a function of φ, for the first scenario. Note that for lisibility purposes the results for each value of number of nodes

Notation	Description	Values
β	Parameters controlling $f(x)$ function's shape	0.4
ϵ	Parameter used to find out x_l such that $f(x_l) = \varphi - \epsilon$ and x_u such that $f(x_u) = -\varphi + \epsilon$	10^{-3}
θ_{init}	Initial threshold value after reset	0
θ_{min}	Minimum possible threshold value θ_{ij} (refer to equation **??**)	-50
θ_{max}	Maximum possible threshold value θ_{ij} (refer to equation **??**)	$+50$
n_{min}	Minimum simultaneous number of SCs used per active node	1
n_{max}	Maximum simultaneous number of SCs used per active node	1
M	Number of nodes	[10 : 10 : 100]
N	Number of available SCs to allocate	20
*The **first scenario** will vary φ while θ_{min} and θ_{max} remain constant. During the entire simulation, there is an SNR change and SC re-negotiation before each new DATA transmission. Also, a total of $K = 100$ DATA frames are transmitted*		

Table 7.5: Numerical Values for the **First Scenario** Simulation Parameters (refer to figures 7.13 and 7.14)

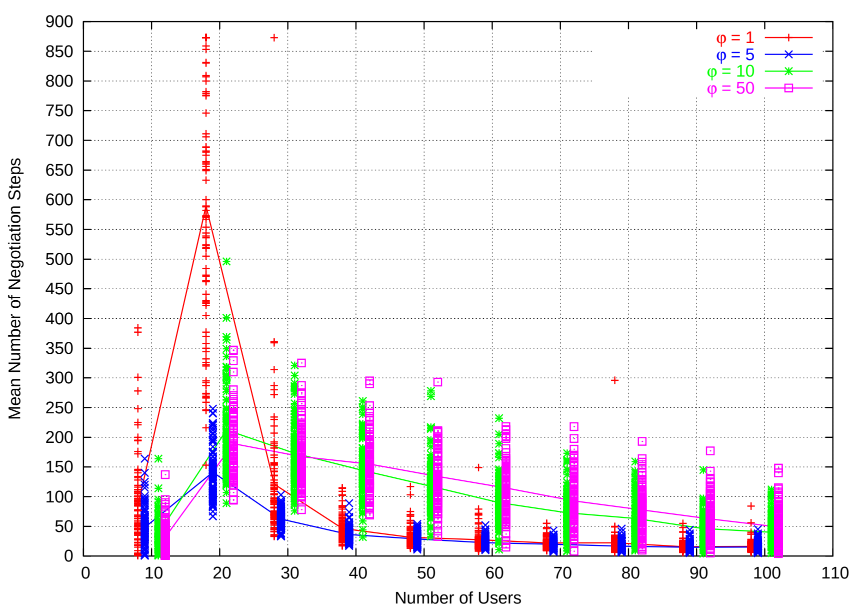


Figure 7.13: Number of Negotiation Steps as a Function of φ, for the First Scenario

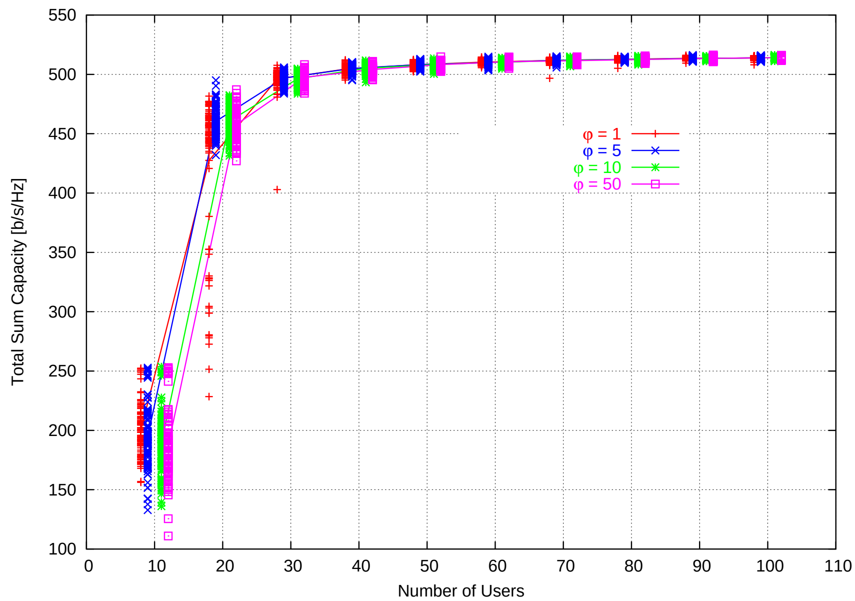


Figure 7.14: Total Sum Capacity as a Function of φ, for the First Scenario

has a small offset in the horizontal axis. The standard deviation reduces as the number of nodes increase. Also, Figure 7.14 shows that as far as the total sum capacity is concerned, there is almost no impact of the memory (no major difference in the standard deviation of values as M increases).

The numerical values used in the **second scenario** simulations are summarized in table 7.6. The results are presented in figures 7.15 and 7.16.

Figure 7.15 shows a slight improvement on the speed of convergence when compared to figure 7.13 indicating that it is beneficial to have a sufficient system memory for a given φ value. Also, the amplitude of the dispersion of values is reduced.

Figure 7.16 shows almost no impact of the system memory on the final total UL sum capacity. In addition, the total capacity values are less dispersed than in the first scenario.

7.6.3 Controlling the System's Learning and Forgetting Capabilities between Negotiation Phases

As represented in figure 7.17, updating the θ_{ij} value for each node i and SC j, after the end of the DATA phase and before a negotiation phase, can be performed in multiple possibilities and we will compare the following ones:

1. Keep the previous θ_{ij} value.
2. Reinitialize θ_{ij} to θ_{init}.
3. Randomly change θ_{ij} to a value between θ_{min} and θ_{max}.

Notation	Description	Values
β	Parameters controlling $f(x)$ function's shape	0.4
ϵ	Parameter used to find out x_l such that $f(x_l) = \varphi - \epsilon$ and x_u such that $f(x_u) = -\varphi + \epsilon$	10^{-3}
θ_{init}	Initial threshold value after reset	0
θ_{min}	Minimum possible threshold value θ_{ij} (refer to equation **??**)	$-10 \times \varphi$
θ_{max}	Maximum possible threshold value θ_{ij} (refer to equation **??**)	$+10 \times \varphi$
n_{min}	Minimum simultaneous number of SCs used per active node	1
n_{max}	Maximum simultaneous number of SCs used per active node	1
M	Number of nodes	[10 : 10 : 100]
N	Number of available SCs to allocate	20
The **second scenario** *will vary φ while θ_{min} and θ_{max} are varied as a function of φ to keep enough memory depth while investigating the impact of the θ variation amplitude. During the entire simulation, there is an SNR change and SC re-negotiation before each new DATA transmission. Also, a total of $K = 100$ DATA frames are transmitted*		

Table 7.6: Numerical Values for the **Second Scenario** Simulation Parameters (refer to figures 7.15 and 7.16)

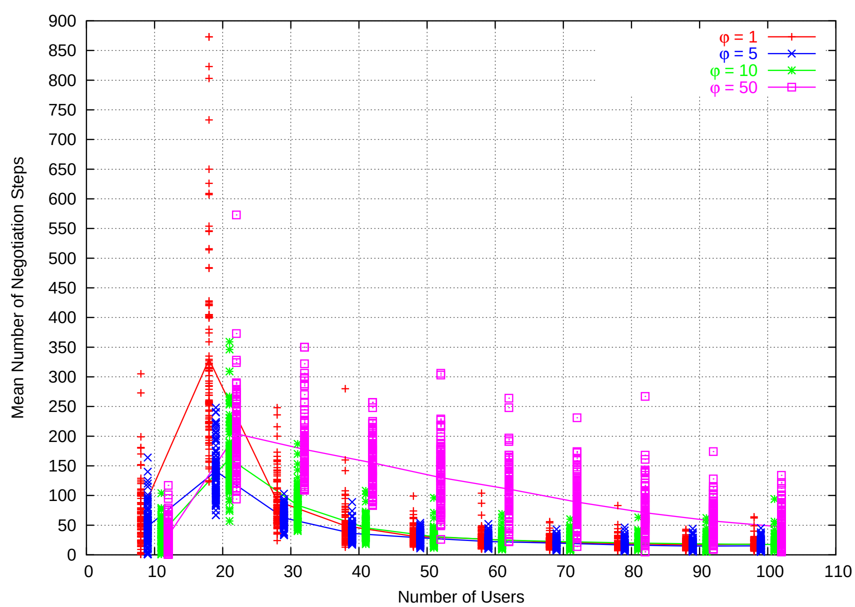


Figure 7.15: Number of Negotiation Steps as a Function of φ, for the Second Scenario

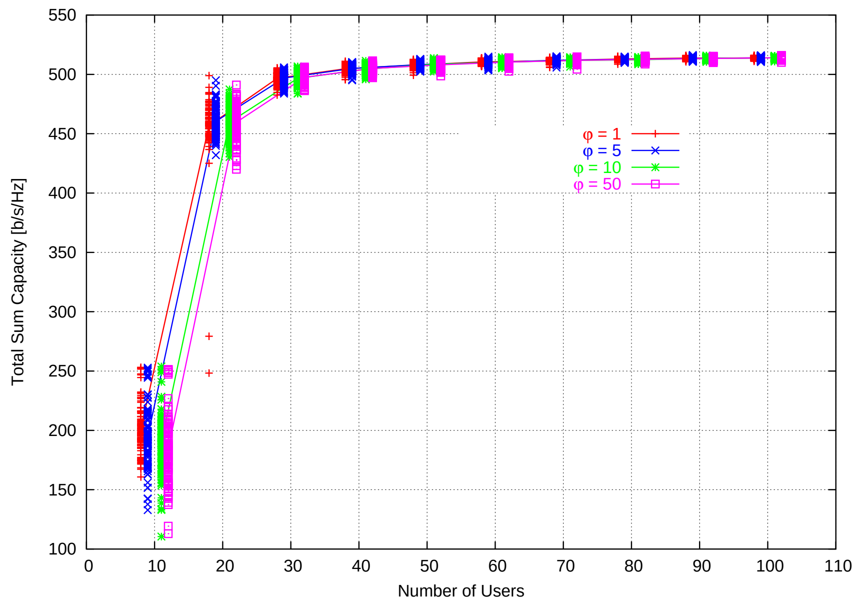


Figure 7.16: Total Sum Capacity as a Function of φ, for the Second Scenario

4. θ_{ij} takes a value reflecting the quality of the capacity value c_{ij} by doing an appropriate scaling.

5. Change θ_{ij} according to a given function taking for example into account channel variations or other metrics.

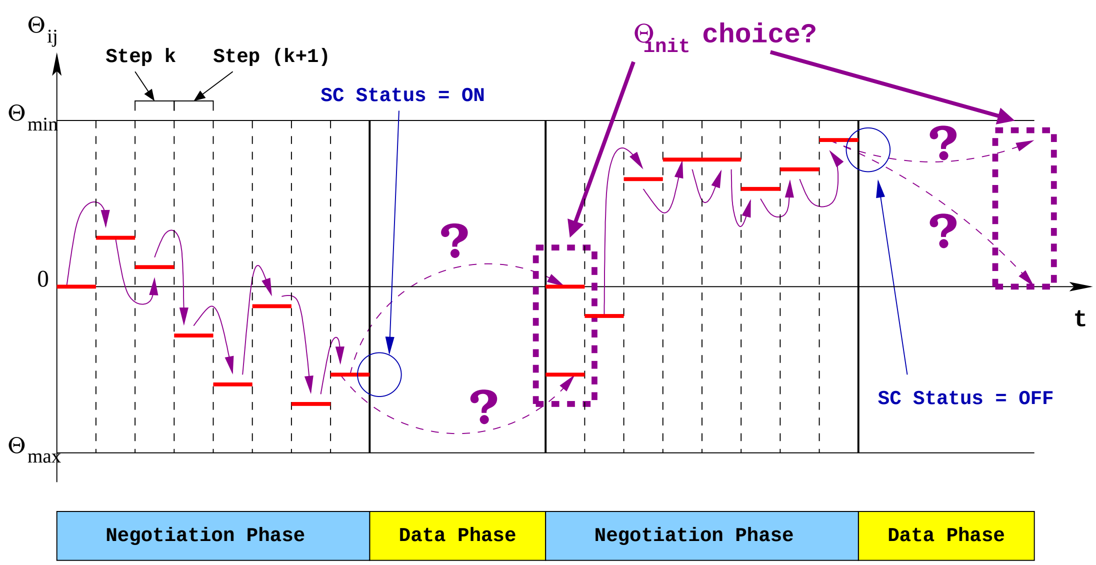


Figure 7.17: Representation of the θ_{init} Update Challenge

Then, we compare methods 1. to 4. whose results are presented in figures 7.18 and 7.19. The numerical values used in these simulations are summarized in table 7.7.

Figure 7.18 shows the impact of each θ_{ij} update method on the speed of convergence.

Notation	Description	Values
β	Parameters controlling $f(x)$ function's shape	0.4
ϵ	Parameter used to find out x_l such that $f(x_l) = \varphi - \epsilon$ and x_u such that $f(x_u) = -\varphi + \epsilon$	10^{-3}
θ_{min}	Minimum possible threshold value θ_{ij} (refer to equation **??**)	-50
θ_{max}	Maximum possible threshold value θ_{ij} (refer to equation **??**)	$+50$
φ	Parameter controlling the maximum $\Delta\theta_{ij}$ variation per negotiation time slot	5.0
n_{min}	Minimum simultaneous number of SCs used per active node	1
n_{max}	Maximum simultaneous number of SCs used per active node	1
M	Number of nodes	[10 : 10 : 100]
N	Number of available SCs to allocate	20
During the entire simulation, there is an SNR change and SC re-negotiation before each new DATA transmission. Also, a total of $K = 100$ DATA frames are transmitted		

Table 7.7: Numerical Values for the Simulation Parameters (refer to figures 7.18 and 7.19

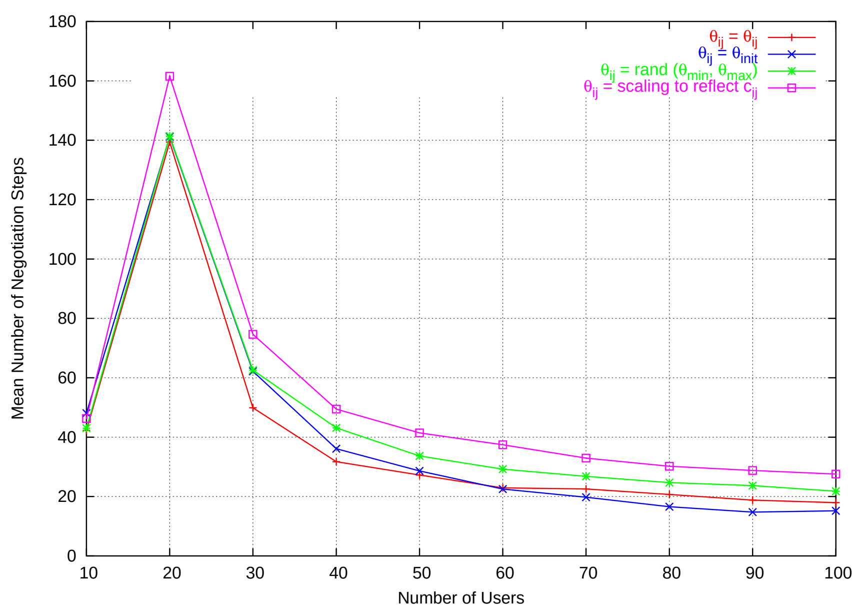


Figure 7.18: Impact of Various θ_{ij} Update Methods on the Speed of Convergence

As seen in figure 7.18, the method used to update θ_{ij} between two spectrum reallocations (after a channel change) has an impact on the negotiation duration. Remember that in our simulations the channel conditions are randomly updated. The two best methods are to reset θ_{ij} to 0 or to its previous value (no change), corresponding respectively to resetting the system's memory or not changing it at all. When the channel conditions do not change during the negotiation phase then the system memory (distributed in all the nodes and cooperatively updated) is powerful. However, having a memory for the system is mostly beneficial when the input conditions do not change too abruptly, otherwise it is a useless or even performance limiting factor.

Accordingly, when the new channel conditions are randomly drawn, then the best solution is to make no assumption and to draw a θ_{init} value located in the middle of the range, i.e. here it is 0.

Figure 7.19 shows the impact of each θ_{ij} update method on the total sum capacity.

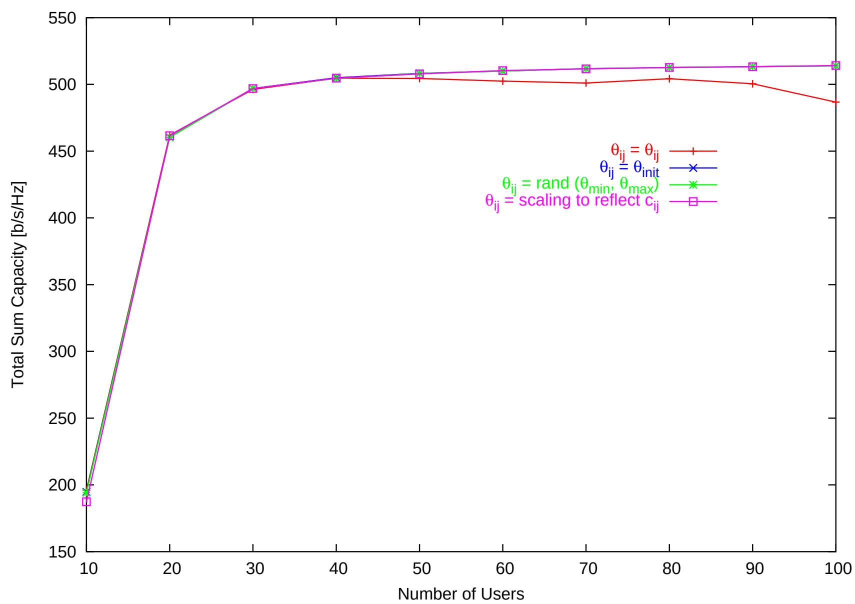


Figure 7.19: Impact of Various θ_{ij} Update Methods on the Total Sum Capacity

As seen in figure 7.19, all the methods find equivalent solutions in terms of system sum capacity. As M increases the method that do not update θ_{ij} gives lower results compared to the other methods.

As a conclusion, in the absence of a channel model (and here with randomly varying channels) in order to minimize negotiation duration and maximize system total sum capacity, the best method is to set θ_{ij} to 0 between two channel variations. This section has also shown that the learning process is beneficial for the system, but resetting the system's memory shall be done using a method adapted to the input parameters' variations. Otherwise, an out of date memory can do more harm than good, and slows down the search for a new best solution.

7.6.4 Conclusion

As a conclusion, this part was not intended for comparison with another method, but rather to show the algorithm's reaction under different changes in its control parameters. We saw the importance of controlling the system's learning capabilities in accordance with the channel conditions, in order to quickly find the best solutions.

7.7 Impact of Varying the Problem's Input Parameters

The following sections show the impact of varying each of the algorithm's input data parameters on the speed of convergence and quality of the final solution.

7.7.1 Sudden Channel Variations between Negotiation Phases

In this section, the channels change only after some time, but always between phases, especially between a data phase and a negotiation phase. The numerical values used in these simulations are summarized in table 7.8. The results are presented in figures 7.20 and 7.21.

Notation	Description	Values
β	Parameters controlling $f(x)$ function's shape	0.4
ϵ	Parameter used to find out x_l such that $f(x_l) = \varphi - \epsilon$ and x_u such that $f(x_u) = -\varphi + \epsilon$	10^{-3}
θ_{init}	Initial threshold value after reset	0
θ_{min}	Minimum possible threshold value θ_{ij} (refer to equation **??**)	-50
θ_{max}	Maximum possible threshold value θ_{ij} (refer to equation **??**)	$+50$
φ	Parameter controlling the maximum $\Delta\theta_{ij}$ variation per negotiation time slot	5.0
n_{min}	Minimum simultaneous number of SCs used per active node	1
n_{max}	Maximum simultaneous number of SCs used per active node	1
M	Number of nodes	100
N	Number of available SCs to allocate	20
During the simulation, a total of $K = 1000$ DATA frames are transmitted. θ_{ij} is reset to θ_{init} after each DATA frame		

Table 7.8: Numerical Values for the Simulation Parameters (refer to figures 7.20 and 7.21

Figures 7.20 and 7.21 compare the convergence speed of the algorithm between three cases:

1. Case 1:
    - There is a channel change after each DATA frame.
    - The nodes reset their SC status after each DATA frame.

2. Case 2:
   - There is a channel change only after 300^{th} and 700^{th} DATA frame.
   - The nodes reset their SC status after each DATA frame.
3. Case 3:
   - There is a channel change only after 300^{th} and 700^{th} DATA frame.
   - The nodes reset their SC status only after 300^{th} and 700^{th} DATA frame.

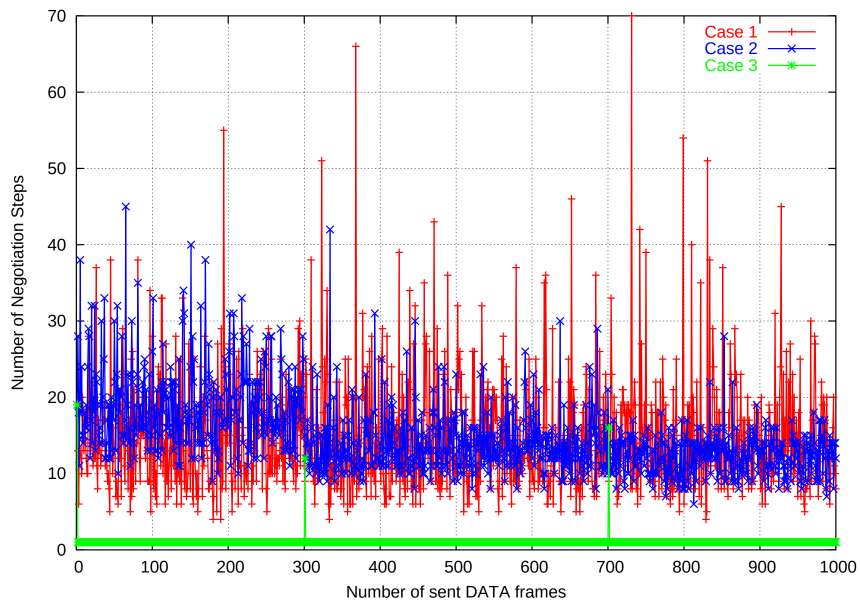


Figure 7.20: Impact of the Channel Variations between Negotiation Phases on the Convergence Speed

Figure 7.20 shows the algorithm's stability under live channel variations. In case 1, the channel is changing very often and the algorithm reacts to these changes. In case 2, the channel changes only twice but the algorithm still resets its memory and found solution as if it was each time a new problem. Compared to case 1, the standard deviation in number of negotiation steps is smaller in case 2. Indeed, when the channel does not change, the problem to solve has the same complexity until the next change. In case 3, the channel changes as often as in case 2, but the system keeps its memory until the next channel change. Accordingly, it takes only a few negotiation steps to converge after each change and then only 1 step.

Figure 7.21 shows that in case 1 a new solution needs to be found after each channel change. On the contrary, in case 2, when there is no channel change, the system hesitates between a few close solutions (4 before the first change, 2 after the first change and then only 1 after the second change). Of course, in case 3 the system keeps the same found solution between changes.

As a conclusion, the system adapts to the channel changes and tries to find new solutions when necessary (case 1), or finds the same best solutions when there is no change in a small number of negotiation steps (case 2) or immediately when the system knows that no change occurred (case 3). This shows the algorithm's potential to adapt its search possibilities and resulting performances according to its knowledge of the problem at hand.

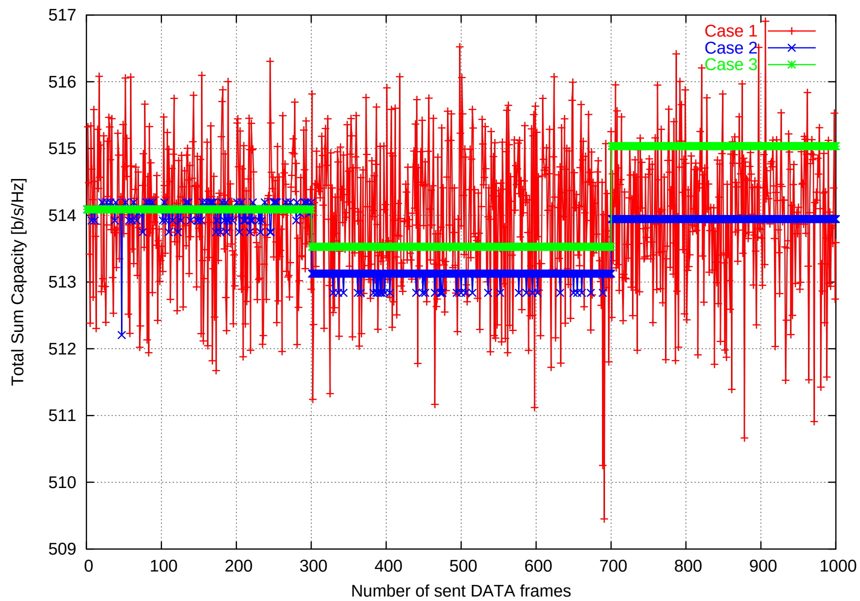


Figure 7.21: Impact of the Channel Variations between Negotiation Phases on the Total Sum Capacity

7.7.2 Channel Variations within Negotiation Phases

In this section we study the potential for the algorithm to adapt to channel variations at random instants during the negotiation phases, in terms of convergence speed.

The numerical values used in these simulations are summarized in table . The results are presented in figure 7.22.

We vary the probability to have a channel change at the next UL/DL timeslot. A channel change corresponds to an abrupt change for all SCs of all the nodes (contending or not). Some contending nodes might switch to non contending or the contrary. The entire system changes. In addition, the θ_{ij} are not reinitialized. Rather, we let the algorithm discover the change by itself and react to find the new best solution as quickly as possible before the next change occurs.

Figure 7.22 presents the maximum, mean, and minimum number of negotiation steps for each value of channel change probability.

As seen in figure 7.22, increasing the number of channel changes results in a longer mean negotiation phase duration. In fact, an increased number of changes increases the negotiation duration, which in turn increases the probability of having another change before convergence, etc. This is explains the important increase in maximum negotiation length. Thus, there is an avalanche effect. In case of an abrupt change, resetting the θ_{ij} values after a change would gain some time in negotiation, rather than trying to fully forget and readapt to the new input values. There is tradeoff to consider.

Notation	Description	Values
β	Parameters controlling $f(x)$ function's shape	0.4
ϵ	Parameter used to find out x_l such that $f(x_l) = \varphi - \epsilon$ and x_u such that $f(x_u) = -\varphi + \epsilon$	10^{-3}
θ_{init}	Initial threshold value after reset	0
θ_{min}	Minimum possible threshold value θ_{ij} (refer to equation **??**)	-50
θ_{max}	Maximum possible threshold value θ_{ij} (refer to equation **??**)	$+50$
φ	Parameter controlling the maximum $\Delta\theta_{ij}$ variation per negotiation time slot	5.0
n_{min}	Minimum simultaneous number of SCs used per active node	1
n_{max}	Maximum simultaneous number of SCs used per active node	1
M	Number of nodes	100
N	Number of available SCs to allocate	20
During the simulation, a total of $K = 1000$ DATA frames are transmitted		

Table 7.9: Numerical Values for the Simulation Parameters (refer to figure 7.22)

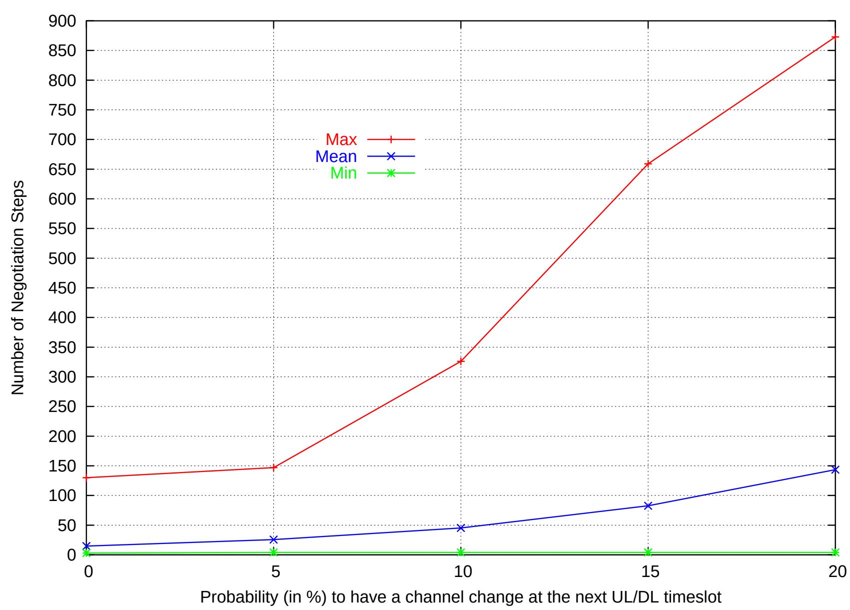


Figure 7.22: Impact of the Channel Variations within Negotiation Phases on the Convergence Speed

7.7.3 Impact of M and N variations

In this section we study the impact of varying the M and N parameters, in terms of convergence speed and total UL sum capacity. The following constant numerical values are used: $n_{min} = 1$ and $n_{max} = 10$. The results are illustrated in figures 7.23 and 7.24.

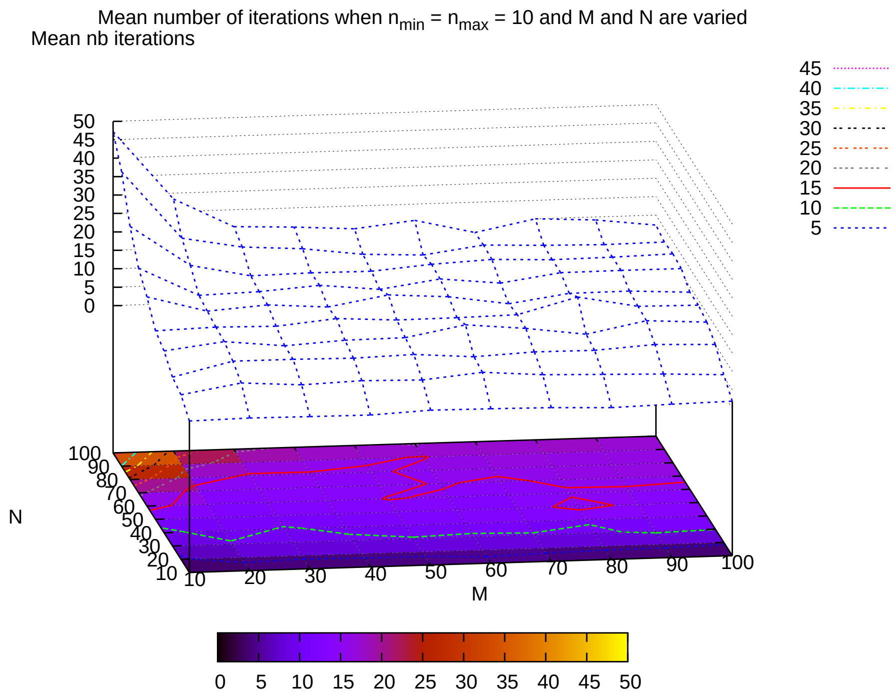


Figure 7.23: Impact of the M and N Variations on the Mean Number of Negotiation Steps

Figure 7.23 shows the algorithm's scalability in number of nodes: the higher M the smaller the mean convergence duration. Also, the higher N the higher the mean convergence duration. However, the duration stays reasonable considering the problem size.

As seen in figure 7.24, for a given N value, the total sum capacity benefits from increasing M (multi-use diversity) as it brings new nodes with some better channel conditions. In addition, the total sum capacity increases with N.

7.7.4 Impact of n_{min} and n_{max} Variations

In this section we study the impact of varying the n_{min} and n_{max} parameters, in terms of convergence speed. The following constant values are used: $M = 100$ and $N = 100$. The results are illustrated in figure 7.25.

As seen in figure 7.25, some n_{min}; n_{max} combinations require more negotiation steps than others, especially the case 1; 1 for which each node must have at the end only a single SC. It can

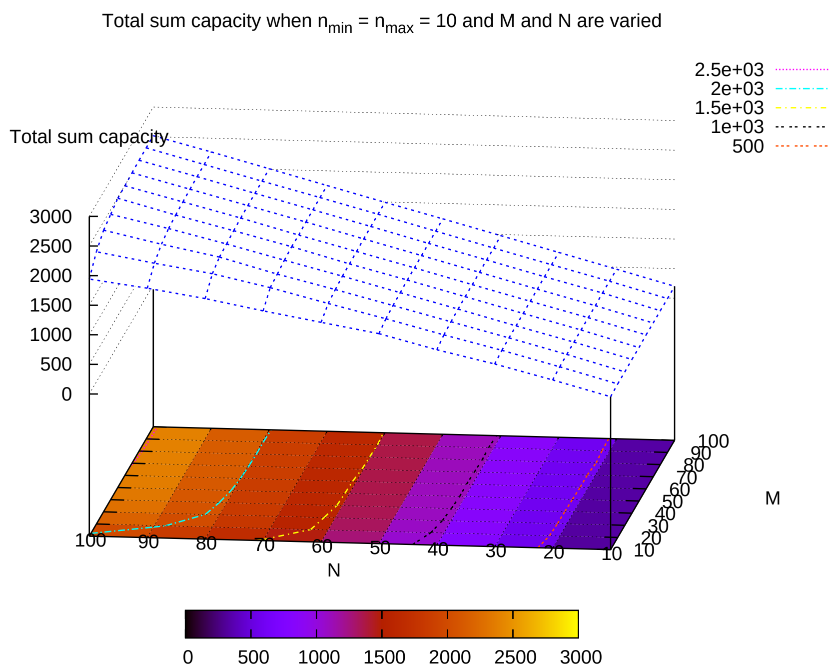


Figure 7.24: Impact of the M and N Variations on the Mean Number of Negotiation Steps

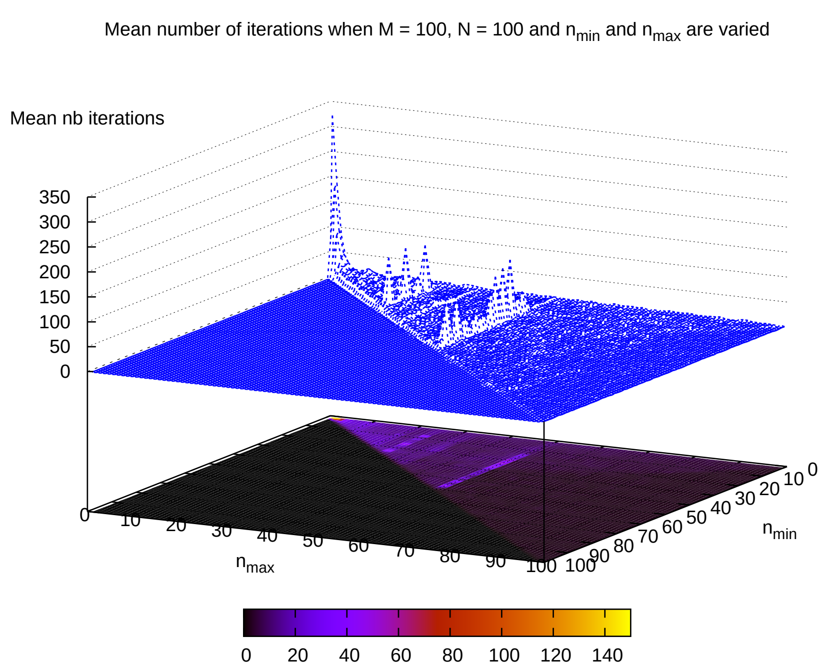


Figure 7.25: Impact of the n_{min} and n_{max} Variations on the Mean Number of Negotiation Steps

be noted that whenever $n_{min} = 1$ and $n_{max}! = 1$, the convergence is always quickly achieved. This is simply explained by the fact that there is always some place left to fit least a single SC.

7.7.5 Impact of an Abrupt Change of M

In this section we show the good adaptation capabilities of the algorithm to a load variation in terms of quality of the reached solution and convergence time.

The numerical values used in these simulations are summarized in table . Figures 7.26 and 7.27 show the adaptation of the algorithm to the modified cell load (in number of nodes).

Notation	Description	Values
n_{min}	Minimum simultaneous number of SCs used per active node	1
n_{max}	Maximum simultaneous number of SCs used per active node	10
M	Number of nodes	varied
N	Number of available SCs to allocate	100
The total number of transmitted DATA frames over time is $K = 1200$. The number of contending cell nodes M is abruptly changed during the simulation. M is abruptly changed as follows: at $D = 0$, $M = 20$, then at $D = 400$, $M = 100$ and finally at $D = 800$, $M = 60$		

Table 7.10: Numerical Values for the Simulation Parameters (refer to figure 7.26 and 7.27)

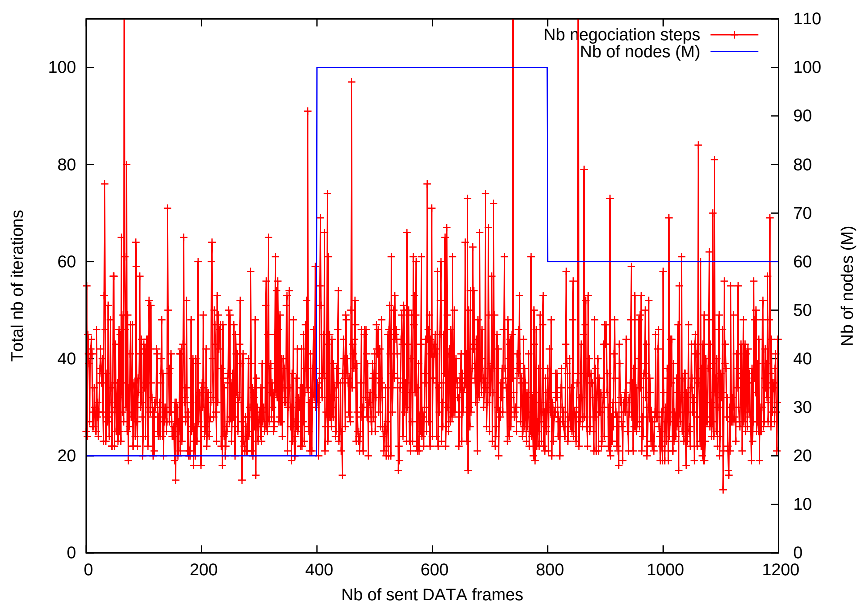


Figure 7.26: Number of Iterations for $N = 100$

As seen in figures 7.26 and 7.27, the agents were successfully able to adapt to a modified (increase and decrease) number of cell nodes. The number of negotiation steps remained small (mean value = 34.8), considering the optimization problem size. Also, the number of negotiation steps remained almost independent from M variation and constant even for a population

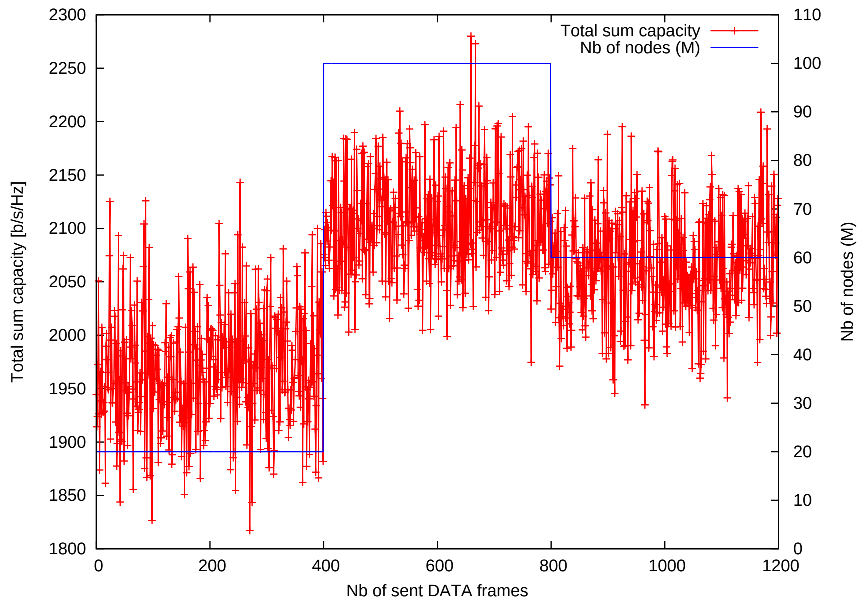


Figure 7.27: Total Sum Capacity for $N = 100$

multiplied by 5 and then decreased by $\frac{5}{3}$. The number of negotiation steps does not appear to depend on M, thus proving the scalability of our distributed algorithm and its great interest.

In Figure 7.27, the obtained system sum capacity was increased with an increased value of M, proving the multi-user diversity interest to increase the system sum capacity.

7.7.6 Conclusion

As a conclusion, these sections have been useful in showing the many interesting algorithm's properties under several input parameters' variations. Indeed, the scalability with the number of nodes, the robustness against changes, adaptability, possibility to optimize the system under any respective value of: n_{min}, n_{max}, M and N. The multi-user diversity helped in increasing the capacity with M. Also, as increasing M does not significantly increase the negotiation duration, accordingly, we obtain an increasing throughput with M.

7.8 Time-based Fairness

7.8.1 Introduction

In this section we implement a distributed time-based fairness algorithm to solve the spatial unfairness problem in channel access, efficient for both uniform and non-uniform nodes' distribution. Our contribution is built upon the mechanisms presented in section 7.5, enhanced to offer a distributed time-based fairness control in SC access. Accordingly, it provides a self-controlled algorithm for QoS management, without the need for the cell to know the current number of contending nodes, or for the nodes to maintain a list of neighbors.

The optimization problem presented in equation E.5 of section 7.5.2 was augmented to include a time-based fairness constraint (constraint 5). We refer to $\alpha_{ij}(t_1, t_2)$ as the service time (for DATA transmission) of node i on SC j between t_1 and t_2. We refer to $f(i, j, k, x)$ as a given function controlling the amount of time-based fairness in the system. The optimization problem can be formulated (including the additional constraint) in mathematical terms as follows. The objective of the optimization problem is to (for the notations refer to equation E.5 of section 7.5.2):

$$\begin{aligned} &\max \sum_i \sum_j a_{ij} c_{ij} \\ &\text{s.t.:} \begin{cases} (1)-(4) & \text{Exactly the same constraints as in equation E.5 of section 7.5.2} \\ (5) & Pr(|\alpha_{ik}(t_1,t_2) - \alpha_{jk}(t_1,t_2)| \geq x) \leq f(i,j,k,x) \end{cases} \end{aligned} \tag{7.14}$$

7.8.2 Distributed Time-based Fairness Algorithm

The time-based fairness mechanism implemented consists in a modification of the centralized version presented in [65] adapted to a distributed OFDMA context. The algorithm allows a controllable statistical bound for the fairness by means of a single parameter ξ (called β in the original article), which changes the cost of using a given SC. Our version allows each node to differ its contention or not, depending on the fairness threshold probability. The algorithm is executed internally by each node and requires no exchange of information between the nodes. As a result, the geographical dispersion of the nodes is more or less hidden as a function of the desired level of fairness. Note that our algorithm is scalable with the number of nodes in the cell.

After the end of the resources use, all nodes update their fairness values for all the SCs separately, based on the waited/transmitted time, γ_{ij} quantity and duration of the previous resource use.

In the legacy algorithm, the central fairness controller schedules the node with the minimum value V such that:

$$\begin{aligned} &V = min_i \frac{L_i(p) - K_i(p) + U_i(p)}{\phi_i} \\ &\text{with:} \\ &\begin{cases} \phi_i \text{ the weight of flow } i \\ L_i(p) \text{ the head-of-line packet length for flow } i \text{ at time } p \\ K_i(p) \text{ credit counter of flow } i \text{ at time } p \\ U_i(p) \text{ the estimated cost for node } i \text{ to transmit the } p^{th} \text{ packet} \end{cases} \end{aligned} \tag{7.15}$$

In our study we assumed $\phi_i = 1, \forall i$. In the distributed version, we use a modified Fermi-Dirac distribution to find the fairness threshold probability f_{ij} for node i to contend for SC j, such that:

$$f_{ij} = \frac{1}{1 + \alpha_{Fair} e^{\beta_{Fair}(L_i(p) - K_i(p) + U_i(p))}} \tag{7.16}$$

This algorithm is achieved internally by each nodes and require no exchange of information between nodes. The only required information for the nodes after a use of resource j by node i is to know the value: $K_i(p) - L_i(p)$, which can simply be included in a broadcast message as an additional field.

This algorithm hides the geographical dispersion of the nodes. In addition, due to its distributed aspect, it is scalable with the number of nodes. Note that in the algorithm we can either use the fairness or not, and when using it we can control the amount of fairness required.

The next section presents the simulation results obtained in the context of several nodes distributed in a cell.

7.8.3 Results Analysis

The numerical values used in the following scenario are modified values taken from an IEEE 802.11a system. However, other values could be taken without altering the conclusions.

Taking as an example an 802.11a system [2] in the band $[5.25 - 5.35]$ GHz, we calculated the maximum range of SNR values for any cell node. The maximum range of capacity values per SC was deduced. After calculations and taking into account the maximum transmit power regulations, the SC bandwidth and the set of possible transmit modes, we obtain that all SNR values γ_{ij} $(\forall i, j)$ are within: $[\gamma_{min}; \gamma_{max}] = [13.5; 78.07]$ dB . In our study, only the continuous capacity case is considered. We can now transform these SNR values to find out the maximum capacity range c_{ij} $(\forall i, j)$ are within: $[c_{min}; c_{max}] = [4.54; 25.9]$ b/s/Hz (for a SC bandwidth of 1 Hz). c_{ij} values are uniformly drawn within $[c_{min}; c_{max}]$. Note, that the aim of these studies is not to focus on the channel model, but rather to demonstrate the benefit of the proposed algorithm.

The **first results** compare a case when the fairness is not taken into account with a case when it is taken into account with a variable level of fairness. We are interested in comparing in each case the performance of the algorithm in terms of speed of convergence and quality of the solution.

The results presented are obtained for the following range of values: $N = 90$, $M = 20$, $n_{min} = 1$ and $n_{max} = 5$, ξ was varied. Also nodes are split between 3 categories of channel conditions: "bad", "neutral", "good". Each category is drawing its SNR values from a third of the total range of values.

Figures 7.28 and 7.29 show the obtained results.

We can see that the impact of ensuring fairness compared to a system not controlling fairness at all, is a reduced system capacity and an increase in the number of negotiation steps. Also note that the results for the case with no fairness control remained independent from ξ.

Another considered scenario is studied and consists in $M = 50$, $N = 10$, $n_{min} = 1$ and $n_{max} = 2$. Also, not all the nodes are in the same radio conditions, to simulate nodes close to the AP and others far from it. The following SNR values are imposed:

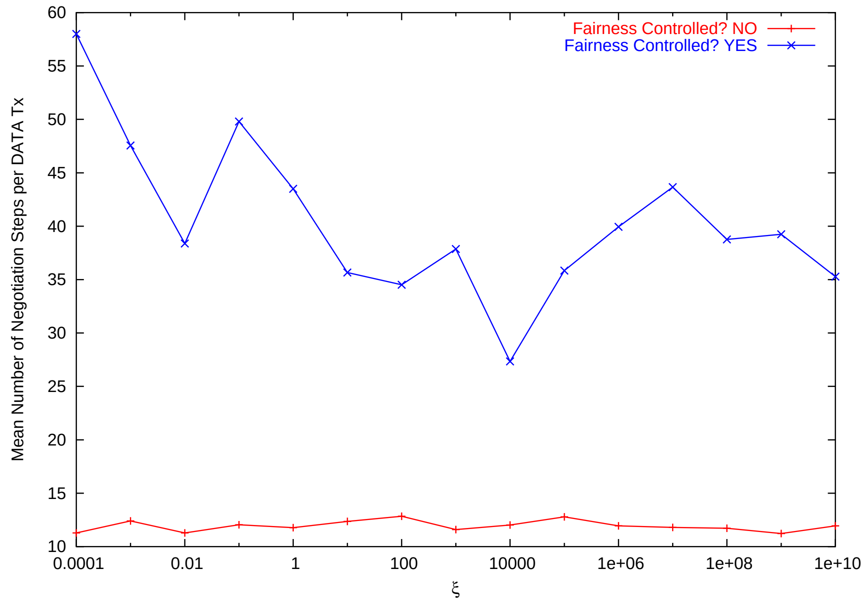


Figure 7.28: Mean Number of Negotiation Steps per DATA Tx as a function of ξ

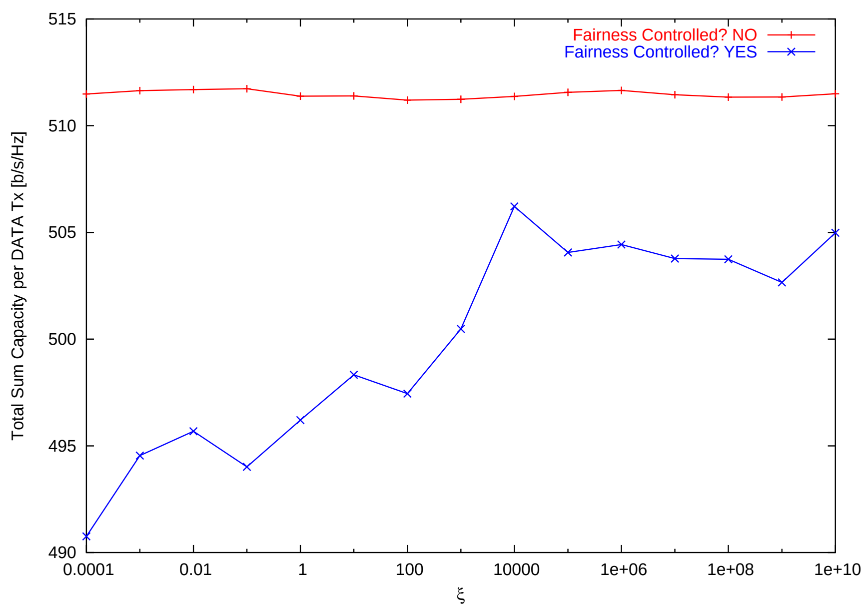


Figure 7.29: Total Sum Capacity per DATA Tx [b/s/Hz] as a function of ξ

- $M = [1:5]$ and $N = [1:5]$: Good radio conditions such that $\gamma_{ij} \in [\gamma_{max} - 28.07; \gamma_{max}]$ dB .
- $M = [1:5]$ and $N = [6:10]$: Bad radio conditions such that $\gamma_{ij} \in [\gamma_{min}; \gamma_{min} + 20.0]$ dB .
- $M = [6:50]$ and $N = [1:10]$: Mean radio conditions such that $\gamma_{ij} \in [\gamma_{min}; \gamma_{max}]$ dB .

To estimate the fairness in spectrum access, we compute Jain's fairness index F_j on SC j, the standard traditional measure of network fairness [52]. We refer to D_{ij} as the number of times node i used SC j. Thus, we obtain the expression:

$$F_j = \frac{(\sum_{i=1}^{M} D_{ij})^2}{M \cdot \sum_{i=1}^{M} {D_{ij}}^2}$$

When $F_j = 0$ there is no fairness, and when $F_j = 1$ there is perfect fairness.

Figure 7.30 presents the obtained level of fairness when ξ varies.

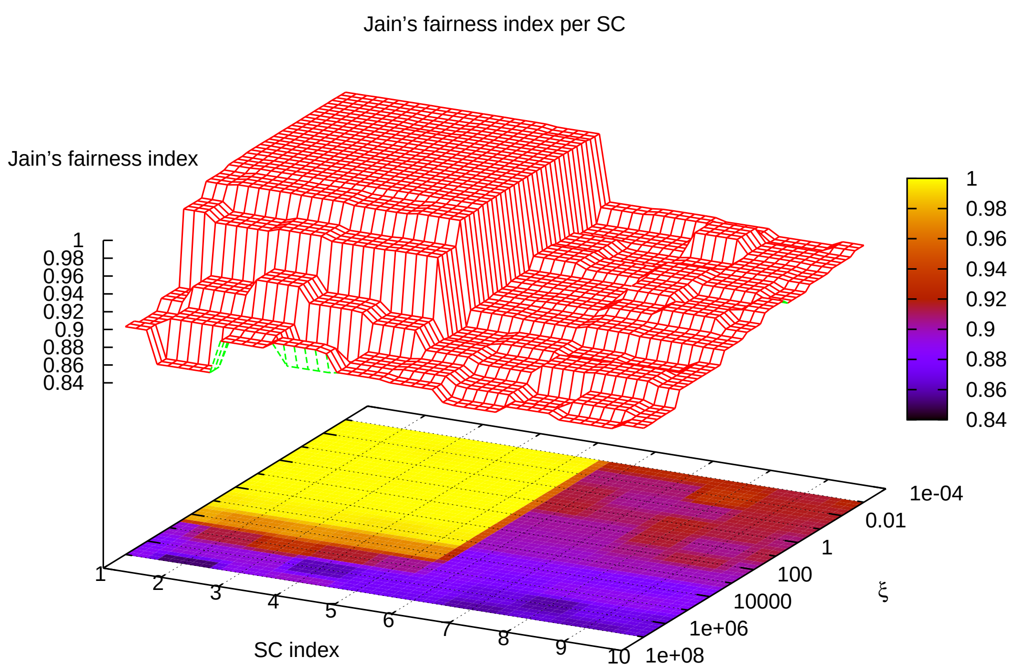


Figure 7.30: Jain's fairness index per SC as a function of ξ

The level of fairness is successfully controlled by ξ: for small values of ξ there is almost perfect fairness. For high values of ξ, we find out the initial radio conditions chosen of this scenario. Each point represents an average over 1000 sent data frames in the system.

Figure 7.31 presents the obtained performance when ξ varies.

As seen in figure 7.31, the lower the level of fairness (ξ increases), the lower the number of iterations and the better the capacity, because the best nodes are favored.

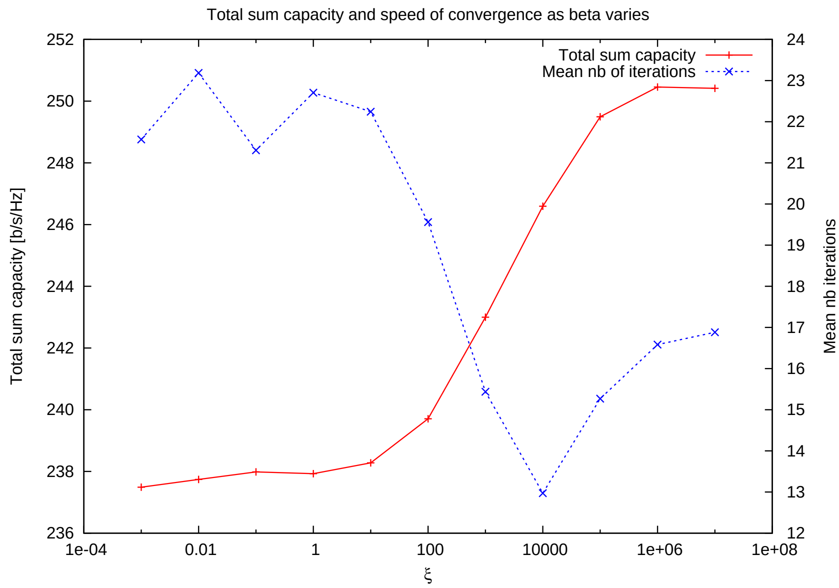


Figure 7.31: Total sum capacity and speed of convergence as a function of ξ

Accordingly, the level of fairness can be effectively controlled by the single parameter ξ and its impact on the performance has been evaluated. There is a tradeoff between ensuring a good level of fairness and having maximum capacity for the system.

7.8.4 Conclusion

In this section, we presented a distributed time-based fairness algorithm added on the top of a collaborative spectrum negotiation algorithm. The operator is able to control the desired level of fairness for his system by tuning the parameter ξ. Then, nodes can self-control that this operator-required level of fairness is ensured. This can be motivated to adapt to several conditions of channel speed and amplitude of variation. Such an algorithm is very efficient for next generations of WLAN systems, to opportunistically use the appropriate part of the spectrum according to the nodes' needs and with a fairness control.

7.9 Scenarios with 2 Competing Classes of Service

In this section we present the results concerning several scenarios with 2 competing classes of service. Three scenarios have been studied. In each scenario, the classes of service have their own set of n_i requirements as shown below:

- Scenario 1: in this scenario Service 1 and Service 2 have fully overlapping n_i requirements such that:
  - Service 1: $n_{min} = 1$ and $n_{max} = 6$.

- Service 2: $n_{min} = 1$ and $n_{max} = 15$.

- Scenario 2: in this scenario Service 1 and Service 2 have partially overlapping n_i requirements such that:
  - Service 1: $n_{min} = 1$ and $n_{max} = 10$.
  - Service 2: $n_{min} = 5$ and $n_{max} = 15$.

- Scenario 3: in this scenario Service 1 and Service 2 have fully non overlapping n_i requirements such that:
  - Service 1: $n_{min} = 1$ and $n_{max} = 6$.
  - Service 2: $n_{min} = 7$ and $n_{max} = 15$.

Each class of service could correspond to a different service with specific requirements in amount of spectrum to use. As explained in section 7.5.2, the n_i constraints correspond to the minimum n_{min} and maximum n_{max} number of SCs allowed per node for DATA transmission. Each scenario is expected to give different results, as different constraints are present.

The numerical values used in these simulations are summarized in table .

Notation	Description	Values
ξ	Parameter controlling the level of time-based fairness in the system	1
n_{min}	Minimum simultaneous number of SCs used per active node	varied per service
n_{max}	Maximum simultaneous number of SCs used per active node	varied per service
M	Number of nodes	100
N	Number of available SCs to allocate	20
Penetration rate: we always have 50The simulation duration corresponds to a total of $K = 10000$ *transmitted DATA frames. The SNR conditions change after each DATA frame transmission, regardless of the service used*		

Table 7.11: Numerical Values for the Simulation Parameters (refer to figures 7.32 to E.8)

The results are organized by scenario. For each scenario, we compare the results obtained without fairness with the ones obtained with fairness. The time-based fairness mechanism used is similar to the one presented in section 7.8. In each case, the following results are provided:

1. The total number of times each node used a group of i SCs for its DATA Tx.
2. The total number of SCs used by each node for DATA Tx, aggregating all the cases when using a group of i SCs. As an example, a node using i SCs for a DATA Tx would add i to this number.

The figures are organized as follows:

- Figure 7.32 presents the results for scenario 1 without fairness.

- Figure 7.33 presents the results for scenario 1 with fairness.
- Figure 7.34 presents the results for scenario 2 without fairness.
- Figure 7.35 presents the results for scenario 2 with fairness.
- Figure E.7 presents the results for scenario 3 without fairness.
- Figure E.8 presents the results for scenario 3 with fairness.

The following main conclusions that can be drawn from the results presented in figures 7.32 to E.8 are:

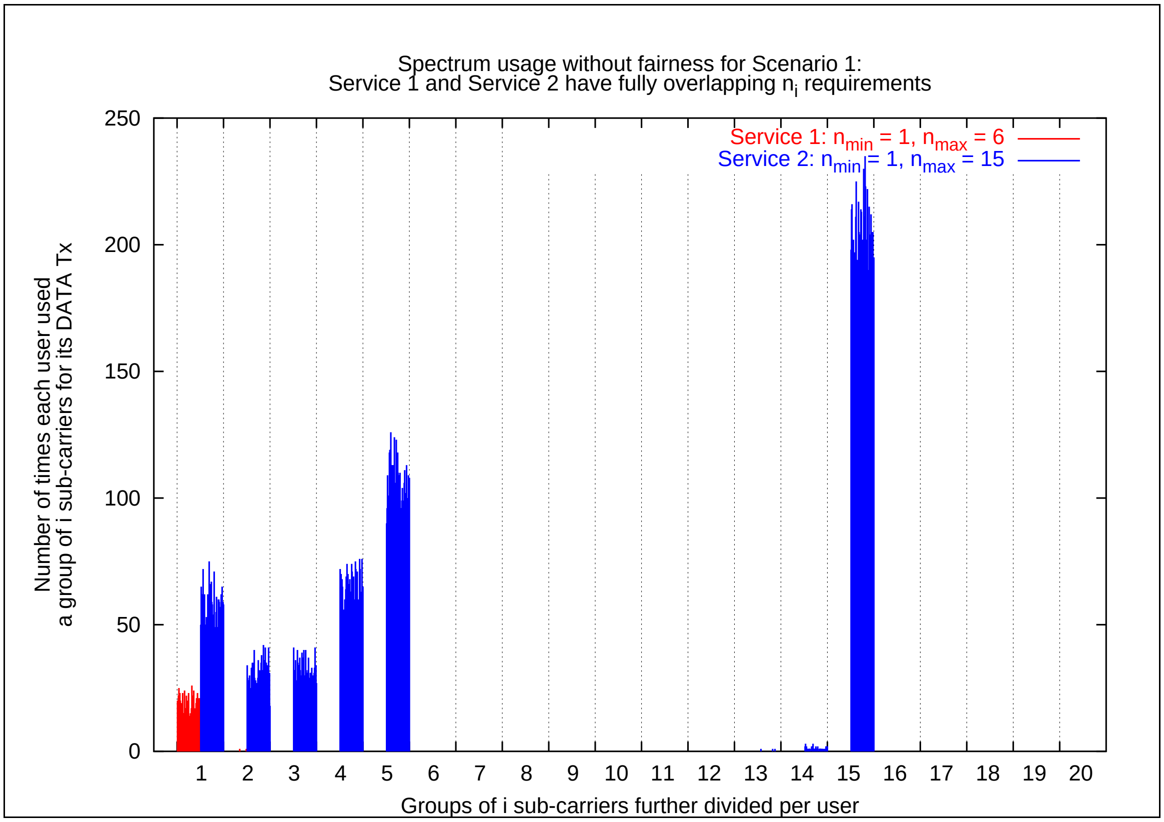
- Without fairness the competing classes of service do not properly coexist even if in each class of service all the nodes reach approximately the same total amount of spectrum resource. Indeed, some allowed values of n_i are sometimes not even reached by any node from a given class of service.
- Adding the time-based fairness was fully beneficial to have a proper coexistence of the services, each with its own set of requirements. Each of the three scenarios was successfully corrected by adding the fairness. Also, each node received approximately the same total amount of spectrum resource as the others from the same class of service as well as from the other class of service. Indeed, the time-based fairness is managed by SC for each node.
- Note that a stricter level of fairness (smaller ξ value) as well as a longer simulation duration would allow to better equalize the results between all the nodes, but would not change the conclusions.

Accordingly, the algorithm was successfully able to distribute the spectrum resources for UL transmission between several nodes each one with its own set of requirements. In addition, a great result is that each node does not need to know how many other nodes are simultaneously contending, and how many different classes of service (and associated spectrum requirements) are present in the cell. This was achieved in a completely distributed fashion.

7.10 Comparison with Other Methods

In this section we compare our method with other methods in terms of total reached UL sum capacity after completing an allocation. This will be evaluated while the total number of nodes M in the system varies. The following methods are used for comparison:

- *Max-Total*: it corresponds to the allocation of N values at c_{max}, which is where the system tends for an infinite number of nodes when using multi-user diversity. But this level cannot be reached in reality, it is intended to give an upper bound for the total sum capacity.
- *OFDMA*: this is our distributed method.



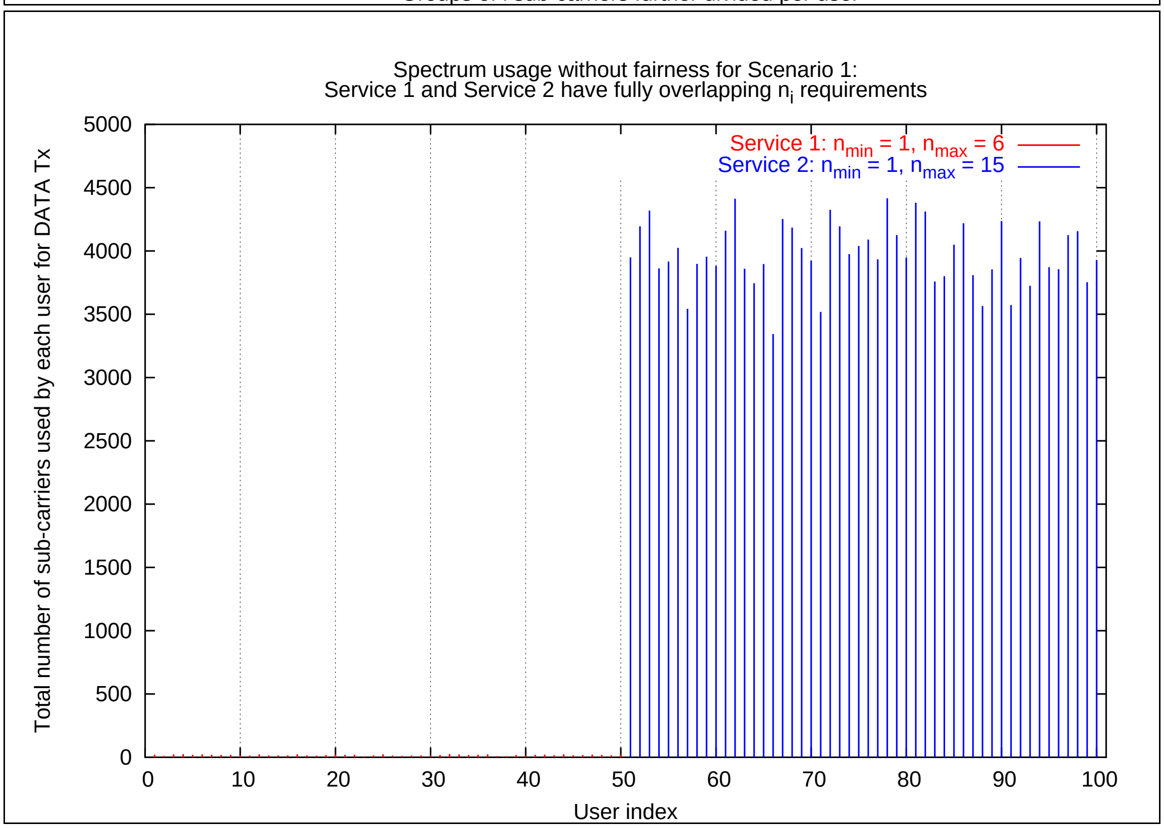


Figure 7.32: Results for Scenario 1 Without Fairness

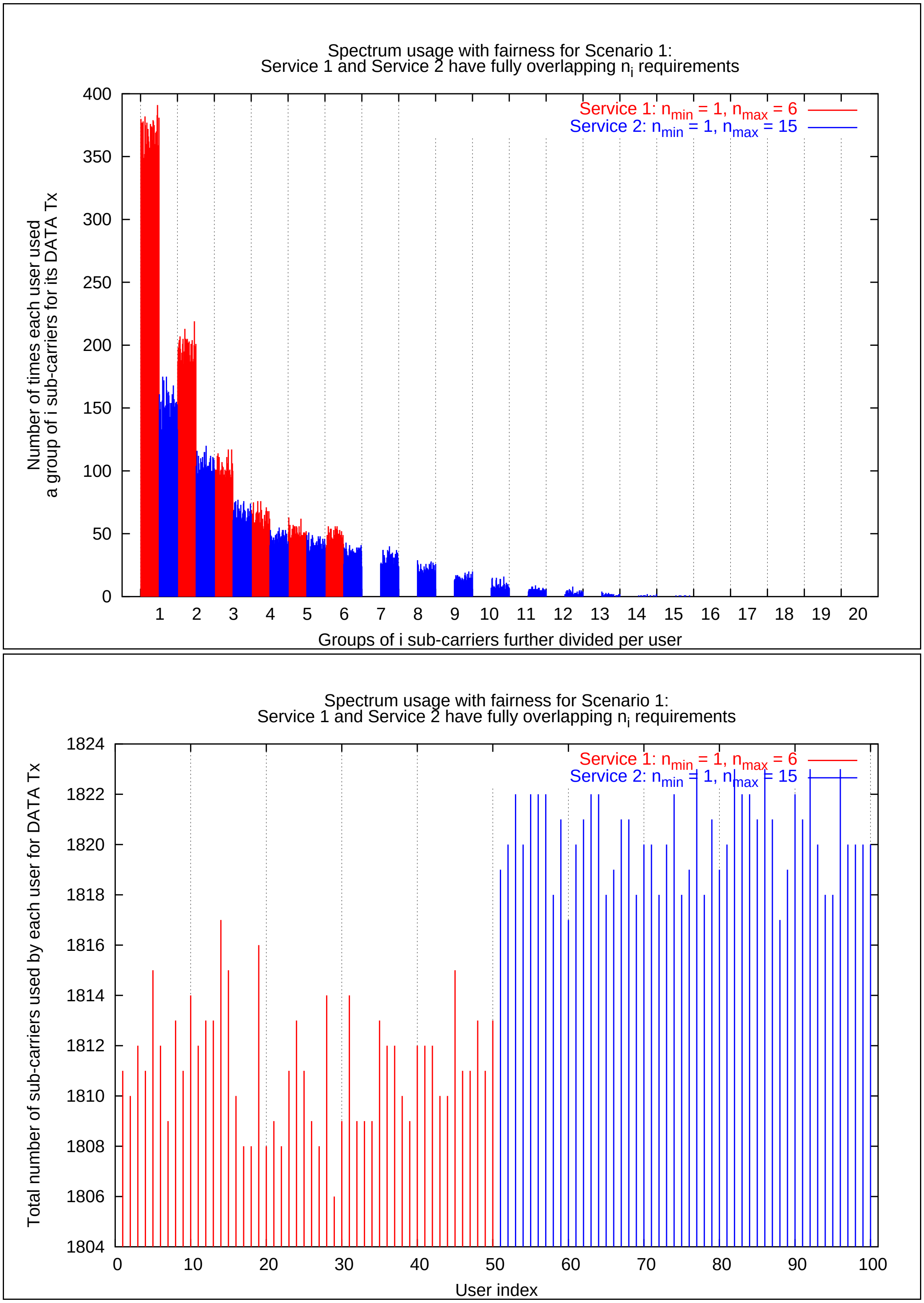


Figure 7.33: Results for Scenario 1 With Fairness

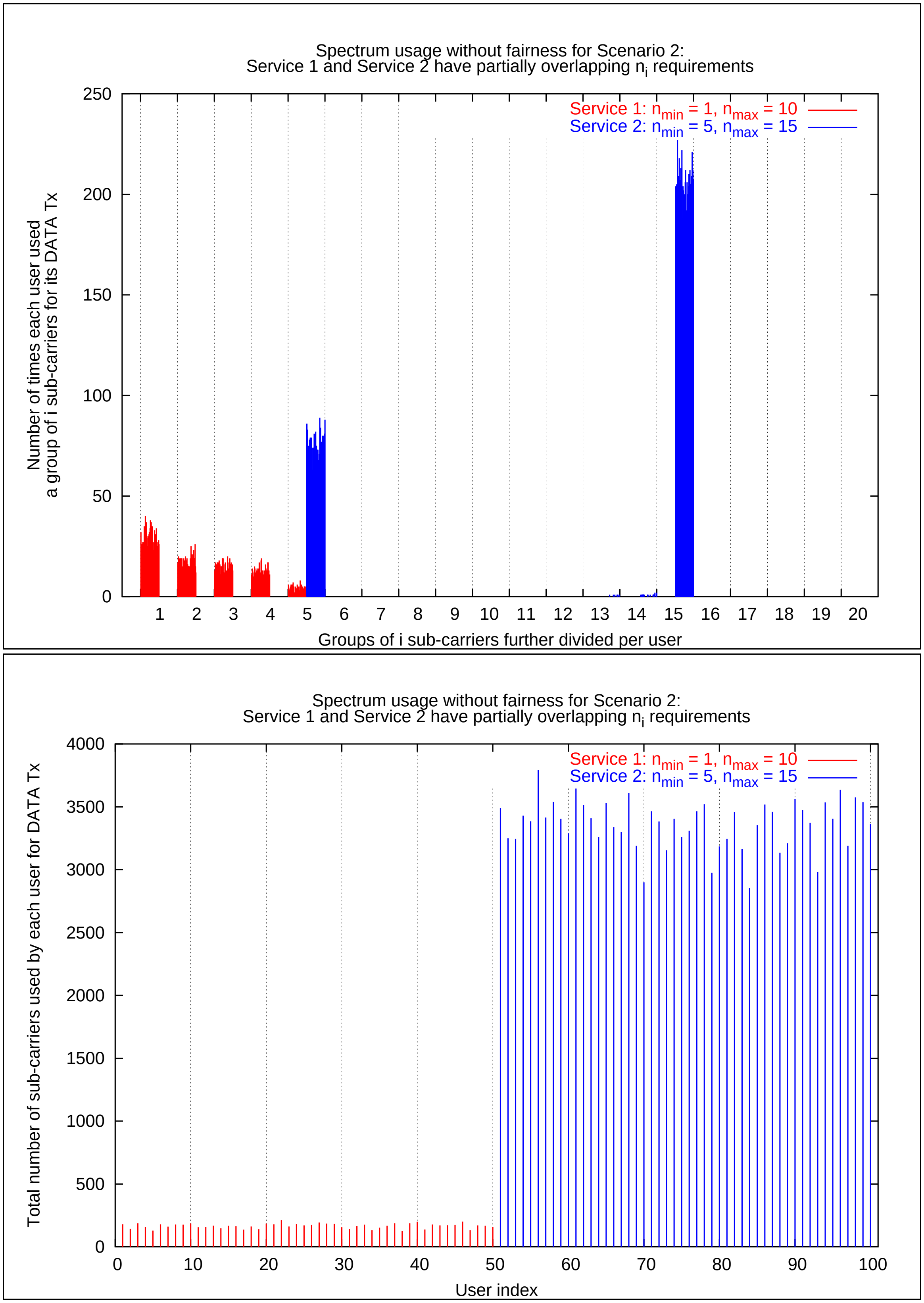


Figure 7.34: Results for Scenario 2 Without Fairness

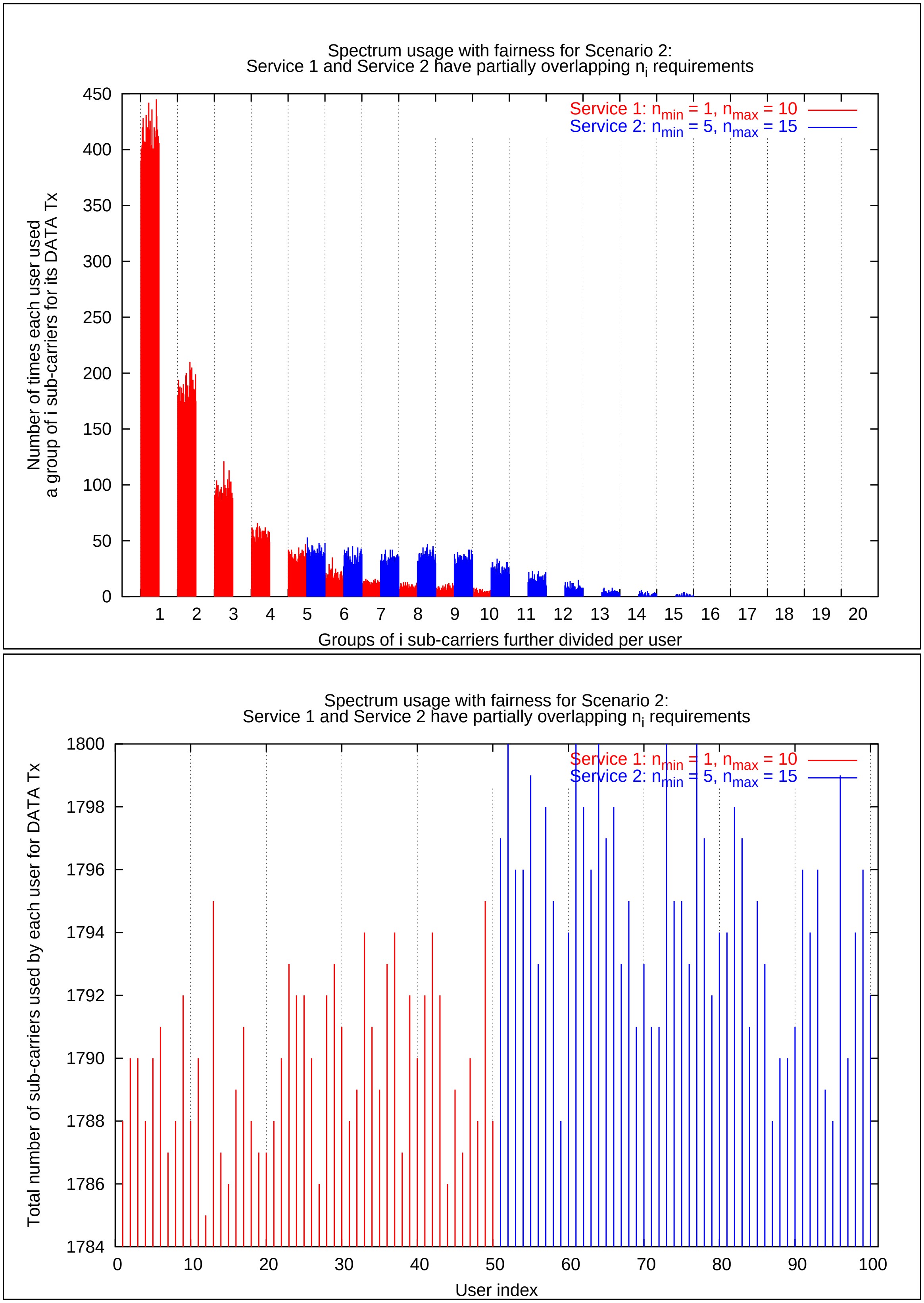


Figure 7.35: Results for Scenario 2 With Fairness

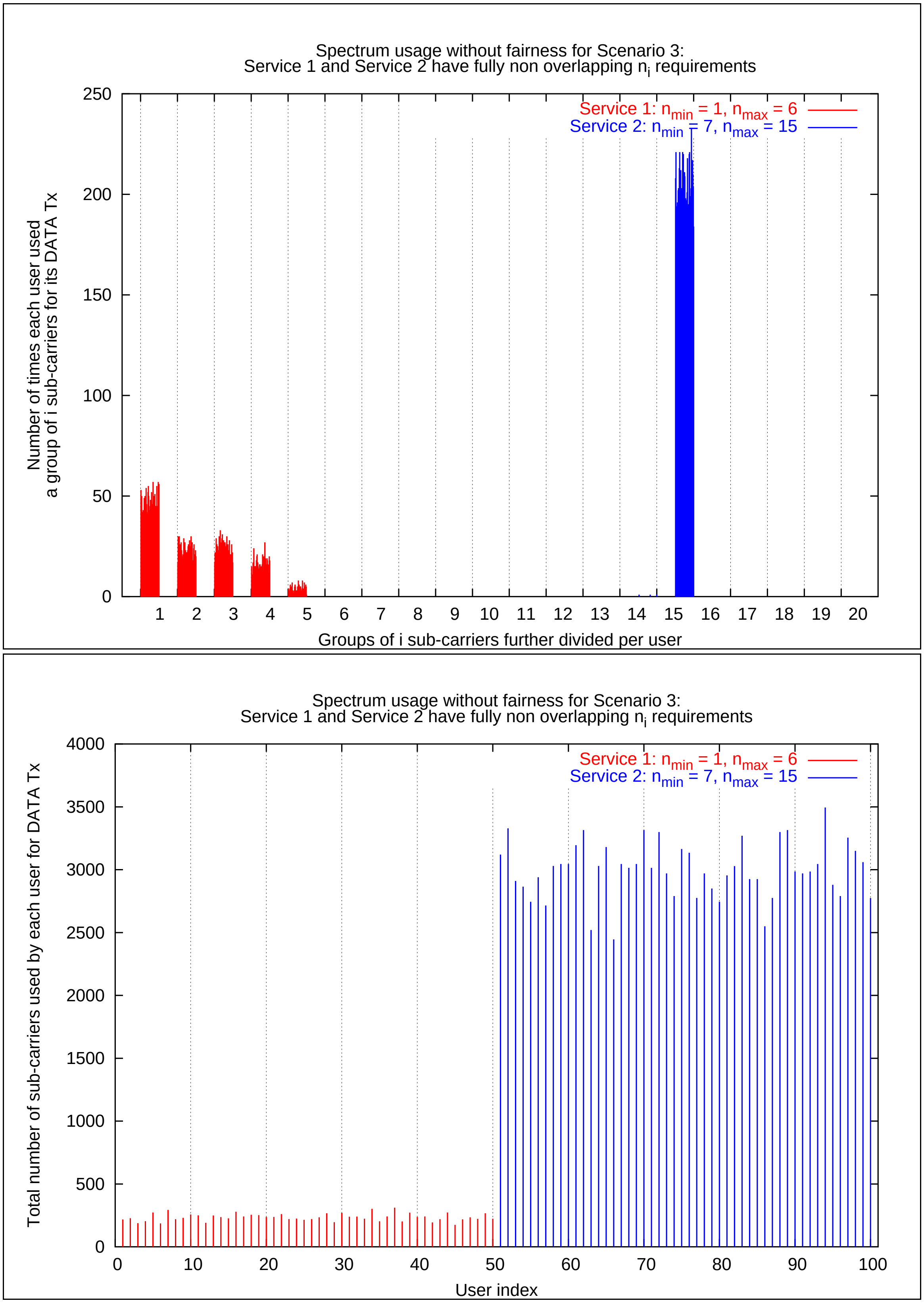


Figure 7.36: Results for Scenario 3 Without Fairness

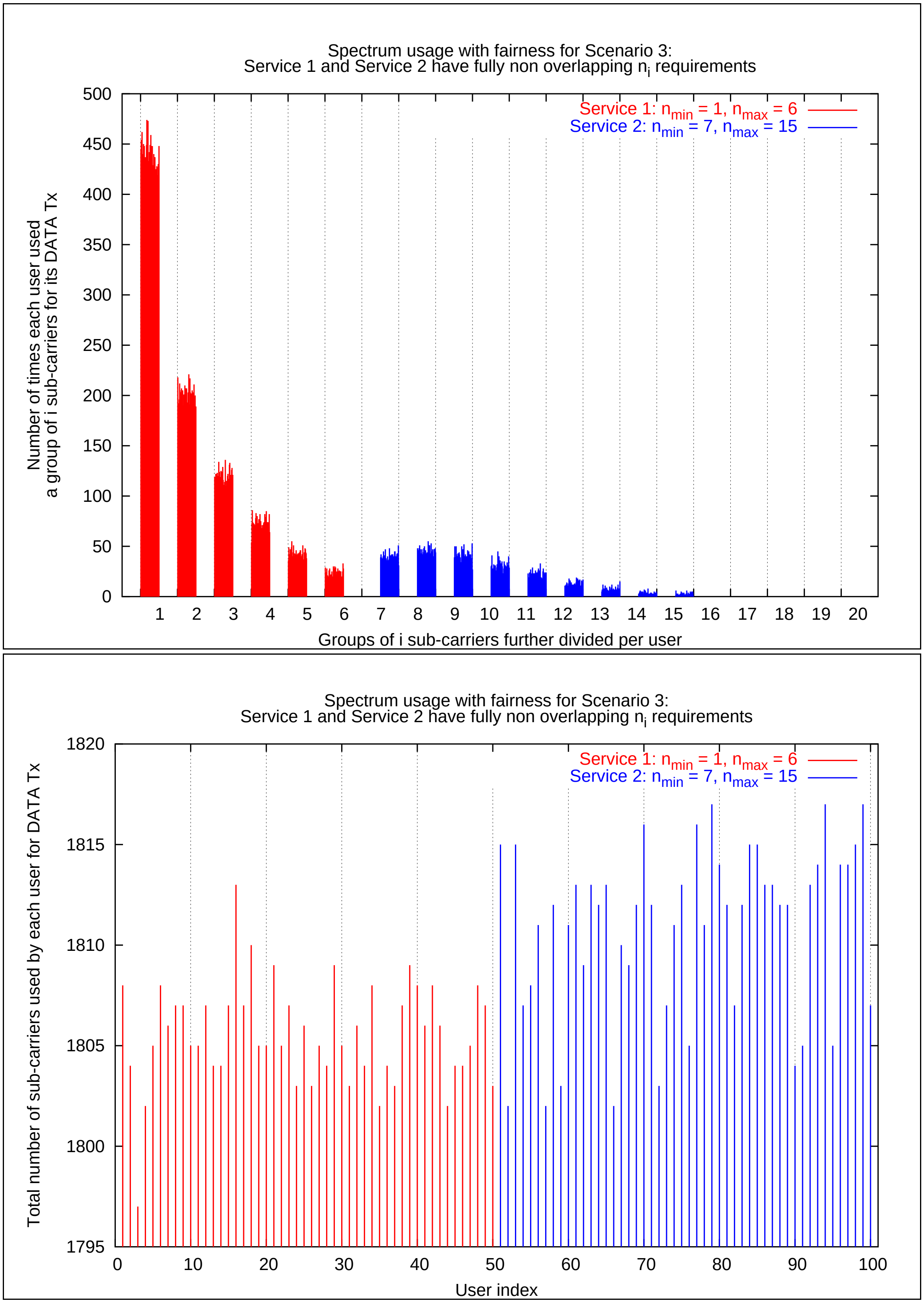


Figure 7.37: Results for Scenario 3 With Fairness

- *Random*: for each SC, a random node is taken to be allocated to this SC. In other words, there is absolutely no intelligence in this choice. It is similar to doing a "blind" choice. It is expected to provide a lower bound for the total sum capacity.
- *Max-PerNode-OverAllNodes*: for each SC we allocate the node having the maximum c_{ij} value over all the other nodes. Then, this node is removed from the future choices, and the decision goes on with the next SC until the complete allocation of all the SCs.

The numerical values used in these simulations are summarized in table .

Notation	Description	Values
n_{min}	Minimum simultaneous number of SCs used per active node	1
n_{max}	Maximum simultaneous number of SCs used per active node	1
M	Number of nodes	$[20:10:100]$
N	Number of available SCs to allocate	20
A total of $K = 1000$ *DATA frames are transmitted for each* M *value*		

Table 7.12: Numerical Values for the Simulation Parameters (refer to figure 7.38)

For each M value, all the methods calculate the mean total UL sum capacity as a result of the K transmitted DATA frames.

Figure 7.38 presents the obtained results.

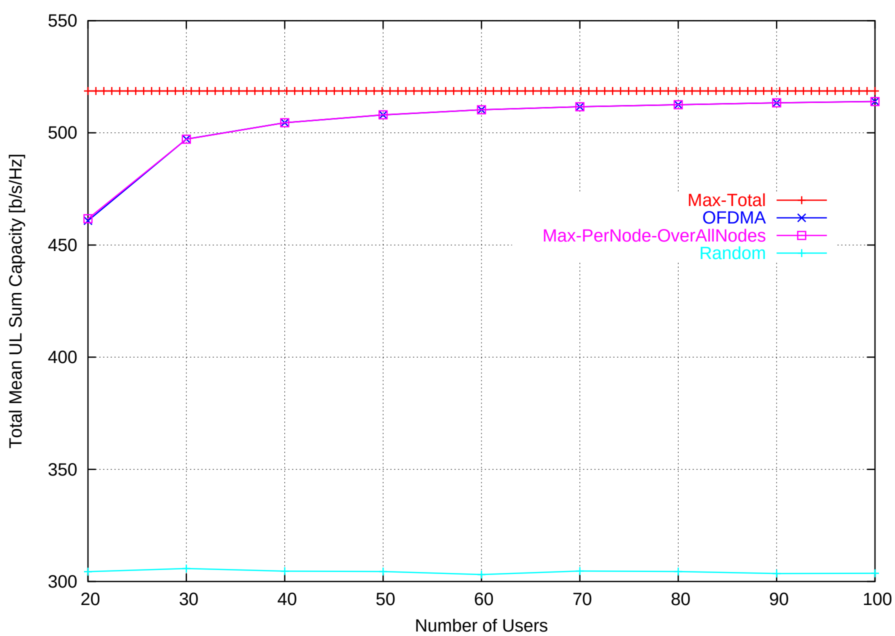


Figure 7.38: Comparison of Methods for Total Sum Capacity for a Varying M

As expected, the *Random* method constantly gives the worst results. In addition, both the *Max-PerNode-OverAllNodes* and the *OFDMA* methods result in an increasing capacity as M increases due to having more nodes with better channel conditions. These two methods asymptotically tend to the *Max-Total* level representing the best possible results. The methods *Max-PerNode-OverAllNodes* and *OFDMA* give comparable results in this scenario. However, the *Max-PerNode-OverAllNodes* method would correspond to a centralized method, whereas our *OFDMA* method is a distributed method in which nodes do not know about each other. Accordingly, our algorithm gives very good results when compared to some other methods and can be implemented in a distributed way.

7.11 Conclusion and Future Opportunities

In this chapter our collective UL OFDMA solution for performance improvement was exposed and studied in details. We have described the modeling challenge achieved to adapt the swarm intelligence to an OFDMA context. The main algorithm's control parameters have been varied and their impact on the performances have been evaluated. As seen from the results in this chapter, our distributed algorithm does not have a single behavior but can be varied by changing some of the main control parameters to better adapt to the problem at hand.

The key concept in this bottom-up approach is to control the system from the inside. The complexity of the system can thus be recreated as a result of the many but simple interactions taking place between the lowest components of the system. In other words, the complexity is not held by each node's device. The burden of computation can thus be distributed between the many nodes.

In this study, we tried to minimize the coordination overhead between the nodes, making use of a parallel reservation mechanism and a variable quantification window to limit the UL negotiation timeslot duration. This corresponds to doing cross-layer optimization. We strongly believe that great gains can result from doing cross-layer optimization instead of several sub-optimizations at each layer.

By simulations we validated the following interesting properties for our algorithm:

- Robustness against changes in the environment (number of nodes, channels changes).
- Flexibility: the algorithm can adapt to various configurations of M, N, n_{min}, n_{max} and various services.
- Scalability in number of simultaneously contending nodes: the convergence time remained reasonable and varied not significantly with an increasing number of nodes.
- Efficiency: the quality of the solution was very good when compared to other methods. The speed of convergence is reasonable considering the size of the considered problems. In addition, the convergence can be accelerated by adjusting the appropriate control parameters.
- Fair: The algorithm can control the fairness between the nodes using a time-based fairness mechanism.

Our algorithm, based on the swarm intelligence approach, has a great potential for future generations of wireless systems using cognitive radios. Indeed, it is an algorithm of dynamic spectrum allocation that can be implemented in devices to better adapt the spectrum use according to the nodes' needs in a network.

Chapter 8

Conclusions and Suggestions for Future Work

8.1 Conclusions

Since a few years, using a mobile phone is no longer solely limited to business people, but spans the whole range of ages from children to older people. Indeed, more and more people are using wireless means of communication to exchange information, for different purposes. The possibility of communication on the move is one major advantage of wireless communications over wired ones. Wireless users always ask for an increasing number of services, a better QoS, and a service availability anytime and anywhere. Note that wireless exchanges of information are also being extensively used by several domains for diverse applications such as: stock management, supermarket price display management, airplane landing system (GPS), search and rescue, etc. All the wireless applications use spectrum to work and are subject to spectrum reuse constraints to avoid radio interference. Considering the above as well as the fact that the earth's population has an estimated increasing rate (for the period 1994-2025) of 1.3% per year, thus the wireless community urgently needs to find innovative solutions to expand the already congested spectrum bands.

In most countries, the currently allocated spectrum bands are almost already saturated, while measurements campaigns have shown their sparse use both in time and space. In other words, only a small portion of the allocated spectrum is used everywhere and all the time. As a consequence, the so-called "spectrum scarcity" exists only in the spectrum allocation paper chart but not in real. Accordingly, the new regulatory spectrum access rules to develop should be more flexible to better reflect the possibility (as well as the technical capability) to use spectrum according a system needs (where and when needed). Of high importance, regulators should be open to reflect in their regulatory framework the current scientific and technical evolutions. Especially, they should be very careful when setting up regulatory rules in which the spectrum resource is considered as a material resource. Such a possibility should be only temporary, waiting for the next technological breakthrough, and should certainly not stand forever or be considered as an intimate property of the spectrum. Nowadays, it is technologically possible to detect spectrum holes and occupy them, even for a short period of time. However, before allowing such an opportunistic spectrum use, regulators expect insurances that no harm (interference) will affect the primary users (the license holders of a considered spectrum band) coming from the secondary users (the non-licensed holders willing to opportunistically use the considered licensed spectrum band). Providing intelligent algorithms to make the best use of a spectrum part is what we have proposed in this thesis.

Historically, cellular operators have been assigned spectrum bands by regulators (usually limited in size, in parts due to its important cost) to use a given technology. Then, operators have tried to maximize the spectrum use by applying frequency reuse planning. However, nowadays, this centralized "command and control" rule imposed on operators have created an artificial spectrum scarcity, which in turn also impacts the regulators themselves due to highly congested spectrum allocation charts. Accordingly, regulators are better eager to look at new ways of allocating the spectrum. The new perspectives of spectrum allocation and use such as: spectrum trading, spectrum exchange, spectrum pool, will be rendered successful with the development of the cognitive radios (smarter devices).

In this thesis, we have compared several methods (dynamic, fixed) of assigning channels to users over time. This was achieved using the great analytical tool of queuing theory. Among all the studied methods (based on a FIFO queuing policy), the highest system throughput came from the method controlling the channel access at the user level as a function of the channel conditions. The benefit comes from a dynamic and negotiated ordering of the users to access idle channels, leading to the best possible system throughput. Next, we present a more realistic context than the FIFO waiting policy.

In this thesis, we have successfully explored a part of the DSA challenge using a collaborative approach for negotiating spectrum resources between users. It gave good results and especially we managed, after modeling, to include the swarm intelligence's following properties in spectrum allocation between several users using several SCs: scalability, robustness to change, flexibility in negotiation. Owing to the fact that the swarm intelligence is a meta-heuristic, it is thus possible to use it to solve other problems related to spectrum allocation. This work insisted on pursuing the work of seeing the wireless modeling as a bottom-up approach.

An important effort was necessary to adapt the swarm intelligence to an OFDMA context. We managed to provide a powerful distributed algorithm for achieving SCs allocation between several users. The algorithm has also the capability to control the fairness in spectrum use between the users using an adapted time-based algorithm. We used cross-layer optimization to design our algorithm so that we could minimize the negotiation duration while testing many different combinations of users on SCs. Accordingly, with this algorithm we have successfully contributed to giving an answer (among other possible) to the question (DSA challenge): how can we achieve a self-adapting spectrum allocation between many users using many SCs, such that the spectrum will be the best allocated to the users in accordance with their needs? We found the swarm intelligence approach very powerful for solving DSA optimization problems in an OFDMA context.

8.2 Suggestions for Future Work

It could be interesting to investigate a slight change in our optimization problem and to allow overlap between several transmissions, in a CDMA-like way. The algorithm would have to include in its choice the possibility to have sometimes overlapping transmissions on a part of the spectrum, in addition to what we already studied. The approach would have to be still cooperative.

As the swarm intelligence is a meta-heuristic it can be used to solve other problems from the wireless world. A possible axis of research could be to use the swarm intelligence meta-heuristic to solve inter-AP or inter-BS coordination problems. Note that other methods could be used such as the game theory, which is currently intensively studied by several researchers to solve problems from wireless systems. Currently, only a limited number of problems can be solved using the game theory. However, its mathematical framework is rapidly progressing and will soon propose interesting solutions to difficult problems. Anyhow, using a collaborative bottom-up approach for modeling systems is a winning path to follow in the future.

We strongly believe that tremendous progresses/gains cannot be achieved solely by improving the MAC layer. Indeed, the ultimate is to benefit at the upper layers from a better understanding of the intimate properties of electromagnetic waves. For me, there appear that something important is missing in our understanding of the electromagnetic waves which has led us to build "dumb receivers". A major discovery will surely take place in this direction in the future, thus creating a technological breakthrough in terms of achievable performances, as spectrum is the support for wireless communication.

Also, new discoveries in information theory and group communication theory will benefit to upper layers (including the MAC layer). Accordingly, cross-layer optimization is an important key for unleashing the full potential of spectrum for the future generations of wireless systems.

In addition, lateral thinking is an important state of mind to have, because it can bring a fresh new start to a domain, simply by looking at the discoveries from other domains and creating a bridge with your domain. As an example, lateral thinking have helped in developing several meta-heuristics which are inspired from nature or physics (simulated annealing, genetic algorithms, swarm intelligence, etc). There are several phenomena that we don't understand in nature, but which are already working properly!!! We need to gain a greater cross-domain interaction and gain a greater inspiration from natural systems or species.

"Spectrum is not like wood, even if it is hard to properly slice it!"

Anonymous

Appendix A

International Efforts on Swarm Intelligence

Several examples of successful attempts of using the SI for several applications are provided in this Appendix. The SI metaheuristic has previously being used, based either on the techniques of Ant Colony Optimization (ACO) [66] or Particle Swarm Optimization (PSO) [67], to solve both static and dynamic optimization problems:

- Static Combinatorial Optimization Problems (assignment-type problems):
  - Traveling Salesman Problem [66].
  - Quadratic assignment problem [68].
  - Dynamic Task Allocation and Scheduling Problem [69], [70], [71], [72], [73], [70], [71].
  - Graph-Coloring Problem [74].
- Dynamic Combinatorial Optimization Problems (routing problems found in communication networks):
  - Connection-Oriented Network Routing [75], [76].
  - Connectionless Network Routing (Internet-type networks) [77], [78], [79], [80].

As a real-world application example and to highlight the benefit of the SI, General Motors has already realized several millions of dollars of savings by solving a factory optimization problem [62] using the SI metaheuristic. The problem was a real-time optimization problem of assigning trucks to paint booths for a truck painting operation. They wanted to have a simple but efficient algorithm, while trucks were going out from the production lines, telling which truck should be painted by which booth, in a minimum amount of time and with a minimum amount of wasted paint.

The SI is creating a growing interest in both research and industry communities from several domains, even including the wireless community. Testifying to that effort are several international initiatives described below:

- Research groups in universities:
  - Université Libre de Bruxelles (Belgium), with Jean-Louis Deneubourg and Marco Dorigo.
  - Stanford University (USA), with Deborah Gordon.
  - CNRS, Toulouse (France), with Guy Theraulaz.
  - EPFL Swarm-Intelligent Research Group (founded in September 2003), with Alcherio Martinoli.
- Companies having used or studying the SI:
  - Deutsche Telekom AG (Germany), with Steffen Lipperts.
  - Motorola, at least with Jean-Christophe Dunat in the Motorola Labs (France).
  - General Motors.
- US-government funded organizations:
  - NASA.
  - DARPA supporting university projects through grants (e.g.: Bio-Networking Architecture Project).
- International workshops, conferences and journals:
  - International Workshop on Ant Algorithms (ANTS).
  - International Workshop on Biologically Inspired Approaches to Advanced Information Technology (Bio-ADIT).
  - IEEE Swarm Intelligence Symposium (SIS).
  - IEEE International Conference on Evolutionary Computing, Fuzzy Computing, and Neurocomputing.
  - IEEE Transactions on Evolutionary Computation.

Appendix B

Simulation Tool

B.1 Introduction

Nowadays, wireless users require increasing spectrum resource to handle services anywhere and anytime, with seamless continuity of service. Developing new techniques to better manage spectrum resource can increase its usage efficiency, reduce radio interferences and network congestions. As a result, the needs for studies dealing whether with the coexistence or the interoperability of multiple RATs, whether with the performance of a single RAT at the system level keep on increasing. Radio oriented system level simulation models are necessary for spectrum management research to validate and reinforce theoretical studies at the user, cell and network levels prior to deployment. In addition, they help in identifying issues (interference, radio resource access unfairness, etc) and providing solutions (radio resource management schemes, enhanced protocol, etc) to anticipate problems that could occur in real. As such, a simulation platform can help in increasing cost savings.

In this thesis, a multi-RAT simulation platform has been designed. The next sections will describe its architecture, its functionalities and to what extend it has been used in our research.

B.2 Multi-RAT Platform Organization

Our system level simulation platform fulfills the simulation needs and helps in studying (static or dynamic scenarios) the performance of a single RAT (isolated or co-existing with several radio systems), or the interoperability of multiple RATs.

Our platform is decomposed into two complementary tools as shown in figure B.1:

- A simulation tool: contains the models and performs the simulations.
- A display tool: helps in visualizing and exploiting the simulation results using a given software (e.g.: Matlab, Gnuplot, Excel, etc).

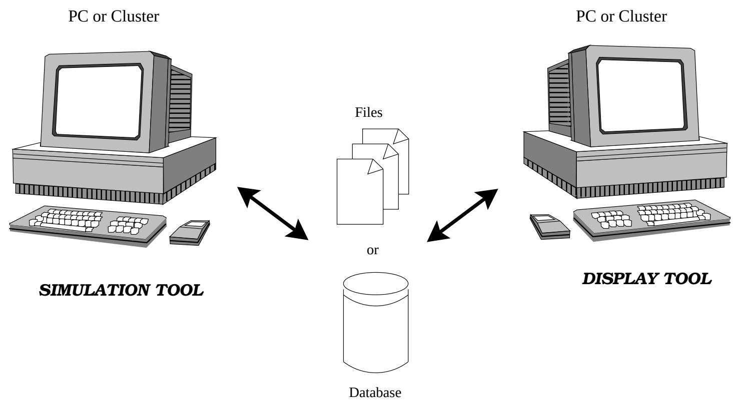


Figure B.1: Multi-RAT Simulation Platform Overview

Both tools communicate by exchanging text files formatted according to a given format common to both tools. Examples of simulation output datas, either collected over time or as distributions, and either at the user, the cell or the network level are included below:

- Radio parameters: coverage maps, desired and interfering signals received at some devices, etc.
- Protocol performance: number of packets sent, throughput, waiting delay before successful channel access, waiting delay to transmit a given payload size, etc.
- Etc.

The innovative aspects of this generic platform reside in the simulation tool itself. A challenging effort was gone through to get a fully modular and flexible platform architecture and to organize all the scenarios elements into several orthogonal libraries (evolving as independently as possible) corresponding to various levels of abstraction. A scenario selects several elements from several of these libraries. Figure B.2 presents a scenario organization involving these libraries.

This platform is targeted towards system performance analysis at the user, cell and network levels. The simulation tool uses an object-oriented approach (then coded using the C++ language for its rapidity of execution over Java and its ability to handle object-oriented modeling) to bring modularity (easy to use and validate), flexibility (easily upgradeable) and code re-use capabilities. Table B.1 gives more details about each library.

The next section describes more precisely the multi-RAT aspects of the simulation platform.

B.3 RAT Description

All along the lifetime of the platform, several RATs will be modeled and used either isolated or in a cooperative way. Accordingly, being able to introduce new RATs easily is a must. Having several RATs modeled in the same simulation would made it possible to study inter-system interference aspects or inter-system handover mechanisms.

Library Name	Library Description
System	Used to run the simulations: a PC or the cluster. Provides some general information regarding the simulation date and time, the computing activity, the system status, etc.
Simulator's Core	Organizes the simulation scenarios. Can be operated in two modes: static and dynamic. Manages all the different layers involved in the simulation. Examples: physical environment(s), communicating entity deployment(s) pattern(s), RAT(s), etc.
Geographical	All simulation scenarios use a simple orthonormal coordinate system to express all coordinates Example: simulated area.
Space	Some basic geometrical shapes have been defined (e.g.: point, vector, circle, rectangle, hexagon, etc). These simple shapes might then be represented as objects in upper layers (e.g.: a square could represent a 2D building in the physical environment layer).
Time	Defines clock counters with different levels of granularity.
Physical Environment	Usually of three types: urban (e.g.: Manhattan environment), sub-urban and rural. It is a complement to the propagation model, and does not incorporate any radio aspect but rather it may constraint users' paths. Examples: buildings, streets, etc.
Communicating Entity Deployment Pattern	List of all the communicating entities of the same type.
Infrastructure Deployment Pattern	List of the infrastructure equipments deployed in the simulated physical environment. Examples: network (list) of BSs, APs, etc.
User Deployment Pattern	List of the users distributed in the simulated physical environment. Their distribution can be: uniform, hot-spot, etc. Example: wireless users distributed in a city.
Communicating Entity	It represents the equipments supporting the radio communication and are of two kinds: radio access infrastructure equipment and user equipment.
Infrastructure Element	It is a kind of communicating entity. An infrastructure element is having a location, is attached to a RAT and is able to support all the implemented RAT services. Example: BSs, APs, etc.
User Element	It is a kind of communicating entity. Each user has a unique ID, a mobility profile, a service profile and a RAT ID. Examples: UMTS user, WLAN user, etc.
Mobility Profile	Characterized by position, speed and direction. Various mobility profiles are defined using specific rules on the evolution of speed and direction over time. Examples: along streets, random move, stationary, etc.
Radio	Contains several propagation models. Examples: free space, Hata, shadowing, Manhattan Berg, etc.
RAT	This library is detailed in the following section. Examples: UMTS, WLAN, etc.
Service Profile	A user is attached to a RAT for a particular service it offers. The infrastructure and the users have a service profile, either circuit or packet-switched to support voice or data. Examples of traffic models: simple saturated, Web, etc.
Mathematics	Some generic and extensively used mathematical functions have been implemented in this part. Examples: (e.g.: pseudo-random number generator, probability density functions, etc).
Input / Output	Contains various functions to manipulate input and output files.

Table B.1: Multi-RAT Platform Libraries Description

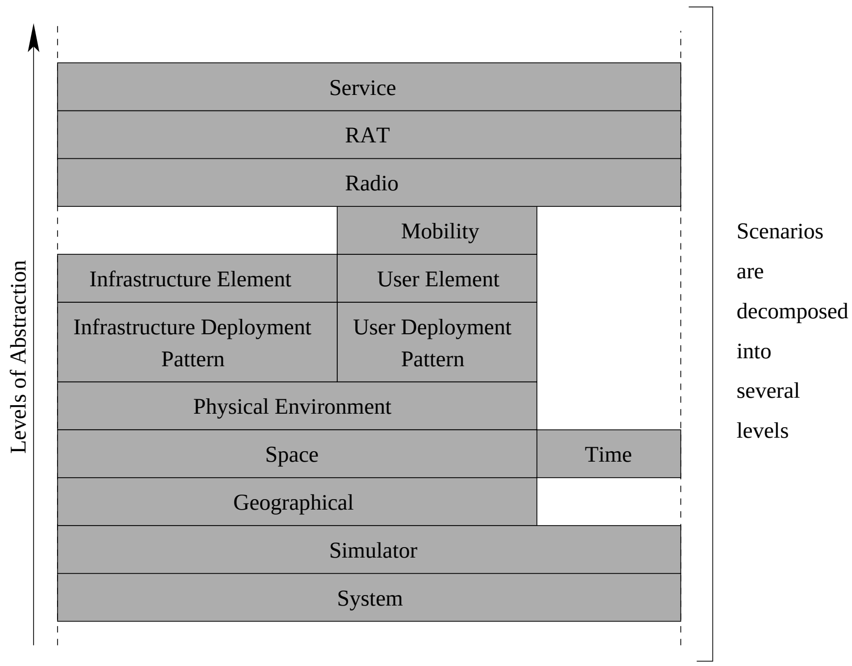


Figure B.2: Scenario Organization

Due to each RAT's specific parameters, finding a generic method to model several RATs in the same coding environment is a real challenge. Some methods help introduce real flexibility in the platform evolution such as the object-oriented philosophy, the Standard Template Library (STL), design patterns and template functions. For flexibility, simplicity and memory saving, we came up with the following architecture to efficiently organize a RAT's parameters, as well as with a mechanism for a user to be attached to a RAT.

A RAT is composed of a user part, an infrastructure part and a model for the RAT (protocol, specific functions). However, some features are sometimes RAT specific, making it inefficient to use a general model for all the RATs. Hopefully, the following method defines a common procedure working for all the RATs. Let us assume that a user, User i, is willing to be attached to RAT j, the following steps are gone through (also refer to figure B.3):

1. User i asks RAT j to register.

2. If it is possible, RAT j creates a copy of a user's RAT parameters for User i (it is a copy of the necessary parameters for a user in RAT j).

3. Then, User i is advertised that he is now attached to RAT j, and thus he updates his RAT address accordingly. The link between User i and his set of RAT's parameters in RAT j (copy) is achieved by his unique ID as well as by the RAT address.

When User i will no more be part of RAT j (disconnection), the copy of his RAT parameters will be removed from RAT j. This method allows a user to successively use several RATs without him being aware of the specific parameters used for him by each RAT.

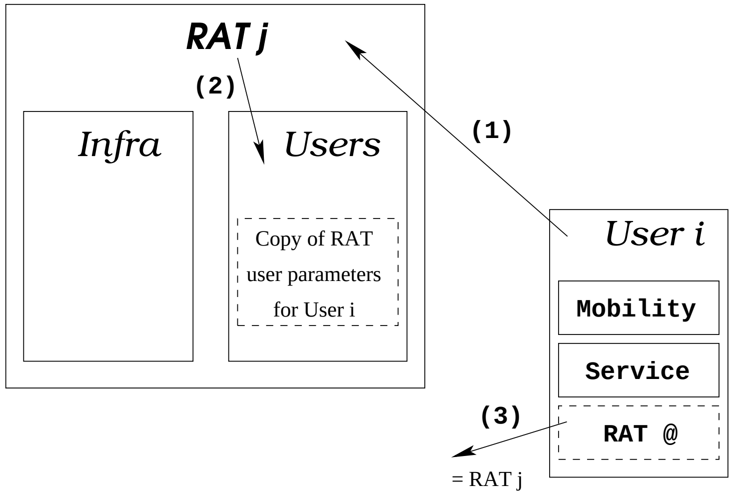


Figure B.3: User Attachment Procedure to a RAT

B.4 Simulation Platform in Use

Following the previously presented architecture, we have used the simulation platform for several studies during the PhD researches. We have only investigated dynamic simulations (not static) involving deployments of users and of a network of cells.

Even though included in the platform architecture, we have not tested scenarios involving multiple RATs in the same scenario. However, deployments of several cells and users have been investigated. The studies realized using the simulation platform are presented hereafter:

- IEEE 802.11 WLAN Model (see Appendix C and chapter 6): we have modeled the MAC and PHY layers of the IEEE 802.11 standard (especially the IEEE 802.11a) with the possibility to study the impact of the WLAN protocol on the performance results in the case of a deployment of several users in a network of multiple cells. The IEEE 802.11 standard uses a packet-based, asynchronous transmission of packets using several transmit modes. The results obtained were mainly focused on the impact of inter-cell interference on the user performance (frames' collision rate, mean wait time, throughput) at cell level, in a network of multiple unequally loaded cells. Indeed, several limitations exist with WLAN systems: hidden nodes, inter-cell interference, spatial unfairness, etc. As an extension, the current scenarios, 3G/WLAN scenarios could be investigated, assuming the 3G standard would be modeled in the platform fitting within the previously presented architecture.
- Simulator to validate the queuing model results (see chapter 4): in the preliminary analysis, we have validated our equations using the "queue" container offered by the STL and by interconnecting the different queues to create a network of connected queues. For this study, some functions were taken from the platform libraries.
- Distributed WLAN Model (see chapter 7): to implement our swarm intelligence distributed optimization algorithm, we have used some functions taken from the platform libraries.
- Etc.

B.5 Conclusion

The realized multi-RAT platform has been structured and extensively used during this PhD. More details are provided in article [81]. Each time, a given part of the platform libraries was used to build a simulation model.

Appendix C

IEEE 802.11 Simulation Model

C.1 Introduction

In this chapter, we describe the IEEE 802.11a/h model implemented for our WLAN studies. We have implemented the parts of the standard having an impact from a radio standpoint, as explained in the following sections. Each time, we have indicated the reference to the corresponding paragraph of the standard involved.

C.2 802.11a Simulation Model

This chapter presents the model of the 802.11a system implemented and used in the simulations. This model is as close as possible to the current 802.11a standard [1], [2]. However, some assumptions have been made. Indeed, the standard contains some "implementation dependent" parts which have been assumed to work a certain way. Unless otherwise specified, all the sections presented below hold true for both the AP and the user of a 802.11a network.

C.2.1 General Description

C.2.1.1 Network Architecture (§5.2 of [1])

A 802.11 network can be operated either in infrastructure mode or in ad-hoc mode. In ad-hoc mode the network forms an Independent Basic Service Set (IBSS) where MTs directly communicate with each other without the need for an AP.

All the simulations will use the infrastructure mode. In such as configuration a cell is equivalent to a BSS controlled by an AP. Usually, a network would be formed of several cells (each one controlled by an AP), and the APs would be connected through the same backbone (called Distribution System (DS)). The whole interconnected WLAN, including the different cells, their respective AP and the DS is seen as a single 802 network to the upper layers of the Open Systems Interconnection (OSI) model and is known in the standard as Extended Service Set (ESS). Figure C.1 presents an example of a 802.11 LAN architecture including the components described above and the two possible modes of operation.

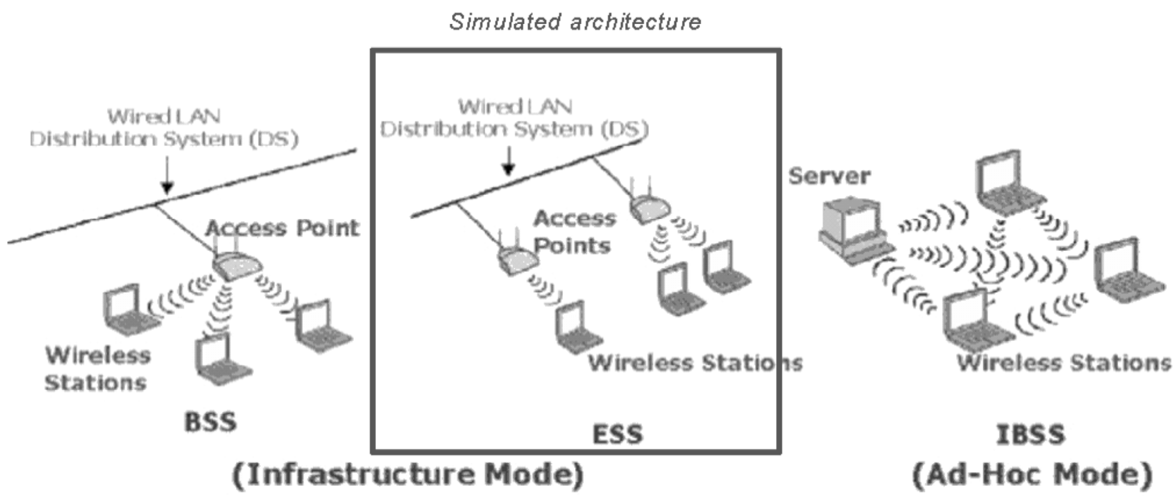


Figure C.1: 802.11 Network Architecture

The simulated architecture is located within the rectangle in figure C.1. No DS is considered in the simulations as we are only interested in the radio link between users and APs and not studying the routing protocols.

C.2.1.2 Reference Model (§5.8 of [1])

The 802.11 standard presents the architectural view, emphasizing the separation of the system into two major parts: the MAC of the data link layer and the PHY. These layers are intended to correspond closely to the lowest layers of the ISO/IEC basic reference model of OSI (ISO/IEC 7498-1: 1994). The layers and sub layers described in this standard are shown in the figure C.2.

Figure C.2 gives an overview of the data path though the layers.

C.2.2 MAC Layer

The MAC layer is mainly described in the initial 802.11 standard [1].

C.2.2.1 MAC Frames (§7 of [1])

In the 802.11 standard, the MAC frames can be of several types (management, control and data) and subtypes. A MAC frame is composed of a header, data part and a tail part (FCS).

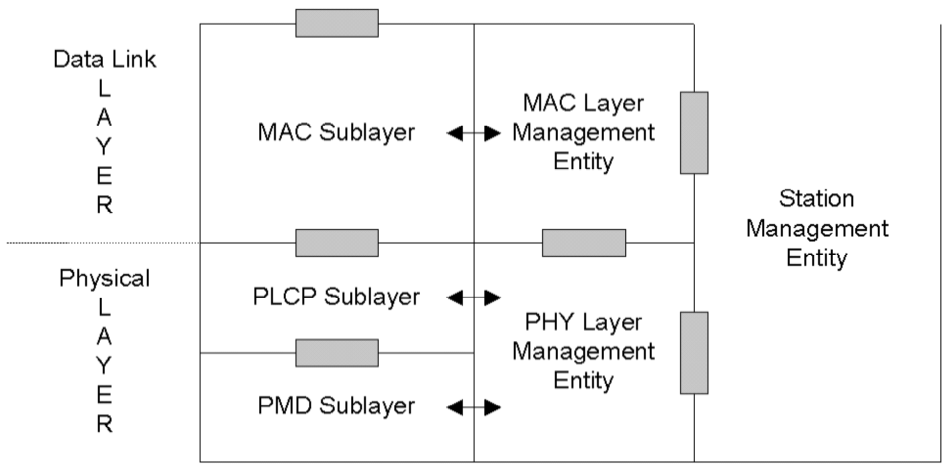


Figure C.2: Portion of the ISO/IEC Basic Reference Model Covered by the Standard [1]

The only MAC frame types and subtypes used in the 802.11a simulation model are presented in table C.1. The data part of the MAC frame may be of variable size and comes from the upper layers (applications). Depending on the frame type and subtype, the MAC frame header may be of variable size. (L is the data payload length in octets and is in the range of 0-2312 octets)

Frame Type	**Frame Subtype**	**Header size [octets]**	**Data size [octets]**	**Footer size [octets]**	**Total frame size [octets]**
Control	RTS	16	0	4	20
Control	CTS	10	0	4	14
Control	ACK	16	0	4	20
Data	Data	30	L	4	34 + L

Table C.1: MAC Frame Size

A MAC frame duration is PHY dependent. Section C.2.4 will present the case of the 802.11a implementation [2].

C.2.3 MAC Protocol (§9 of [1])

CSMA/CA - DCF

In the well-known Carrier Sensing Multiple Access / Collision Detection (CSMA/CD) protocol, a user listens to the medium before transmitting a data. However detecting the collisions is not always possible in a wireless medium due to hidden nodes. Also, as only one radio channel is used for both transmit and receive, a STA cannot Tx and Rx at the same time. Thus, the detection of a simultaneous transmission is not possible for the Tx, it can only be interpreted by no Rx of an acknowledgment. All the packets involved in a collision are lost at the reception. This is why the CSMA/CA mechanism is used.

The CSMA/CA protocol is a random access method based on the "listen before send" rule. All our simulations make use of the infrastructure mode (always an AP) and the compulsory DCF mode (see figure C.3), but not the Point Coordination Function (PCF) mode.

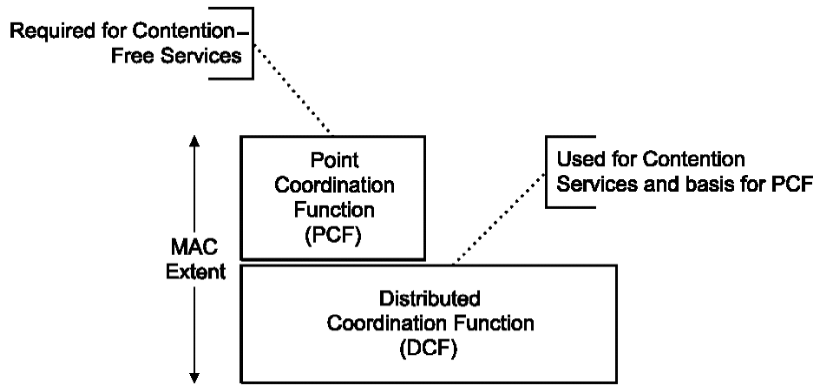


Figure C.3: MAC DCF / PCF [1]

In DCF mode, there is no central point of coordination in the cell, instead the coordination is distributed over all the STAs. In DCF, the AP has the same priority level to access the medium than every user of its BSS. For more details about the DCF protocol refer to the standard itself [1]. The simulations always make use the RTS/CTS mechanism. The threshold to choose between the basic access (no RTS/CTS mechanism) and the RTS/CTS depends on the data frame size to transmit. Indeed, below $RTS_{Threshold}$ (we set $RTS_{Threshold} = 1500$ bits as studied in the literature [54]) the basic access method is used instead of the RTS/CTS mechanism.

In DCF, the channel access and reservation method is said to be an ad-hoc slotted method, even tough a cell might be using the infrastructure mode. It is ad-hoc because the control is not centralized. It is slotted when the STAs decrement their backoff counter.

Using the 802.11 MAC protocol, a data transmission encompasses three main phases (as shown in figure C.4):

- *Channel Access*: After a null BC,
- *Medium Reservation*: RTS and CTS frames exchanges,
- *Data Transmission*: DATA frame (DATA) and ACK frames exchanges.

The protocol is based on the monitoring of the medium state by each STA. The medium state is either:

- IDLE = No current Tx in the cell. And it means a possibility to decrement the backoff counter and when it is null, the STA transmits its data,
- BUSY = Means that a current transmission is taking place in the cell. Thus, a STA will defer its BC decrement until the end of the current Tx plus a DIFS period of time.

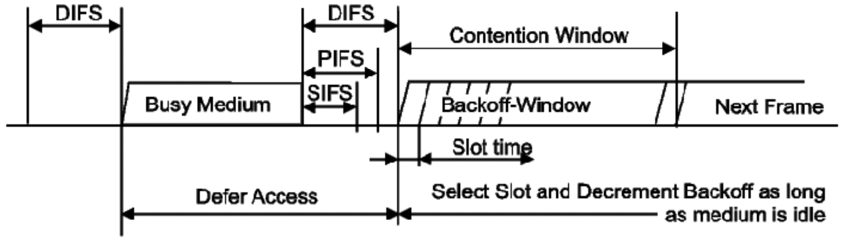


Figure C.4: MAC Protocol Phases [1]

The next protocol state and action is conditioned by the medium state around each STA. Due to a STA limited transmit range, the result of the carrier sensing operation is only valid for a certain region around the STA (hidden node problem), and thus is location dependent leading to some limitations as described in section 6.6.

NAV

The NAV mechanism is used in combination of the RTS/CTS mechanism (see figure C.5). In addition to the PHY carrier sensing operation, a MAC virtual carrier sensing is performed by all the non transmitting active STAs.

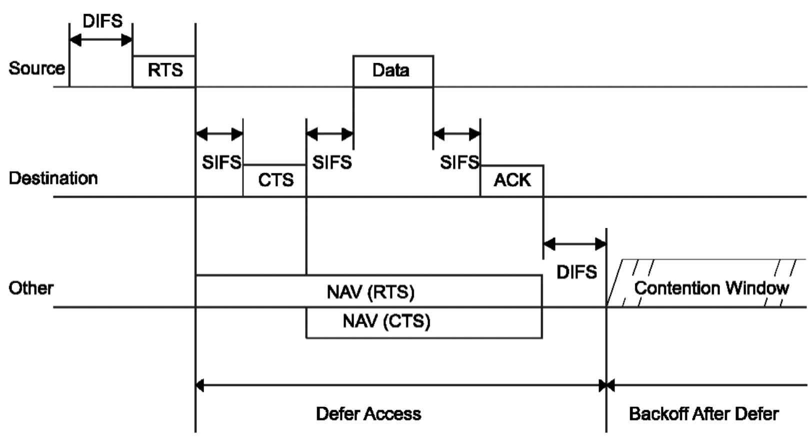


Figure C.5: NAV Mechanism [1]

The NAV information is included in all the transmitted frames (control and data) and is readable by all STAs (mode 1). The NAV information indicates the time when the complete data transmission (reception of a ACK frame) process is finished, corresponding to the start of the next DIFS wait period prior to the next contending period. As such, all the listening STAs become aware of the time for the next contention period of the medium.

The NAV duration is equal to the following value:

- RTS frame received: $T_{TX} = T_{SIFS} + T_{CTS} + T_{SIFS} + T_{DATA} + T_{SIFS} + T_{ACK}$
- CTS frame received: $T_{TX} = T_{SIFS} + T_{DATA} + T_{SIFS} + T_{ACK}$
- DATA frame received: $T_{TX} = T_{SIFS} + T_{ACK}$

T_x corresponds to wait time durations or frame time spent in the air. For more details and the parameters' values refer to the 802.11a standard [2].

BC Algorithm and Variable CW

The CSMA/CA protocol used by contending users is a random access method. Every period prior to a transmission is slotted (decomposed in time slots) due to the BC value. Thus, there is no pre-established transmission schedule. Each STA delays its transmission start time by some random BC value (generated according to a uniform distribution) within a variable CW. A failure to transmit exponentially increases the CW size in order to reduce the probability that two users or more draw the same BC value, and simultaneously transmit. This minimizes collisions between multiple STAs by temporally spreading their transmission start times. The duration of these waiting periods is random, and depends on the number of contending users and the medium state around each STA. The algorithm used is presented below:

$$CW_{n+1} = 2 \times (CW_n + 1) - 1$$

where n is the number of transmit attempts. The maximum number of transmission attempts is limited by counters as presented in the following paragraph (Short Retry Counter (SRC)/Long Retry Counter (LRC)). As a consequence, the MAC protocol waiting philosophy is the more a STA fails to transmit a frame, the more it waits whatever its data frame size and transmit mode".

SRC / LRC

Each unsuccessful attempt to transmit a frame increases a retry counter. There are two counters depending on the frame size (control or data frames):

- SRC = limits the number of transmission attempts for control frames and for data frames whose size is less than the $RTS_{Threshold}$,
- LRC = limits the number of transmission attempts of data frames whose size exceeds the $RTS_{Threshold}$.

After each successful transmission of a frame, the corresponding retry counter is reset. But after reaching the maximum value of either the SRC or LRC, the corresponding data frame is discarded.

RTS/CTS mechanism

To reduce the performance degradation due to hidden terminals, a medium reservation technique based on a reserve (RTS frame) and confirm (CTS frame) mechanism between the source and the destination is proposed in DCF mode. Once the medium is successfully reserved for a STA, the data frame can be transmitted with a higher chance of success. Also, when a collision occurs between several RTS frames, far less bandwidth is wasted when compared with a larger data frame collision. Thus, using the RTS/CTS mechanism is recommended when many users are contending for the medium to transmit large data frames. However, simulations show that the hidden node problem is not completely resolved by the RTS/CTS mechanism, even if it is greatly reduced.

C.2.3.1 MAC Sublayer Management Entity (§11 of [1])

Presentation of the MAC sublayer management entity:

- **a) Synchronization (§11.1 of [1]):** An ESS is composed of several BSSs each one controlled by an AP. BSSs are not synchronized in operation neither in time. However, inside a BSS all stations are supposed to keep the same time value, thus being synchronized with the AP they are associated with. Indeed, each AP regularly sends beacon frames containing an update of the time value, and solely intended for the BSS it controls. The mechanism of sending beacon frames is not modeled in the simulator because it is not directly linked with the transfer of data.

- **b) Power management (§11.2 of [1]):** The power save (sleep) mode capability for a STA is not modeled in the simulation model, thus, all the STAs are supposed to be always active. Also this mechanism might increase the battery autonomy this is not a topic studied in the simulations.

- **c) Association and reassociation (§11.3 of [1]):** Refer to sections C.2.5.2 and C.2.5.4 for more details.

C.2.4 PHY Layer

The 802.11a technology OFDM PHY layer is described in the 802.11a standard [2]. As shown in section 5.8 of [1], the PHY layer is decomposed into two layers: the PHYsical Layer Convergence Protocol (PLCP) and the Physical Medium Dependent (PMD).

C.2.4.1 Multirate Support (§9.6 of [1] and §17.3.2.2 of [2])

Some PHY layers (especially the OFDM PHY of the 802.11a standard) provide multiple data transfer rate capabilities that allow implementations to perform dynamic rate switching with the objective of improving performance. The higher the transmit mode, the faster the data transmission as well as the higher the necessary $\left(\frac{C}{I}\right)$ ratio. According to our LA algorithm (section C.2.5.3), a STA will always try to use the highest possible transmit mode, in order to achieve a better goodput. Table C.2 presents all the possible modes of operation. All the transmit modes are modeled in the simulation model except mode 2 because of its worse performance compared to the other modes. In our simulation model, all control frames (RTS, CTS, ACK) are always transmitted using mode 1, whereas data frames are transmitted using a mode driven by the link adaptation algorithm.

C.2.4.2 PHY Layer BSS Parameters

Table C.3 recapitulates the main BSS parameters used in the simulation model.

Mode	**PHY Data Rate [Mbps]**	**Modulation**	**Coding Rate (R)**	**Coded Bits per Subcarrier (N_{BPSC})**	**Coded Bits per OFDM Symbol (N_{CBPS})**	**Data Bits per OFDM Symbol (N_{DBPS})**
1	6	BPSK	1/2	1	48	24
2	9	BPSK	3/4	1	48	36
3	12	QPSK	1/2	2	96	48
4	18	QPSK	3/4	2	96	72
5	24	16-QAM	1/2	4	192	96
6	36	16-QAM	3/4	4	192	144
7	48	64-QAM	2/3	6	288	192
8	54	64-QAM	3/4	6	288	216

Table C.2: OFDM PHY RATE-Dependent Parameters

The explanations in table C.3 were found in §9.2.10, 10.4.3.2 of [1], and the corresponding values were found in §17.5.2 of [2]. Thus, these values are also applicable for the 802.11a extension.

C.2.4.3 PLCP Frame Format (§17.3.2 of [2])

This sub-clause provides a convergence procedure in which PLCP SDU (PSDU)s are converted to and from PLCP Protocol Data Unit (PPDU)s. During transmission, the PSDU shall be provided with a PLCP preamble and header to create the PPDU. At the receiver, the PLCP preamble and header are processed to aid in demodulation and delivery of the PSDU.

Figure C.6 shows the interaction of the PLCP and the PMD.

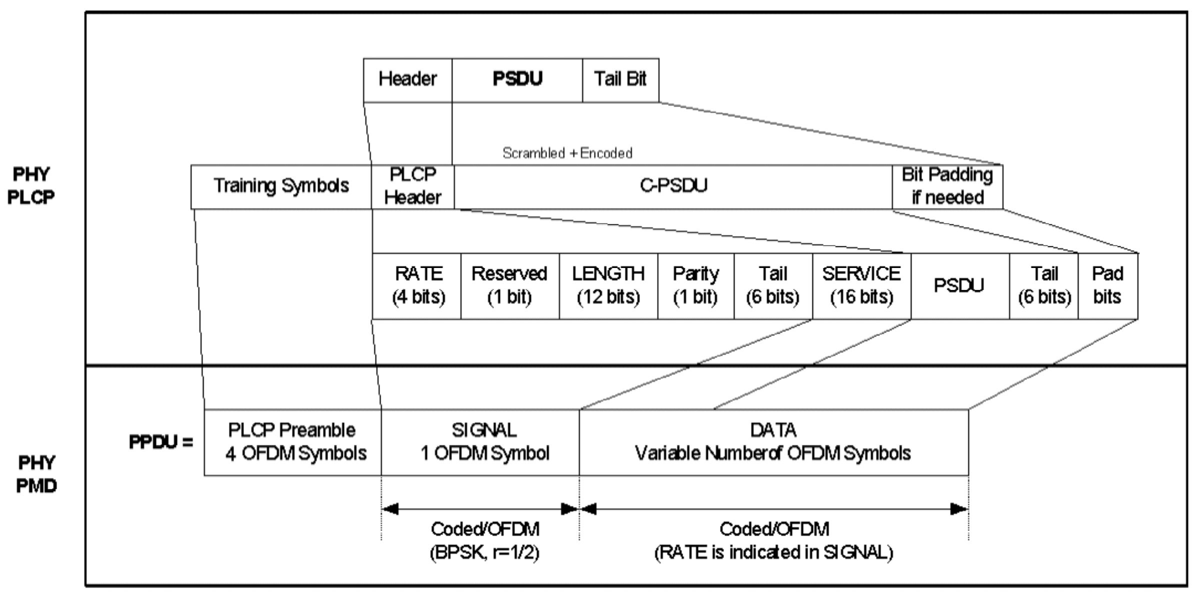


Figure C.6: PSDU and PPDU Frame Format of the OFDM PHY [2]

Name	Description	Value from OFDM PHY 802.11a standard
Air Propagation Time	The anticipated time (in μs) it takes a transmitted signal to go from the transmitting STA to the receiving STA.	$\ll 1$ μs, neglected
Slot Time	The Slot Time (in μs) that the MAC will use to define the PIFS and DIFS periods.	9 μs
SIFS Time	The nominal time (in μs) that the MAC and PHY will require to receive the last symbol of a frame at the air interface, process the frame, and respond with the first symbol on the air interface of the earliest possible response frame.	16 μs
DIFS Time	DCF interframe space. $T_{DIFS} = T_{SIFS} + 2 \times T_{TS}$.	34 μs
EIFS Time	Extended interframe space. $T_{EIFS} = T_{SIFS} + T_{ACK} + 13 \times T_{ODFM} + T_{DIFS}$	146 μs
CTS Timeout	Time to wait before considering that the CTS frame will never be received. $T_{CTS\_Timeout} = T_{SIFS} + T_{CTS} + T_{DIFS}$	94 μs
ACK Timeout	Time to wait before considering that the ACK frame will never be received. $T_{ACK\_Timeout} = T_{SIFS} + T_{ACK} + T_{DIFS}$	94 μs
CCA Time	The minimum time (in μs) the CCA mechanism has available to assess the medium within every time slot to determine whether the medium is busy or idle.	< 4 μs
CW Min	The minimum size of the CW, in units of SlotTime.	15
CW Max	The maximum size of the CW, in units of SlotTime.	1023
SRC Limit	Lower limit for a frame retransmission attempts	7†
LRC Limit	Upper limit for a frame retransmission attempts	4†
RTS Threshold	For data frames larger than this threshold, the RTS/CTS mechanism is used	1000 octets†

†Values based on articles readings, not found in the standard.

Table C.3: Main PHY Layer BSS Parameters

In the simulation model, PLCP PREAMBLE and SIGNAL are always Tx in mode 1 whereas the DATA part (only for data frames) uses a mode determined by the link adaptation algorithm.

C.2.4.4 PLCP Main Parameters (§17.3.8.1 and §17.3.2.3 of [2])

The equation below gives the time a packet spends in the air (the transmission time). The PHY layer reports this value to the MAC layer before sending the packet.

$$T_{Tx} = T_{PREAMBLE} + T_{SIGNAL} + T_{SYM} \times Ceiling\left(\frac{16 + 8 \times LENGTH + 6}{N_{DBPS}}\right)$$

where:

- N_{DBPS} is derived from the DATARATE parameter. (Ceiling is a function that returns the smallest integer value greater than or equal to its argument value.) LENGTH is the PSDU length in octets.
- $T_{PREAMBLE}$ = PLCP preamble duration = 16 μs,
- T_{SIGNAL} = Duration of the signal BPSK-OFDM symbol = T_{OFDM} = 4 μs,
- T_{SYM} = Symbol interval = 4 μs.

Using the formulas of section OFDM TXTIME calculation (§17.4.3 of [2]) and the MAC frames sizes presented in section MAC frames (§7 of [1]), each MAC frame duration can be deduced as presented in the table C.4.

Frame Type	**Frame Sub-type**	**Total PSDU frame size [octets]**	**PPDU frame duration depending on the mode used to transmit [μs]**	**PPDU frame duration range (mode 1 - mode 8) [μs]**
Control	RTS	20	$T_{RTS} = 20 + 4 \times Ceiling\left(\frac{182}{N_{DBPS}}\right)$	(52 - 24)
Control	CTS	14	$T_{CTS} = 20 + 4 \times Ceiling\left(\frac{134}{N_{DBPS}}\right)$	(44 - 24)
Control	ACK	14	$T_{ACK} = 20 + 4 \times Ceiling\left(\frac{134}{N_{DBPS}}\right)$	(44 - 24)
Data	Data	34 + L	$T_{DATA} = 20 + 4 \times Ceiling\left(\frac{294+8L}{N_{DBPS}}\right)$	L = 0: (72 - 28) L = 2312: (3152 - 368)

Table C.4: MAC Frame Size

N_{DBPS} value depends on the transmit mode (expressed in number of data bits per OFDM symbol). As shown in table C.2, the extreme values for N_{DBPS} are 24 bits in mode 1 and 216 bits in mode 8. Also, L is the data information length in octets.

C.2.4.5 Data Transport Through Layers

Figure C.7 shows the transport of the data information inside the MAC and PHY layers before its transmission or after its reception.

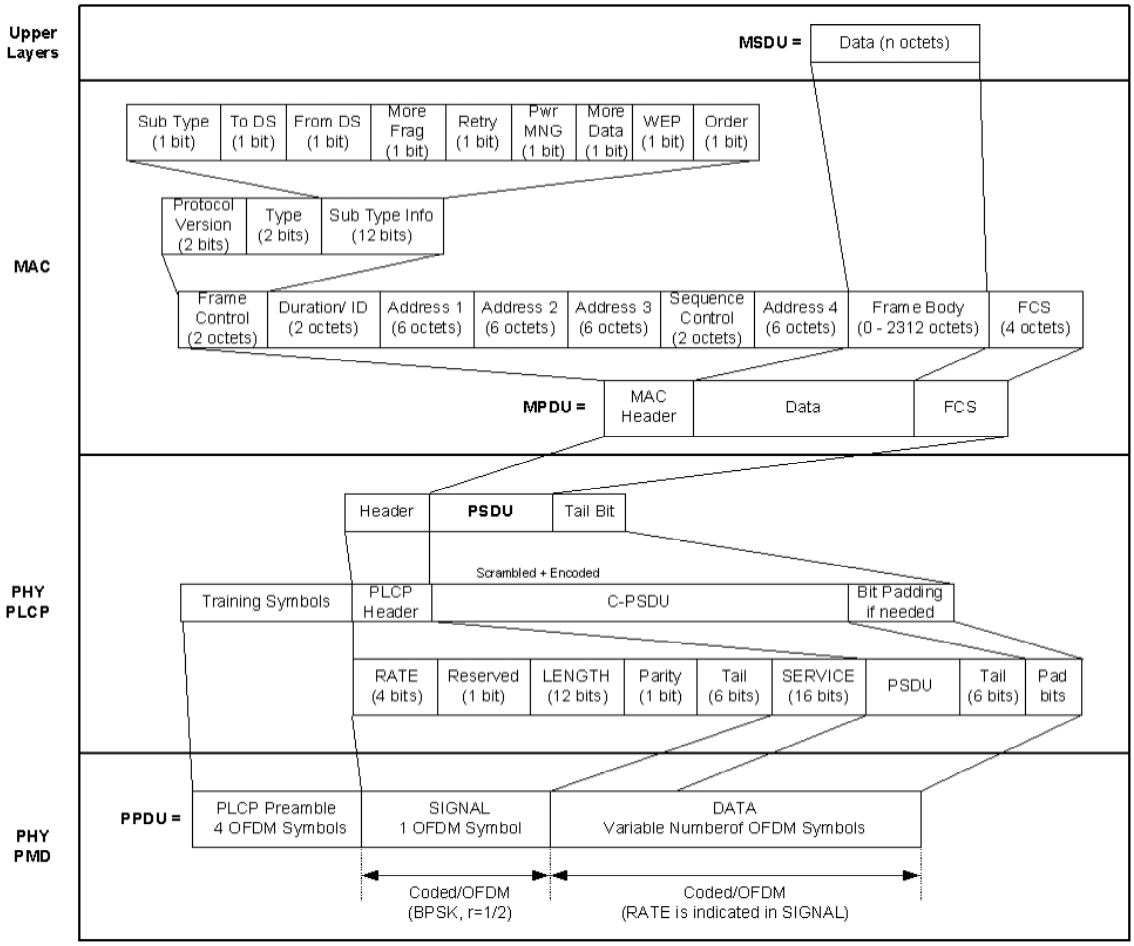


Figure C.7: Data Transport and Formatting Through the Layers [2]

C.2.4.6 CCA Sensitivity (§17.310.5 of [2])

The carrier sensing mechanism is the heart of the CSMA/CA protocol. To decide the medium state, the standard [2] specifies that:
"The start of a valid OFDM transmission at a receive level equal to or greater than the minimum 6 Mbit/s sensitivity (-82 dBm) shall cause CCA to indicate busy with a probability $> 90\%$ within 4 μs. If the preamble portion was missed, the receiver shall hold the carrier sense (CS) signal busy for any signal 20 dB above the minimum 6 Mbit/s sensitivity (-62 dBm)."

In the simulation model the carrier sensing function is implemented as our interpretation of this definition. For example, we decided for control and data frames that:

The start of a valid OFDM transmission at a receive level equal to or greater than the minimum 6 Mbit/s sensitivity (i.e. -82 dBm) shall cause CCA to indicate busy with a **probability 100% for the duration of the detected signal in the air.** If the preamble portion was missed, **the transmitted frame will never be correctly received**, but the receiver shall hold the CS signal busy for any signal 20 dB above the minimum 6 Mbit/s sensitivity (i.e. -62 dBm), idle otherwise. Refer to section 6.6.5 for a more complete discussion about the carrier sensing operation.

C.2.4.7 Channelization (modification of §17.3.8.3.2 and §17.3.8.3.3 of [2])

Channel center frequencies are defined at every integral multiple of 5 MHz above 5 GHz. The relationship between center frequency and channel number is given by the following equation:

$$\text{Channel center frequency} = 5000 + 5 \times N_{Channel} \text{ (MHz), where } N_{Channel} = 0, 1, \cdots, 200.$$

This definition provides a unique numbering system for all channels with 5 MHz spacing from 5 GHz to 6 GHz. The set of valid operating channel numbers for Europe is defined in table C.5.

Regulatory domain	**Band [GHz]**	**Operating channel numbers**	**Channel center frequencies [MHz]**
Europe	5.15 - 5.35 (8 channels)	36	5180
		40	5200
		44	5220
		48	5240
		52	5260
		56	5280
		60	5300
		64	5320
Europe	5.47 - 5.725 (11 channels)	100	5500
		104	5520
		108	5540
		112	5560
		116	5580
		120	5600
		124	5620
		128	5640
		132	5660
		134	5680
		138	5700

Table C.5: Valid Operating Channel Numbers by Regulatory Domain and Band

As a reminder, inside a BSS all stations are contending for the medium using the same channel to Tx and Rx. Other surrounding BSSs may use the same or another channel. Figure C.8 shows the spacing of all the bands described in table C.5.

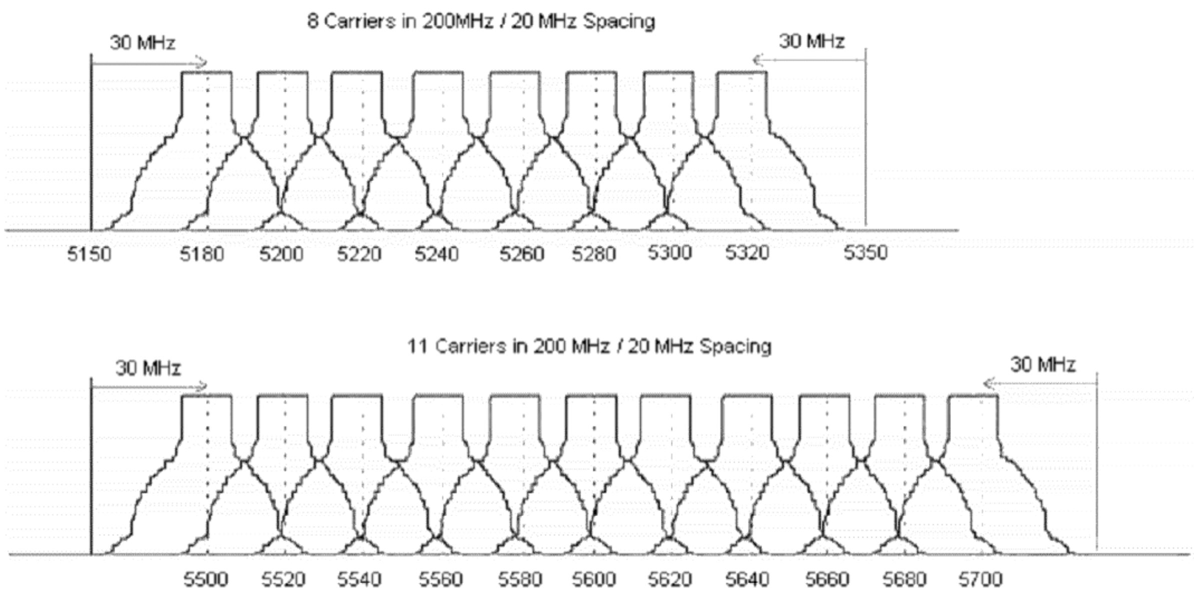


Figure C.8: OFDM PHY Frequency Channels Plan for Europe

C.2.4.8 Transmit Power Levels (modification of §17.3.9.1 of [2])

The maximum allowable output power according to European regulations is shown in table C.6.

Frequency band [GHz]	**Maximum output power with up to 6 dBi antenna gain**	**Comment**
5.15 - 5.35	200 mW = 23 dBm (12.5 mW/MHz)	Indoor use only
5.47 - 5.725	1 W = 30 dBm (62.5 mW/MHz)	Indoor/outdoor use

Table C.6: Transmit Power Levels for Europe

The minimum transmit power in all bands for a station (AP and MT) will be taken as 0 dBm in the simulations.

C.2.4.9 Transmit Spectrum Mask (§17.3.9.2 of [2])

This section presents the relative power emitted by a transmitting station in each band of the radio spectrum. These values ensure that minimum interference level is produced at each transmission. The term dBr represents the value in dB relative to the maximum spectral density of the signal.

Figure C.9 illustrates the values from table C.7.

All the simulations assume an identical transmit mask for all supported data rates, and all STAs (APs and MTs).

Frequency offset [MHz]	Power value [dBr]
≤ 9	0
11	-20
20	-28
≥ 30	-40

Table C.7: Transmit Power Levels for Europe

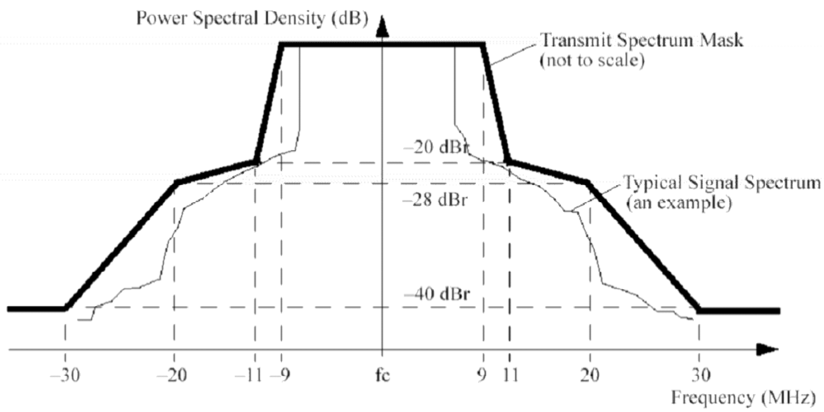


Figure C.9: Transmitter Spectrum Mask [2]

C.2.4.10 Receiver Characteristics (§17.310.1-17.310.4 of [2])

The Packet Error Rate (PER) shall be less than 10% (or 0.1) at a PSDU length of 1000 bytes for rate-dependent input levels shall be the numbers listed in table C.8 or less. The minimum input levels are measured at the antenna connector (NF of 10 dB and 5 dB implementation margins are assumed).

Mode	Minimum sensitivity [dBm]	Adjacent channel rejection [dB]	Alternate Adjacent channel rejection [dB]
1	-82	16	32
2	-81	15	31
3	-79	13	29
4	-77	11	27
5	-74	8	24
6	-70	4	20
7	-66	0	16
8	-65	-1	15

Table C.8: Receiver Performance Requirements

The receiver maximum input level is -30 dBm measured at the antenna for any baseband modulation. The simulations will use the general mask to be used in 5 GHz band for WLAN devices and shown in figure C.10. This mask gives a better rejection than proposed by the standard and is the same for all the modes. Figure C.10 presents this mask, which is a blocking mask.

Frequency offset [MHz]	Power value [dBr]
≤ 9	0
11	-27
≥ 30	-45

Table C.9: General Receiver Rejection Characteristics

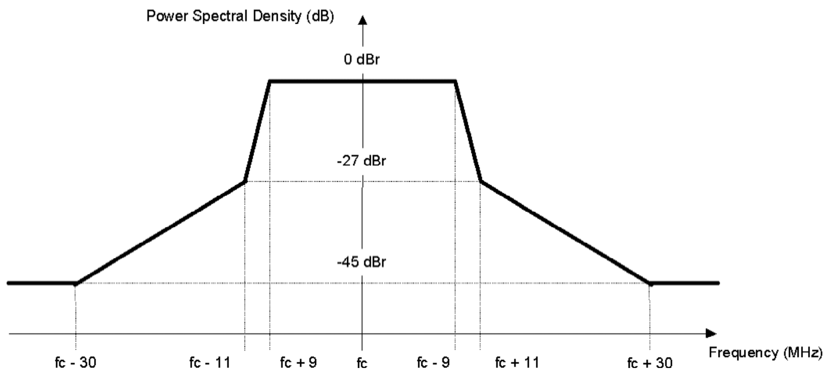


Figure C.10: General Receiver Rejection Mask

dBr: rejection value relative to the one at the current receiver channel (0 dB).

C.2.4.11 Antenna Characteristics

The antenna characteristics for the simulations are presented in table C.10. Only omni-directional antennas are modeled.

	Antenna gain [dBi]	Antenna height [m]
AP	0	2.5
MT	0	1.5

Table C.10: Antenna Characteristics

C.2.5 Specific Implemented Algorithms

C.2.5.1 Cell Channel Selection

An AP choses a single channel out of all the available channels for both UL and DL. As in the 802.11a system no DFS mechanism is implemented, the network planning is done manually. Height channels are available in the band [5.15 - 5.35] GHz (indoor only). When deploying a network composed of several APs, the objective is to maximize inter-channel separation in order to minimize inter-cell interference. This operation is done manually in our simulations: each cell's channel is selected depending on the scenario. We have investigated the impact of several channel offsets.

C.2.5.2 User Cell Attachment

The minimum transmit mode is mode 1. In the simulation model, the MTs and the APs always use transmission of control frames in mode 1. Accordingly, to be attached to a given cell, a user must be able to correctly decode all control frames plus a confidence margin of 3 dB.

C.2.5.3 Link Adaptation

Once attached to a cell, an MT starts with the highest transmit mode (mode 8). Each time the LRC_{Limit} value is reached, the corresponding data is discarded and the transmit mode is decremented by 1, until it reaches mode 1 where is stays. As all control frames are transmitted in mode 1, the LA algorithm is only used for data frames. With the LA algorithm implemented, a STA always transmits the data frame using the highest achievable mode of operation determined by its distance to the AP. Thus, a cell is decomposed in "rings" of maximum modes around the AP as shown in figure C.11.

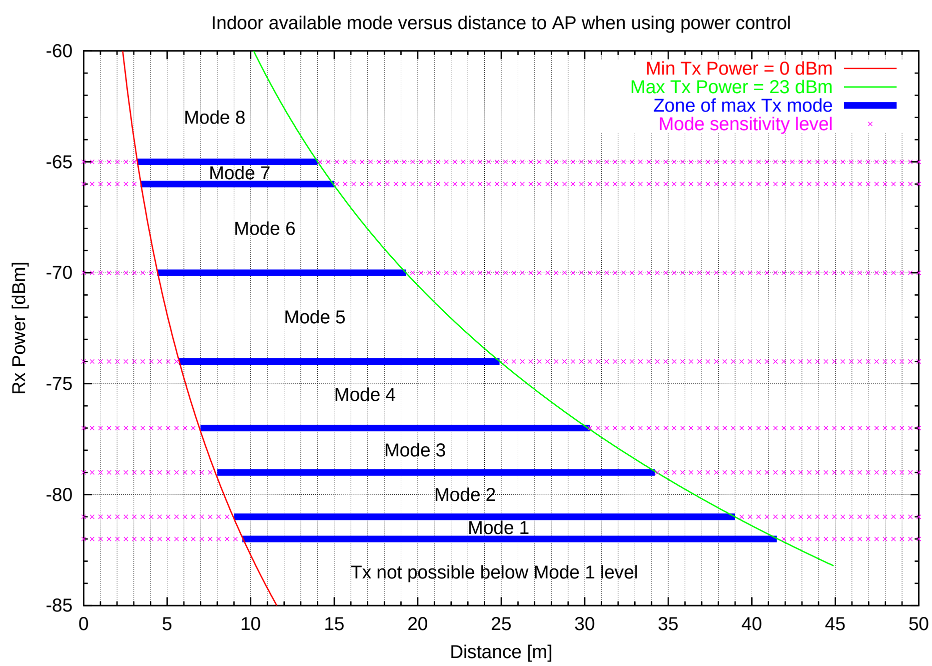


Figure C.11: Maximum Achievable Transmit Mode Versus Distance to AP

Whenever the $\left(\frac{C}{I}\right)$ ratio for a frame is less than the required margin for the frame mode, the frame is unsuccessfully decoded with a probability of 100%. These assumptions explain the sharp transitions from one mode to another observed in figure C.11.

Figure C.11 also gives an indication of the maximum cell coverage in an indoor environment given the simulation assumptions.

C.2.5.4 Handover

All the users are always static and stay attached to their initial cell all along the simulation. Thus, no handover or load balancing algorithm is implemented in the simulation model.

C.3 Conclusion

This chapter has presented our implemented simulation model for the IEEE 802.11a/h technology. To reach such an accurate model, I had to gain a deep understanding of the related standards (IEEE 802.11 [1], IEEE 802.11a [2] and IEEE 802.11h [82]), as well as to propose appropriate algorithms (in accordance with the philosophy of the standard) to model the "implementation dependent" aspects not specified in the standard. Then, using this model, we have investigated several scenarios, whose details and results are detailed in chapter 6.

Appendix D

List of Written Contributions

D.1 International Journals

J.C. Dunat, D. Grandblaise, and C. Bonnet, "Collaborative Allocation of Orthogonal Frequency Division Multiplex Sub-Carriers using the Swarm Intelligence", *to appear in JCM 2006 First Regular Issue*, 2006.

D.2 International Conferences

- J.C. Dunat and D. Grandblaise, "Distributed Time-Based Fairness for Multiple Wireless Swarm Agents in Cognitive Radios", *in WSR 2006,* 4^{th} *Karlsruhe Workshop on Software Radios*, March 22-23, 2006, Karlsruhe, Germany.

- J.C. Dunat and C. Bonnet, "Uplink Distributed Resource Allocation Algorithm for Maximized WLAN Performance", *in EUROCON 2005, IEEE Region 8 Conference*, November 21-24, 2005, Belgrade, Serbia & Montenegro.

- J.C. Dunat, D. Grandblaise, and C. Bonnet, "Efficient OFDMA distributed optimization algorithm exploiting multi-user diversity", *in DySPAN 2005, 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, November 8-11, 2005, Baltimore, US.

- J.C. Dunat, L. Elicegui, and C. Bonnet, "Impact of inter-cell interference in a IEEE 802.11a network with overlapping cells", *in PIMRC 2004, 15th IEEE international symposium on personal, indoor and mobile radio communications*, 5-8, September 2004 -Barcelona, Spain.

- J.C. Dunat, L. Elicegui, and C. Bonnet, "Impact of the users' distribution on the IEEE 802.11a MAC protocol fairness and performance", *in PIMRC 2004, 15th IEEE international symposium on personal, indoor and mobile radio communications*, 5-8, September 2004 -Barcelona, Spain

- J.C. Dunat and L. Elicegui, "Innovative flexible architecture for a multi-Radio Access Technologies (RAT) system level simulation tool", *Motorola S3 Symposium*, 2004.

D.3 Technical Contributions to European Projects

- Jijun Luo and al., "Analysis of Combined Strategies including Concepts, Algorithms and Reconfigurable Architecture Aspects", D5.4 deliverable, *E2R European Project, IST-2003-507995/E2R/WP5/D5.4*, December 2005.
- Clemens Kloeck and al., "Algorithms and Performance (including FSM & RRM/Network Planning)", D5.3 deliverable, *E2R European Project, IST-2003-507995/E2R/WP5/D5.3*, June 2005.
- Technical contribution to the MESA Project.

Appendix E

Résumé en français

Résumé

Depuis quelques années on assiste à un accroissement de la demande pour l'utilisation de services sans fil. Un exemple est le succès des réseaux locaux sans fil tels que IEEE 802.11. Ces systèmes utilisent du spectre radio pour fonctionner. Or, de nombreux pays souffrent déjà d'une pénurie de spectre disponible.

Heureusement, de nombreuses mesures ont montré qu'une grande partie du spectre officiellement alloué à chaque système n'est en réalité pas utilisé partout et tout le temps. Ce constat de "trous" dans le spectre radio ouvre la voie au développement de nouvelles méthodes d'allocation du spectre plus flexibles et opportunistes.

Notre thèse s'est concentrée sur les aspects techniques de tels algorithmes dynamiques d'allocation du spectre. Nous nous sommes placé dans le contexte des radios cognitives (radios capables de détecter et de s'adapter à leur environnement).

Notre principale contribution a été le développement d'un algorithme d'allocation opportuniste de sous-porteuses OFDM en voie montante permettant à plusieurs mobiles de se partager les ressources radio disponibles (plusieurs sous-porteuses OFDM) de façon à maximiser la débit d'une cellule. Cet algorithme réalise une allocation de spectre de manière distribuée et s'inspire des résultats sur les comportements des insectes sociaux pour le partage des tâches. Nous pensons que de tels algorithmes permettant une auto-organisation d'un groupe d'entités sont amenés à se développer pour les prochaines générations de systèmes de communication sans fil.

E.1 Introduction

E.1.1 Motivation et objectifs

Les utilisateurs des technologies sans fil sont de plus en plus exigeants sur la qualité et la disponibilité des services pour lesquels ils paient, certains disposants de terminaux multimodes. Cependant, une technologie ne peut offrir tous les services avec le meilleur débit. Le paysage des télécoms est donc nécessairement voué à être hétérogène, avec des passerelles entre les différentes technologies sans fil sont plus que jamais nécessaires pour assurer une continuité de service à l'utilisateur.

Dans ce travail de thèse nous avons cherché à proposer des algorithmes innovants d'allocation de la ressource spectrale de façon plus dynamique en fonction du temps et de l'espace pour les utilisateurs. Nous avons principalement étudié la transmission de paquets radio. Nous avons obtenu un contrôle distribué de l'allocation de spectre (contexte de l'OFDMA) avec une approche montante (des terminaux des utilisateurs vers les équipements du réseau) et utilisant une coordination intelligente, l'idée étant de propager l'optimisation du niveau de plus bas jusqu'à l'ensemble du réseau. Les conclusions générales contenues dans la section E.9 décrivent les extensions possibles de ce travail à d'autres scenarios, ainsi que les nécessaires améliorations.

E.1.2 Réalisations et contributions

Voici la liste des principales réalisations de cette thèse :

- Nous avons tout d'abord étudié (en théorie et par simulation) les performances de la technologie sans fil IEEE 802.11 (tout particulièrement le IEEE 802.11a). Les problèmes connus de type noeud caché, iniquité spatiale ont été montrés.

- Nous avons ensuite cherché à résoudre ces problèmes tout en proposant un algorithme d'allocation flexible de spectre. La méta-heuristique utilisée pour résoudre le problème d'optimisation difficile de l'allocation de spectre a été celle de "l'intelligence en essaim". Cette méthode est inspirée des insectes sociaux capables de trouver, de façon complètement décentralisée, un optimum global à partir de multiples interactions locales entre individus, et en utilisant des règles simples pour communiquer et se guider mutuellement vers la meilleure solution du point de vue du système. Cette méta-heuristique a les propriétés intéressantes suivantes: scalabilité, robustesse, adaptabilité.

- Afin de pouvoir évaluer par simulations les résultats de nos recherches, un outil de simulation a été créé. Il est suffisamment générique pour permettre d'étudier les performances d'un système isolé ou de plusieurs systèmes interagissant.

- Enfin, ce travail de thèse a donné lieu à la publication de 6 articles dans des conférences, 1 article de journal, ainsi que des contributions dans des projets européens.

E.1.3 Organisation du résumé

Cette thèse se concentre sur le développement d'un algorithme d'allocation opportuniste du spectre radio pour des radios cognitives. Dans cette section nous avons introduit le contexte et les objectifs de notre travail. La section E.2 présente le problème de l'DSA ainsi que celui de l'actuelle pénurie dans l'allocation de spectre. Ensuite, la section E.3 revient sur l'évolution des techniques pour introduire de la flexibilité dans l'allocation de spectre, depuis les premières tentatives jusqu'aux nouvelles tendances. Dans la section E.4, nous décrivons nos premières tentatives d'algorithmes dynamiques d'allocation de spectre. Ensuite nous avons décidé d'utiliser une approche distribuée et montante basée sur "l'intelligence en essaim". Plus de détails sont consignés dans la section E.6 qui est consacrée à la présentation de notre méthode d'allocation de spectre opportuniste distribuée. Des résultats de performances et de limitations des systèmes locaux sans fil (notamment du IEEE 802.11a et IEEE 802.11h) basés sur du OFDM sont présentés dans la section E.7. Dans la section E.8 nous améliorons les performances des systèmes précédents en utilisant la méthode présentée dans la section E.6. Finalement, la section E.9 conclue le travail effectué et propose des pistes de continuation.

E.2 Allocation dynamique de spectre

E.2.1 Introduction

Cette section présente un statut sur l'état actuel du spectre radio et explique pourquoi il y a encore de la place pour d'autres systèmes, mais dans un autre contexte que l'actuel. En effet, le problème actuel de la rareté du spectre se situe au niveau de la réception [3] mais pas de la transmission. En effet, les interférences sont un problème au niveau des récepteurs. Ceux-ci sont dit peu intelligents [4]. Si l'on considère les 2 axes de l'espace et du temps, on constate que le spectre est actuellement sous utilisé. Cependant, les règles actuelles de régulation dans l'allocation du spectre ne permettent pas de bénéficier de ces trous qui existent dans le spectre.

E.2.2 "La tragédie des communs"

Dans "la tragédie des communs" [5], Hardin aborde le problème du partage d'une ressource matérielle par plusieurs utilisateurs de façon excessive. Il montre qu'en l'absence de règles, le résultat est le chaos. Il propose plusieurs règles qui permettent de réguler l'accès à la ressource, choix basé sur la : loterie, loi du premier arrivé premier servi, loi du mérite, etc. Certaines de ces règles sont actuellement appliquées par les opérateurs.

Cependant, la ressource spectrale n'est pas une ressource matérielle mais un ensemble de fréquences électromagnétiques utilisées pour transporter de l'information. Par conséquent, les règles de partage du spectre doivent être différentes de celles employées pour des ressources matérielles. De plus, actuellement, les règles d'allocation du spectre ne sont pas adaptées à la réalité du spectre ni à la réalité des fluctuations des besoins des utilisateurs.

E.2.3 Status actuel de l'allocation du spectre

La figure E.1 illustre le problème actuel de la rareté spectrale tel qu'il existe dans de nombreux pays, en montrant l'état de la table d'allocation aux USA. On constate une saturation de la table d'allocation, ce qui complique l'attribution de fréquences pour de nouveaux systèmes radios.

Heureusement, comme le montrent les mesures de la Shared Spectrum Company, une grande partie du spectre n'est pas entièrement utilisé partout et tout le temps, levant ainsi le voile sur une nouvelle source de gain, avec la même quantité de spectre initiale.

Ces mesures montrent que l'utilisation du spectre est concentrée dans certaines bandes de spectre (jusqu'à 13.1% à New York City) alors que d'autres sont presque inutilisées (1% pour la radio astronomie). Cela montre qu'il existe un contexte favorable pour s'affranchir du problème de la rareté spectrale grâce à : des solutions techniques (algorithmes plus intelligents pour trouver et utiliser ces trous dans le spectre [8]), avec un modèle économique approprié à l'utilisation future du spectre ([9]), et dans un contexte régulatoire plus flexible et qui valorise l'utilisation du spectre plutôt que sa possession ([10]).

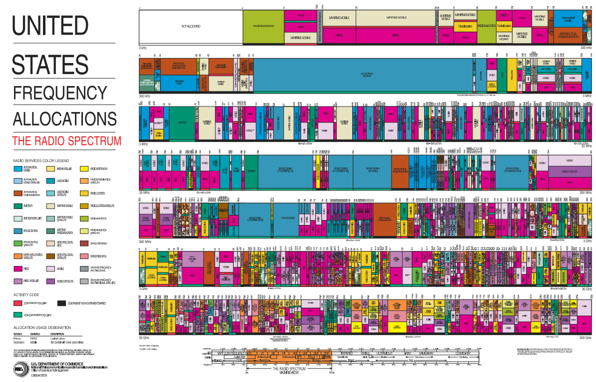


Figure E.1: Table d'allocation des fréquences radio aux USA

E.2.4 Prévisions de croissance de l'utilisation du spectre

Au fur et à mesure des années, les trois indicateurs suivants se sont maintenus pour le monde du sans fil : augmentation du nombre d'abonnés souscrivant à des services sans fil, augmentation du volume des données échangées par utilisateur, et les utilisateurs demandent sans cesse de nouveaux services disponibles n'importe où et tout le temps avec une continuité de service. Des analyses ont montré que l'industrie du sans fil consolidera dans les années à venir son importance. De plus, l'Asie et les pays émergents vont renforcer leur poids (notamment en nombre d'abonnés) par rapport à l'Europe et aux Etats-Unis dans les solutions sans fil. Par conséquent, toujours plus de spectre sera nécessaire dans les années à venir. D'où pourra venir ce supplément de spectre, au vue de la pénurie actuelle ?

E.2.5 Conclusion

Malgré les règles rigides actuelles d'allocation du spectre qui ont artificiellement créé une rareté du spectre, il existe des méthodes permettant d'accroître encore l'utilisation de bandes de spectre déjà très saturées. En effet, certains services ont une utilisation décorrélée dans l'espace et le temps et créent des trous dans le spectre qui peuvent être détectés par des algorithmes intelligents. Par conséquent, les algorithmes d'allocation dynamiques du spectre permettent de réagir face à une saturation de certaines zones du spectre radio, face à d'autres qui sont sous-utilisées.

E.3 Vers plus de flexibilité dans l'allocation de spectre

E.3.1 Introduction

Cette section décrit l'évolution chronologique du contexte de l'allocation de spectre au niveau des régulateurs et des opérateurs, depuis les premières techniques jusqu'aux dernières évolutions (DSA et les radios cognitives). Il est important de noter que les règles employées pour l'allocation (temporelle et spatiale) du spectre d'un système ou d'une bande de fréquence influencent fondamentalement sur l'efficacité et le succès de tels systèmes.

Dans les premiers temps des communications sans fil, les régulateurs ont segmenté la ressource radio en segment réservé comme des propriétés privées pour les systèmes choisis. Ceci afin d'empêcher les problèmes d'interférence entre systèmes. Cette méthode de découpage du spectre sous forme de droits d'utilisation exclusifs a artificiellement créé la rareté actuelle de la ressource spectrale [12] qui atteint son paroxysme dans les zones à fortes concentrations de populations (zones urbaines). Il existe une inégalité dans l'attribution du spectre car une grande partie du spectre est alloué à des agences gouvernementales qui n'utilisent que sporadiquement le spectre en comparaison des systèmes commerciaux. Les régulateurs peuvent décider quelles technologies opèreront dans du spectre avec licence (le détenteur de la licence a un accès exclusif à cette partie du spectre, sans obligation d'utiliser ce spectre) ou bien dans du spectre sans licence (il n'y a pas d'accès exclusif pour cette partie du spectre mais plutôt un accès partagé).

E.3.2 Premières tentatives d'allocation flexible de spectre dans les systèmes cellulaires

Pour pallier à leur manque de spectre et accroître la surface de couverture de leurs réseaux les opérateurs réutilisent leurs canaux radios. Pour ce faire, ils utilisent différentes méthodes [13] :

- FCA: méthodes simples qui n'adaptent pas la répartition des canaux en fonction de la distribution des utilisateurs.
- DCA: pas de relation figée entre les canaux et les cellules. Cela utilise une méthode de pool de canaux (par opérateur) où les cellules puisent quand elles ont besoin de spectre (en s'assurant que les contraintes d'interférence avec les cellules voisines sont respectées). Ces méthodes modélisent bien ce qui se passe dans des bandes sans licence.
- HCA: combinent les avantages des stratégies FCA et DCA. Les canaux sont divisés en deux groupes : fixes (canaux appartenant à une cellule comme en FCA) et dynamiques (ces canaux peuvent être utilisés par différentes cellules comme en DCA).

E.3.3 Brève description de l'allocation dynamique de spectre(DSA)

L'idée de l'allocation dynamique de spectre est d'utiliser le spectre temporairement non utilisé (libre) pour un autre service que celui pour lequel il est initialement prévu. Le DSA a été étudié dans les projets européens DRiVE [14] et OverDRiVe [15]. Il existe trois types de DSA qui peuvent fournir du gain en profitant de l'orthogonalité temporelle (DSA temporel), spatiale (DSA spatial) ou des 2 (DSA temporel et spatial), entre services qui peuvent donc utiliser le spectre laisser libre par les autres services. L'importance du gain fournit est hautement conditionnée par l'orthogonalité entre les services considérés. L'approche du DSA est intéressante car elle ouvre la voie à une gestion plus dynamique de la ressource radio nécessitant un échange entre systèmes.

E.3.4 Le succès du spectre sans licence

Les bandes de spectre sans licences sont identifiées par les régulations de l'ITU sont : la bande ISM : comprise entre 2400 et 2483.5 MHz, 4.90 GHz et 5.90 GHz, ainsi qu'entre 868 et 928 MHz. De nos jours, un nombre croissants de technologies sans fil opèrent dans ces bandes (WiFi© (IEEE 802.11b/g/a/...), Bluetooth^{TM} (IEEE 802.15.1)) sans avoir besoin de licence d'émission (pas de coût de licence) et sont faciles à déployer.

Même si la saturation de ces bandes crée parfois de l'interférence inter-système, tous les canaux radio sont mis en commun et peuvent être accédés par n'importe quelle entité. Par conséquent, ce contexte prouve qu'une utilisation du spectre basée sur les besoins réels et sans notion de propriété peut fonctionner.

E.3.5 Open Spectrum: une nouvelle vision de la ressource radio

L'initiative "Open Spectrum" est menée par David Reed [16] et propose une vision futuriste de l'utilisation du spectre et de l'organisation des réseaux. Le cadre de l'"Open Spectrum" est de parvenir à un réel partage dynamique de la ressource spectrale entre les utilisateurs, et entre les systèmes, au fil de leurs besoins, et en s'affranchissant des règles de la régulation, moyennant un ensemble minimum de règles d'utilisation.

Les supporters de cette vision de l'"Open Spectrum" véhiculent les idées suivantes :

- Le problème de l'interférence n'est pas une loi intrinsèque de la nature mais plutôt un problème au niveau du récepteur (stupide) [4] qui est incapable de séparer les signaux provenant de plusieurs sources.
- L'accès au spectre ne devrait pas être réglementé comme le sont les ressources matérielles, sur la base par exemple d'un droit d'accès exclusif à la propriété.
- L'"Open Spectrum" ne va pas créer plus de spectre, mais par des communications coopératives, ce contexte devrait accroître le gain que l'on peut obtenir avec le spectre actuel. Ainsi, la capacité du système va s'accroître à mesure que le nombre d'utilisateurs augmentent [4].

En temps qu'exemple, notons qu'un contrôle centralisé et une régulation rigide aurait empêché l'Internet de devenir ce qu'il est actuellement, alors qu'il faut mettre l'intelligence a la périphérie [21].

E.3.6 De nouvelles opportunités avec la radio cognitive

La vision de la radio cognitive fut introduite par Mitola [22]. C'est une vision assez vaste qui comprend de nombreux concepts [22] [23] [24], dont celui d'une nouvelle façon de gérer le spectre. La vision de la radio cognitive s'appuie sur une architecture de radio logicielle SDR.

E.3.6.1 SDR

Comme définit en essence par la FCC [25], une radio logicielle SDR est un équipement qui peut modifier certaines de ses caractéristiques de façon logicielle. En d'autres termes, c'est un équipement qui peut synthétiser de nouvelles fonctionnalités au cours du temps et devenir un autre équipement, en étant ainsi potentiellement plusieurs équipements à la fois. La reconfiguration serait commandée par les algorithmes d'allocation flexible de spectre. Ce concept de radio logicielle, envisagé dans les années 90, est en train de devenir réalité, même si de nombreux progrès technologiques existent encore tels que celui de trouver des convertisseurs adéquats (analogique-numérique) et des antennes avec des bonnes propriétés. Le gain d'une reconfiguration d'un équipement sera important si les algorithmes d'allocation flexible de spectre sont efficaces. De plus, ce gain doit compenser le coût de la reconfiguration ainsi que le surcoût de ces terminaux par rapport à des terminaux classiques.

E.3.6.2 CR

La radio cognitive CR introduit un langage appelé RKRL [22] et permet une communication intelligente entre terminaux au sein d'un réseau sans fil, de l'apprentissage, une connaissance et une interprétation de l'environnement permettant aux terminaux d'optimiser leur profil.

Ainsi, de nouvelles approches de gestion du spectre radio voient le jour. On peut citer comme exemple la technique du pool de spectre [29], [30] qui consiste à permettre un accès à une ressource radio partagée en fonction de la disponibilité du spectre et des besoins réels, et non plus basé sur une notion de priorité.

E.3.6.3 La technologie de la XG DARPA

Les initiatives de la XG Darpa [31] [32] traitent des radios cognitives. Il s'agit d'un nouveau contexte qui accroît la part de contrôle des terminaux utilisateurs sur le choix de leur accès au spectre. Ainsi, l'approche XG, appelée accès opportuniste au spectre radio, repose sur deux parties :

- Agilité spectrale grâce à des terminaux qui s'adaptent à leur environnement.
- Dynamicité de la politique de régulation grâce à des règles de régulation traduites dans un langage compréhensible par les terminaux (politique d'allocation logicielle qui évolue au cours du temps et de l'espace).

Bien que simple à comprendre, la réalisation matérielle de cet accès opportuniste au spectre est très difficile à mettre en IJuvre et prendra quelques années avant d'être démontrée. En effet, de nombreux problèmes doivent être résolus : détection des trous dans le spectre, communication coordonnée entre terminaux d'un système, etc.

E.3.7 Commerce de spectre

Le commerce de spectre (ou "spectrum trading" en anglais) est un cadre qui offre des techniques d'échanges de spectre entre plusieurs systèmes en fonction de leurs besoins respectifs. Ces mécanismes sont basés sur le mécanisme de régulation par le jeu de l'offre et de la demande. Le but étant d'amorcer une transition vers plus de flexibilité dans l'allocation du spectre, tout en assurant une certaine stabilité du système global. Les économistes sont très intéressés pour appliquer ces modèles économiques à n'importe quel échange de ressource, y compris le spectre. Dans ce contexte, le spectre est donc considéré comme une ressource matérielle que l'on peut échanger, vendre, louer, etc. Cette approche de commerce de spectre est essentiellement plébiscitée par l'OFCOM (UK), les USA et la Nouvelle Zélande.

E.3.8 Conclusion

De nombreuses nouvelles initiatives voient le jour pour améliorer l'efficacité des méthodes de partage du spectre. Cet élan est motivé par l'état de congestion actuel des tables d'allocation des régulateurs. En effet, tous les acteurs du monde du sans fil sont aujourd'hui conscients de l'enjeu majeur que représente le spectre. C'est pourquoi régulateurs, économistes et ingénieurs cherchent ensemble à trouver des solutions aux problèmes actuels, ces solutions qui feront le succès des systèmes de radio communications de demain.

E.4 Etudes comparatives entre plusieurs méthodes d'accès radio

E.4.1 Liste des notations pour cette section

Le tableau E.1 présente les notations utilisées dans cette section.

Notation	Description
β	Paramètre utilisé pour pondérer le taux de service (ex: $\beta\mu$)
μ	Taux de service moyen (loi exponentielle)
μ_1	Taux de service moyen de la phase de réservation (cf méthode b) ci-après)
μ_2	Taux de service moyen de la phase de transmission des données (cf méthode b) ci-après)
μ_i	Taux de service moyen pour la station i
σ	Durée du décrément du compteur de backoff (cf [2])
a, b	Paramètres entiers pour décomposer M en plusieurs groupes
e_i	Taux de visite de la station i
$p_i(k,m)$	Probabilités marginales à la station i pour un réseau de m clients
C_i	i^{ime} serveur pour une station composée de plusieurs serveurs
CW_{min}	Taille de la fenêtre de contention minimum pour le standard IEEE 802.11a
L	Charge utile des paquets transmis (en bytes)
M	Nombre total de clients (utilisateurs géographiquement distribués) dans le réseau
N	Nombre total de serveurs (canaux radios) dans le réseau
$Q_i(M)$	Nombre moyen de clients à la station i
$Q(M)$	Nombre moyen de clients dans le réseau
$R_i(M)$	Temps moyen de traversée de la station i
$R(M)$	Temps moyen par client entre, après son service et la fin de son prochain service
T_{ACK}	Durée d'un message de contrôle ACK (cf [2])
T_{CTS}	Durée d'un message de contrôle CTS (cf [2])
T_{DATA}	Durée d'un message de DATA (cf [2])
T_{DIFS}	Durée d'une période d'attente DIFS (cf [2])
T_{RTS}	Durée d'un message de contrôle RTS (cf [2])
T_{SIFS}	Durée d'une période d'attente SIFS (cf [2])
$X_i(M)$	Débit de la station i
$X(M)$	Débit de la station de référence

Table E.1: Liste de Notations pour cette Section

E.4.2 Introduction

Dans cette section plusieurs méthodes d'accès au canal radio et de politique de partage du spectre vont être comparées. Notre but étant de trouver la méthode fournissant le débit total le plus élevé suite au partage de spectre entre plusieurs utilisateurs étant donné plusieurs canaux radio disponibles.

E.4.3 Analyse des performances des méthodes de partage de spectre

E.4.3.1 Queuing Model

Dans cette section avons modélisé les techniques de partage du spectre en utilisant la théorie des files d'attentes. Nous supposons un groupe de N canaux radio disponibles, et M utilisateurs géographiquement distribués. De plus, nous supposons que les canaux ne s'interfèrent pas entre eux, qu'il n'y a pas de collisions : tout les équipements monde dans le réseau ont une parfaite connaissance de l'état du système. Les résultats de cette étude sont destinés à une modélisation de la voie montante.

Chacune des méthodes présentées ci-après répond de façon différente à cette question : "à un instant donné, quel utilisateur transmet et en utilisant quel canal radio ?". Notre modèle utilise une politique d'attente des clients (utilisateurs) pour accéder aux serveurs (canaux) de queue de type FIFO. La figure E.2 illustre les notations utilisées par l'étude en cours.

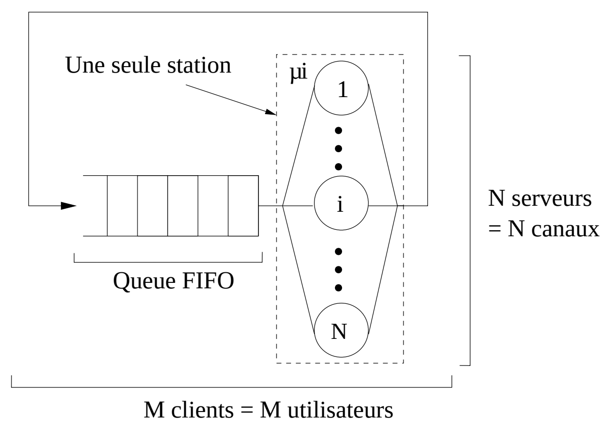


Figure E.2: Notations du modèle de file d'attente

Plusieurs méthodes sont comparées dans les paragraphes suivants :

- a) Fixe : les utilisateurs sont organisés en groupes de plus d'un utilisateur, et il y a au maximum un canal par groupe. Dans chaque groupe, tous les utilisateurs utilisent le même canal pour la transmission et la réception.
- b) Canal commun et canaux de données : 1 canal commun et $N-1$ canaux de données. Avant chaque transmission un utilisateur doit gagner la contention pour pouvoir ensuite chercher un canal de données libre. Il transmettra sur ce canal libre.
- c) Canal unique : un seul canal est disponible pour les M utilisateurs, mais il a un taux de service de $\beta\mu$. Ce canal simple correspond à tous les N canaux mis ensemble.
- d) Séquentiel : un utilisateur choisit aléatoirement un canal parmi N et après la fin de chaque transmission il choisit d'utiliser séquentiellement le canal avec l'index suivant.
- e) Pool aléatoire : le système est consisté de N canaux organisés en un pool commun. Tous les canaux sont accessibles par tous les utilisateurs dès qu'un canal devient libre.

- f) Routage probabiliste : le système est consisté de N canaux et un utilisateur peut accéder à n'importe lequel d'entre eux. Un utilisateur choisit aléatoirement un canal avant une transmission et ensuite il attend que ce canal devienne libre afin de pouvoir l'utiliser pour transmettre.

- g) Opportuniste : le système, consisté de N canaux, choisit toujours d'offrir l'opportunité de transmettre à l'utilisateur qui a le meilleur SINR sur un canal donné.

Les valeurs de SINR varient au cours du temps, ce qui est modélisé par la variation de la durée du paquet (et donc du service). Cependant, les méthodes a) à f) ne prennent pas en compte les valeurs de SINR pour adapter leur stratégie d'allocation. Au contraire, la méthode g) schédule toujours l'utilisateur avec les meilleures conditions radios, ce qui est supposé donner les débits agrégés les plus élevés.

Dans notre étude nous avons modélisé des réseaux de files d'attente fermes, correspondant à un trafic saturé avec nombre constant d'utilisateurs. Pour trouver les performances de chaque méthode de façon analytique nous avons utilisé l'algorithme itératif de MVA [34]. Nous avons considéré les deux cas suivants : réseau composé d'une ou plusieurs stations avec simple serveur par station, et réseau composé d'une ou plusieurs stations avec multiples serveurs par station.

E.4.4 Comparaison des méthodes

Dans les figures suivantes, on a varié N de 1 à 20 par pas de 1 (sauf pour la méthode b) qui commence à $N = 2$), $M = 10$, et la durée simulée est de 10 secondes. Les différentes méthodes sont comparées en terme de débit agrégé total pour le système. Les courbes ont été obtenues en implémentant les résultats analytiques (après les avoir préalablement validés par simulation), sauf pour la méthode g) qui a été obtenue par simulation.

La figure E.3 représente débit agrégé total pour chaque méthode (de a) à f)).

Les principales conclusions sont les suivantes :

1. Toutes les méthodes ne donnent pas lieu au même débits.
2. Pour M donné, un accroissement du nombre de canaux résulte pour toutes les méthodes en un accroissement du débit, mais avec une pente différente suivant les méthodes.
3. Les méthodes fournissant les débits les plus élevés sont la méthode fixe et la méthode du pool de canaux.
4. La méthode avec un canal commun et des canaux dédiés pour les données est moins efficace que la méthode du pool de canaux car elle crée un goulot d'étranglement au niveau du canal commun.
5. Pour $N \geq 4$, la méthode du canal unique fournit le débit de plus faible car elle ne parallélise pas le temps d'attente sur plusieurs canaux.

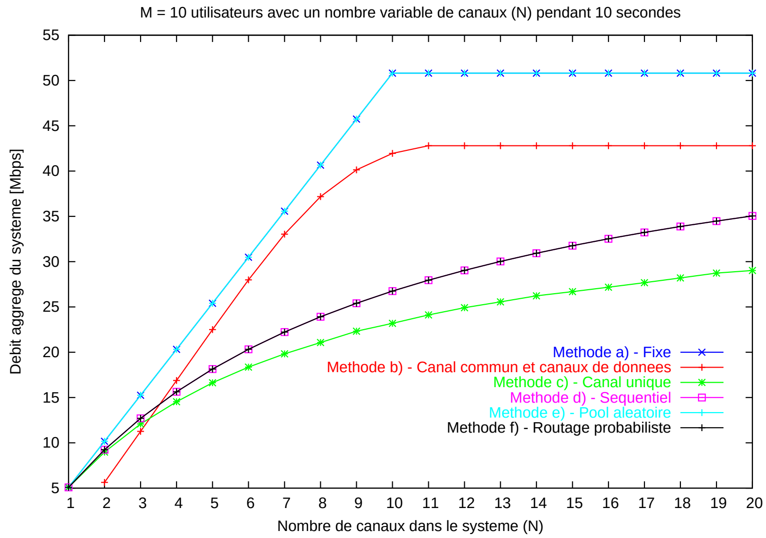


Figure E.3: Comparaison des méthodes a) à f) en terme de débit aggrégé du système [Mbps]

La figure E.4 représente une comparaison entre la méthode e) (la meilleure méthode entre toutes celles de a) à f)) et la méthode g), en terme de débit agrégé du système et de temps moyen d'attente.

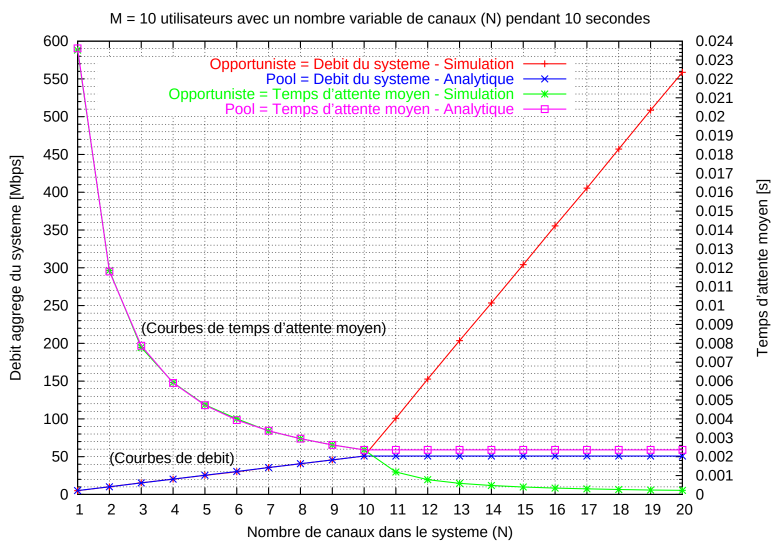


Figure E.4: Comparaison des méthodes e) et g) en terme de débit aggrégé du système [Mbps] et de temps moyen d'attente [s]

Comme observé sur la figure E.4, quand les utilisateurs en tête de queue dans la méthode g) ont le choix entre moins qu'un canal (cas où $N <= M$) le débit est le même que la méthode en pool. Mais dès que $N > M$, le débit de la méthode g) augmente linéairement suivant N, et reste constant pour la méthode e). En utilisant nos hypothèses de cette étude, la méthode g) fournit G ($G = max\{0; N - M\} + 1$) fois plus de débit que la méthode e) ($N > M$). En fait, la pente devient M fois plus importante quand $N > M$ que lorsque $N \leq M$. La raison vient du choix opportuniste que chaque utilisateur peut faire avant d'utiliser un nouveau canal : il choisit d'utiliser le canal libre pour lequel il a les meilleures conditions de canal.

E.5 Conclusion

En conclusion de la comparaison de toutes les méthodes étudiées (basées sur une politique d'attente FIFO), la méthode bénéficiant du débit le plus élevé pour le système est la méthode g). En effet, la diversité multi-canal (cette méthode dispose d'un choix opportuniste du canal radio, contrairement à un choix systématique ou aléatoire qu'ont d'autres méthodes) introduit du gain dans le débit du système. Par conséquent, un contrôle local (effectué au niveau du terminal de l'utilisateur) de l'accès au canal en fonction des conditions radios accroît le débit pour l'ensemble du système.

Une méthode encore plus efficace en débit (mais non présentée dans cette section car elle n'utilise pas une politique d'attente FIFO), est présentée dans la section E.6. Cette méthode introduit de la diversité multi-utilisateurs et multi-canaux.

Ces résultats sont basés sur des hypothèses simplificatrices par rapport à la réalité. Or dans un système réel, des collisions dans l'accès au canal sont inévitables. Par conséquent, la communication entre les utilisateurs est nécessaire pour qu'ils se coordonnent et qu'ils négocient efficacement leur accès au spectre.

E.6 Contexte de résolution du problème d'optimisation de l'allocation de spectre

E.6.1 Liste des notations pour cette section

Le tableau E.2 présente les notations utilisées dans cette section.

Notation	Description
a_{ij}	Valeur binaire (0 or 1) pour indiquer si la tâche i est utilisée par la ressource j
f_{ij}	Valeur de force obtenue suite à l'utilisation de la ressource j par la tâche i
i	Index utilisé pour faire référence aux tâches, notamment la tâche i
j	Index utilisé pour faire référence aux ressources, notamment la ressource j
n_i	Nombre de ressources allouées à la tâche i
n_{max}	Nombre maximum autorisé de ressources simultanément utilisées par tâche
n_{min}	Nombre minimum autorisé de ressources simultanément utilisées par tâche
M	Nombre total de tâches
N	Nombre total de ressources disponibles à être allouées pour des tâches
R_{NA}	Nombre de ressources libres en raison des contraintes du problème
R_A	Nombre de ressources qui vont être allouées en raison des contraintes du problème
$S(N,i)$	Nombre de partitions de l'ensemble N en i groupes
$S(N;k_1,k_2,...,k_N)$	Nombre de partitions de l'ensemble N en k_i blocs de taille i
T_A	Nombre de groupes de ressources qui a pu être alloué

Table E.2: Liste des notations pour cette section

E.6.1.1 Formulation mathématique du problème d'optimisation

Notre problème d'optimisation est un problème d'allocation de ressource. Etant donné N ressources disponibles et à allouer parmi M tâches, chaque tâche i ayant une valeur de force f_{ij} sur la ressource j, nous étudions la meilleure allocation des tâches sur les ressources de façon à maximiser la valeur somme totale pour le système en entier.

En utilisant les notations du tableau E.2, l'objectif du problème d'optimisation peut être formulé en termes mathématiques :

$$
\begin{aligned}
&\text{Maximize } \sum_i \sum_j a_{ij} f_{ij} \\
&\text{s.t.:} \begin{cases}
(1) & \{n_{min}, n_{max}\} \in [1; N] \\
(2) & \forall i,\ n_i = 0 \Leftrightarrow \sum_j a_{ij} = 0 \\
(3) & \forall i,\ n_i \neq 0 \Leftrightarrow \\
 & n_{min} \leq \sum_j a_{ij} \leq n_{max} \\
(4) & \forall j,\ \sum_i a_{ij} \leq 1
\end{cases}
\end{aligned}
\tag{E.1}
$$

L'objectif est de trouver le vecteur d'allocation final pour le système tel qu'il maximise la somme des forces, en supposant que toutes les contraintes du problème sont respectées.

E.6.1.2 Complexité du problème d'optimisation

Les paramètres suivants sont variables pour notre problème : conditions radio, ressource spectrale disponible, nombre de noeuds en compétition, etc. Le problème d'optimisation actuel avec ses contraintes est NP-complet et requiert l'utilisation d'une heuristique (ou d'une méta-heuristique) afin de :

1. Guider la recherche dans l'immense espace des solutions possibles.
2. Chercher une "bonne" solution dans un temps "raisonnable".
3. Permettre d'effectuer des compromis entre la qualité de la solution obtenue et la vitesse de convergence.

Dans des environnements dynamiques, on cherche à utiliser des algorithmes d'optimisation rapides et robustes, plutôt que des algorithmes fins et lents.

E.6.2 Méthodes possibles pour résoudre le problème

Pour résoudre notre problème d'optimisation NP-complet plusieurs heuristiques ou méta-heuristiques peuvent être utilisées. Une heuristique est une méthode utilisant une approche essai-erreur pour trouver une solution acceptable en un temps raisonnable (polynomial) à des problèmes d'optimisation difficiles. Les méta-heuristiques sont des heuristiques de haut niveau qui peuvent être spécialisées pour résoudre différents types de problèmes d'optimisation.

Plusieurs méta-heuristiques sont inspirées de systèmes naturels : recuit simulé, algorithmes génétiques, réseaux de neurones, intelligence en essaim. D'autres théories et méthodes peuvent servir à résoudre des problèmes d'optimisation complexes de façon efficace : théorie des graphes, théorie des jeux, processus de décisions de Markov, logique floue, etc.

Dans cette thèse nous avons choisi de contribuer au champ de la méta-heuristique d'intelligence en essaim en raison de son aspect distribué et de son approche montante de la résolution du problème. Dans cette approche (qui appartient au champ de l'éthologie) plusieurs solutions sont évaluées en parallèle, ce qui accélère la convergence vers la meilleure solution.

En général, les méta-heuristiques suivent itérativement les étapes suivantes :

1. Diversification (exploration) : de nouvelles solutions sont explorées par certains individus, de façon à découvrir une meilleure solution au problème.

2. Intensification (exploitation) : certains individus explorent les meilleures solutions trouvées jusqu'alors.

3. Mémoire et apprentissage : la mémoire permet de se souvenir des bonnes solutions and d'oublier les mauvaises solutions. Ainsi, l'algorithme apprend et devient plus rapide pour trouver les meilleures solutions qu'au démarrage de la recherche.

E.6.3 Allocation opportuniste de spectre avec l'intelligence collective

E.6.3.1 Intérêts et obstacles pour l'optimisation distribuée

Dans le contexte de l'optimisation distribuée, on peut représenter chaque utilisateur individuellement avec ses caractéristiques propres et on n'est plus obligé de le considérer comme un utilisateur moyen. Avec l'optimisation distribuée, on peut étudier des distributions géographiques d'utilisateurs dont les caractéristiques évoluent au cours du temps. Ceci nous rapproche de la réalité ou des entités interagissent entre elles.

Dans la nature on trouve énormément de système avec un fonctionnement décentralisé plutôt que centralisé (ex : fourmis, abeilles, hommes, etc), d'où l'importance de mieux comprendre les propriétés fondamentales de l'interaction d'individus dans un groupe. Dans les groupes où le contrôle est distribué, les individus ont une connaissance imparfaite de leur environnement mais ils interagissent avec leurs plus proches voisins de façon à enrichir leur connaissance et à apprendre.

Les progrès techniques réalisés ces dernières années nous permettent d'imaginer des terminaux plus intelligents et qui fonctionneraient avec un contrôle décentralisé. Cependant nous n'y sommes pas encore. Pour désigner un nouveau système, la méthode d'un concepteur classique est une approche descendante qui cherche à contrôler le système de l'extérieur : comment contrôler le système pour qu'il se comporte d'une certaine façon ? Au contraire, l'approche multi-agent (approche montante) cherche à contrôler le système de l'intérieur : comment contrôler les agents pour arriver à un comportement donné du système par émergence ?

E.6.3.2 SI

Pour résoudre le problème de l'allocation de ressource décrit dans la section E.6.1.1, nous allons utiliser une méta-heuristique distribuée appelée "intelligence en essaim" (expression introduite par Beni et Wang en 1989), inspirée de l'observation des insectes sociaux. On parle ici de colonies auto-organisées de plusieurs millions d'individus réalisant des tâches complexes de manière décentralisée. SI est définie comme "une intelligence collective émergeant de groupes d'individus simples" (Bonabeau and al, 1999).

SI offre un excellent moyen de contrôler des systèmes distribués complexes à l'aide de plusieurs règles simples opérées au niveau local, afin d'aboutir à un comportement souhaité (résolution de problème, tâche complexe, etc) au niveau global. En variant les paramètres de contrôle on peut réaliser un compromis entre la vitesse de convergence et la qualité de la solution obtenue.

Les quatre principes basiques de l'auto-organisation qui sont utilisés sont : feedback positif, feedback négatif, amplification, interactions multiples. De plus, les systèmes qui fonctionnent à base de SI s'appuient sur des interactions de type : directes ou indirectes (en utilisant l'environnement comme une mémoire collective gardant l'historique des explorations passées - phénomène appelé stimergie).

Au final, une colonie d'insectes sociaux à les propriétés intéressantes suivantes quand elle résout un problème de façon collective : scalabilité (en nombre d'individus impliqués), flexibilité (résistance aux perturbations externes), robustesse, décentralisation, auto-organisation (les chemins vers les solutions sont émergeants plutôt que prédéfinis). Notamment, nous avons trouvé le comportement des guêpes (division du travail dynamique) très proches de ce que nous cherchions pour l'allocation flexible de spectre.

E.6.4 Conclusion

Dans cette section nous avons introduit le problème d'optimisation à étudier. Ensuite nous avons décrit l'approche que nous allions suivre pour sa résolution, à savoir l'utilisation de la méta-heuristique d'intelligence en essaim, au vue de ses propriétés intéressantes pour le problème de l'allocation dynamique de spectre.

E.7 Réseaux locaux sans fil basés sur de l'OFDM

E.7.1 Liste des notations pour cette section

Le tableau E.3 présente la liste des notations utilisées dans cette section.

Notation	Description
γ_i	Taux de paquets transmis avec succès pour STA i
Δf	Offset en fréquence entre l'interféreur et le récepteur de la victime
d_0	Distance de référence pour la propagation en espace libre
k	Constante de Boltzmann = 1.381 x 10-23 W/Hz/K
n	Exposant pour la perte par propagation
t	Un instant dans le temps
B	Largeur de bande du récepteur
$C_{Blocking}$	Niveau d'interférence reçu d'un interféreur dû au blocking
$C_{Interf\ Total}$	Niveau total d'interférence reçu à un récepteur et venant d'un seul interféreur
$C_{Spurious}$	Niveau d'interférence reçu d'un interféreur dû au émissions parasites
F_T	Index d'équité de Jain pour estimer l'équité d'une topologie
G_j^{Rx}	Gain de l'antenne de STA j (Rx)
G_i^{Tx}	Gain de l'antenne de STA i (Tx)
G_T	Débit utile agrégé pour le système
I_j	Somme de toutes les signaux interférents à STA j
K	Nombre d'éléments de la IFFT
N	Nombre d'utilisateurs en compétition dans une cellule
$N_{thermal}$	Niveau de bruit thermique
PL_{ij}	Perte par propagation entre STA i et STA j
$PL(d)$	Perte par propagation sur la distance d
$PL(d_0)$	Perte par propagation en espace libre (sur la distance d_0)
PR	Protection Ratio du récepteur en mode 1
$PR(Mode_i)$	Protection Ratio du récepteur en mode i
P_i^{Tx}	Puissance de Tx du paquet de STA i
R_{ij}	Niveau de signal reçu à STA j venant de STA i
T	Fréquence d'échantillonnage OFDM
T_{Kelvin}	Température absolue [Kelvin]
S	Efficacité du protocole IEEE 802.11 MAC
S_1	Niveau de sensibilité du mode 1
S_j	Somme de tous les niveaux de signaux reçus à STA j
S_t	Niveau de sensibilité du mode 1 plus 20 dB

Table E.3: Liste des notations pour cette section

E.7.2 OFDM

OFDM est une technique de transmission qui utilise du multiplexage en fréquence (plusieurs sous-porteuses) où l'information est transmise sur chaque SC. En plus d'offrir des hauts débits pour les systèmes sans fil, la technologie OFDM a les propriétés clés suivantes : offre un bon packing en fréquence ([42], [43], [44]), offre une bonne granularité de découpage en fréquence, bonne résistance aux multi-trajets.

Cependant, l'implémentation de la technologie OFDM doit faire face à de nombreux challenges tels que : des problèmes de non-linéarité, des limitations de PAPR, sensibilité plus importante à l'offset en fréquence et à la dérive que les systèmes à simple porteuse.

Malgré ces limitations, la technologie OFDM est très prometteuse et de nombreux standards vont l'utiliser comme base pour la couche physique tels que : ETSI DAB, ADSL, UWB, IEEE 802.16-2004 standard pour wireless MAN (WiMAX), OFDM est un candidat pour les systèmes de communications 4G mobiles, les standards ETSI DVB-T (1997) et DVB-H (2004), les standards IEEE wireless LAN (IEEE 802.11a [2] (1999) et IEEE 802.11g (2002).

E.7.3 Presentation du IEEE 802.11

Le standard IEEE 802.11 est un standard de réseau local sans fil qui comprend plusieurs versions, comme présenté ci-après avec une explication de leur mission respective :

- 802.11a: pour des débits plus élevés à 5.4 GHz
- 802.11b: standard initial à 2.4 GHz
- 802.11c: opération de pont
- 802.11d: effort global d'harmonisation entre les instances de régulations internationales
- 802.11e: QoS pour les services temps réels incluant de la differentiation de service
- 802.11f: interopérabilité entre les AP
- 802.11g: pour des débits plus élevés à 2.4 GHz
- 802.11h: mécanisme de contrôle de puissance et de sélection dynamique de fréquence pour une cellule, en plus du 802.11a
- 802.11i: amélioration de la sécurité au niveau de la couche MAC
- D'autres extensions: 802.11n pour avoir des très hauts débits (amélioration de la couche MAC et de la couche PHY), et IEEE 802.11y [51] lancé récemment pour arriver à une coexistence avec d'autres systèmes IEEE 802.11.

E.7.4 Limitations et problèmes connus des systèmes 802.11

Voici quelqu'uns des problèmes et limitations rencontrées par les des systèmes 802.11 :

- Coexistence des modes de transmission : Le canal est partagé par tous les utilisateurs de la cellule. Il existe plusieurs modes de Tx donnant lieu à plusieurs durée de Tx. Dans une cellule avec une variété de modes de Tx, la vitesse de Tx des utilisateurs se trouve ralenti par l'utilisateur le plus lent.
- Collisions et noeuds cachés : En mode DCF le contrôle de la coordination est distribué entre toutes les STAs causant des collisions (arrive dès que 2 noeuds non cachés transmettent simultanément ou quand 2 noeuds cachés transmettent des paquets qui se chevauchent, résultant dans aucun paquet correctement décodé). Un noeud caché est à portée de la destination d'un message, mais hors de portée de son transmetteur. La conséquence est un accroissement du nombre de collisions et des performances dégradées.
- Effet de capture : arrive dès qu'un récepteur est capable de décoder correctement un message transmis en même temps que plusieurs autres. L'effet de capture est toujours plus bénéfique pour le débit du système que la collision car au moins un paquet est décodé.

E.7.5 Paramètres étudiés via les simulations

Le but des simulations que nous avons faites était de quantifier les comportement de la technologie 802.11a pour différents déploiements de cellules et d'utilisateurs.

Les métriques étudiées dans les simulations sont :

- Le débit utile: défini comme la rapport entre la charge utile envoyée et le temps total nécessaire pour la transmettre.
- L'efficacité du protocole MAC S: en terme de réservation du canal et de la transmission de la data avec succès (exprimé en %) tel que :

$$S = \frac{\text{Nombre de trames ACK attendues en réception}}{\text{Nombre de RTS trames transmises}}$$

  avec "trame ACK attendue en réception" correspond à une réponse avec succès à une "RTS trame transmise".
- Le taux de collision
- Le délai moyen d'accès: défini comme le temps moyen passé par un paquet dans la queue MAC, i.e., depuis l'instant où il est mis dans la queue jusqu'à ce qu'il soit correctement transmis.

Pour déterminer l'équité du protocole MAC on utilise l'index de Jain [52]. Egalement, pour caractériser la dispersion extrême d'une série de valeurs, on utilise la rapport entre la valeur maximum et minimum parmi tous les utilisateurs.

E.7.6 Résultats de simulation

E.7.6.1 Impact de la distribution des utilisateurs

Le but de ce scenario est d'investiguer l'impact de la distribution des utilisateurs sur les performances du protocole MAC. Nous avons montré par simulation que le protocole MAC était très sensible à la topologie de la cellule, donnant lieu à un système spatialement non équitable dans l'accès au canal radio et favorisant les utilisateurs très près du point d'accès.

Nous avons simulé un temps de 200s, avec un nombre d'utilisateurs (uniformément répartis dans la cellule) dans la cellule variant de 5 à 40. La figure E.5 présente S versus le nombre d'utilisateurs dans la cellule et de la distance à l'AP.

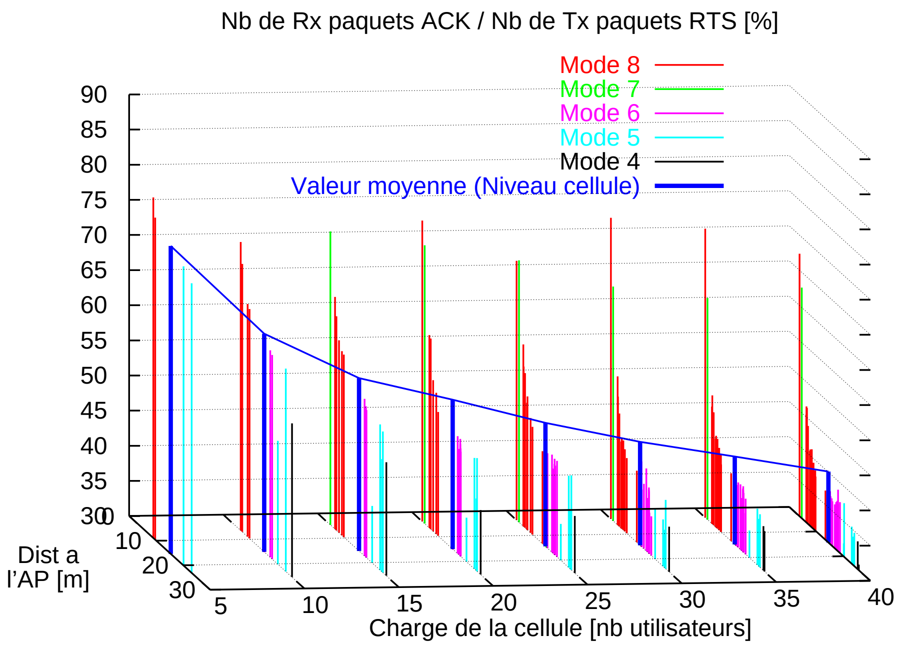


Figure E.5: S (%) Versus nombre de noeuds et distance à l'AP

Les noeuds n'ayant que peu d'opportunités d'accès au canal souffrent d'un pourcentage important de collisions et d'un temps d'attente long entre 2 Tx, donnant lieu à de piètres performances. L'iniquité spatiale est évidente étant donné que même les utilisateurs transmettant la même quantité de data utile ont des opportunités différentes d'accéder au canal.

E.7.6.2 Interférence entre les cellules en 802.11a (pas de contrôle de puissance)

On a simulé plusieurs scénarios avec plusieurs cellules et plusieurs distances entre les accès points des différentes cellules. Le but était d'investiguer s'il y avait des contraintes de déploiements à respecter dans un cas avec plusieurs cellules pour conserver les performances du réseau, si l'on pouvait facilement réutiliser les canaux en pool ou bien s'il fallait plus de coordination.

La conclusion est qu'il faut prendre en compte la séparation en fréquence et en distance entre les APs afin de se prémunir contre des problèmes d'interférence intra-système en 802.11a. Il faut donc plus de coordination entre les APs afin que des utilisateurs situés dans des cellules voisines ne s'interfèrent pas pendant le déroulement du protocole MAC. De plus, une meilleure isolation des canaux pourrait également aider à mieux séparer les canaux et donc mieux isoler des cellules WLAN voisines.

E.7.7 Conclusion

Dans cette section nous présenté la technique OFDM avec ses avantages, ses inconvénients et ses perspectives. Nous avons également présenté quelques simulations réalisées sur la mise en évidence de quelques problèmes du 802.11a.

Dans la suite de notre travail nous essayerons de trouver des algorithmes intelligents d'allocation de spectre pour des systèmes WLAN qui résolvent les problèmes identifiés dans cette section.

E.8 Solution UL collective pour améliorer les performances du système

E.8.1 Liste des notations pour cette section

Le tableau E.4 présente les notations utilisées dans cette section.

E.8.2 OFDMA collectif en voie montante

Nous utilisons de l'OFDMA (une extension multi-utilisateurs de l'OFDM). L'intérêt est de pouvoir adapter l'allocation des SCs entre les utilisateurs au gré des variations des conditions radio. La granularité minimum pour une ressource radio sera l'utilisation d'une sous-porteuse OFDM pendant une période de temps. L'OFDMA offre un découpage du spectre avec une très bonne granularité et sans forcément avoir des blocs de fréquences contigus permettant une meilleure utilisation des trous dans le spectre [39].

Nous nous intéressons aux mécanismes permettant de réaliser une allocation OFDMA en voie montante. En utilisant la méta-heuristique d'intelligence en essaim, nous trouvons un algorithme distribué permettant de trouver une allocation auto-organisée des SCs entre les utilisateurs.

E.8.3 Challenge de modélisation de l'intelligence en essaim avec de l'OFDMA

E.8.3.1 Introduction

Notre algorithme concerne essentiellement la couche MAC, même si nous avons également développé un mécanisme de réservation parallèle qui concerne la couche PHY. Le travail le plus difficile a été celui de la modélisation : passer d'un monde d'insectes sociaux à un monde d'agents situés dans des terminaux sans fil.

On distingue le noeud central (semblable à un AP en 802.11) et autour de lui des noeuds (terminaux utilisateurs). Les noeuds du réseau coopèrent de façon distribuée pour arriver à une allocation de SCs qui maximise la capacité somme totale (somme de toutes les capacités sur chaque SC). L'algorithme exploite la diversité multi-utilisateurs en essayant d'allouer les SCs aux utilisateurs les plus à même de fournir le plus de capacité.

E.8.3.2 Problème d'optimisation

Le problème d'optimisation est le même que celui présenté dans la section E.6.1.1, mais précisé pour le contexte OFDMA. En utilisant les notations du tableau E.4, on formule le problème tel

Notation	Description
α, β	Paramètres contrôlant la forme de la fonction $f(x)$
$\alpha_{ij}(t_1, t_2)$	Temps d'utilisation du service par le noeud i sur la SC j entre t_1 and t_2.
α_{Fair}, β_{Fair}	Paramètres contrôlant forme fonction f_{ij} (équité en temps de service)
α_s, β_s	Paramètres contrôlant la forme de la fonction p_{ij}
ϵ	Paramètres utilisés pour trouver x_l tel que $f(x_l) = \varphi - \epsilon$ et x_u tel que $f(x_u) = -\varphi + \epsilon$
γ_{ij}	Valeur de SNR pour le noeud i sur la SC j
γ_{max}, γ_{min}	Respectivement les valeurs possibles max. et min. de SNR pour le système
θ_{ij}	Valeur de seuil dans l'algorithme
θ_{init}	Valeur initiale de seuil après reset
θ_{max}, θ_{min}	Respectivement les valeurs possibles max. et min. de seuil
ξ	Paramètre contrôlant le niveau d'équité en temps de service pour le système
ϕ_i	Pondération du flux i
φ	Paramètre contrôlant la variation maximum de $\Delta\theta_{ij}$ par time slot de négociation
$\Delta\theta_{ij}$	Variation du seuil dans l'algorithme
Δc	Fenêtre de valeurs de capacité pour le système
Δx	Fenêtre de valeurs de x pour le système
a, b	Paramètres contrôlant la transformation de c_{ij} en T_{ij}
a_{ij}	Valeur binaire utilisée pour indiquer si le noeud i a sélectionné la SC j
c_{ij}	Capacité de Shannon atteignable par le noeud i sur la SC j
c_{max}, c_{min}	Respectivement les valeurs de capacity max. and min. posibles dans le système
$c_{max\ j}$, $c_{min\ j}$	Respectivement les valeurs c_{ij} max. and min. reçues au noeud central pour SC j
c_l, c_u	Respectivement les valeurs possibles max. et min. de capacité pour le système
$f(i, j, k, x)$	Fonction controllant le taux d'équité dans le système
$f(x)$	Fonction de seuillage pour notre algorithme
i, j	Respectivement les index pour les noeuds et pour les SCs
n_i	Nombre de SCs allouées pour le noeud i
n_{max}, n_{min}	Resp. nombres max. et min. de SCs utilisables simultanément par un noeud actif
p_{ij}	Probabilité qu'un agent i réalise la tâche j
q	Pas de quantification
$q(c_{ij})$	Valeur quantifiée c_{ij}
x_{ij}	Résultat de la transformation des valeurs de c_{ij} sur une échelle x
x_l, x_u	Respectivement les valeurs transformées de c_l et c_u
D	Index du nombre de trames de données transmises pendant une simulation
D_{ij}	Nombre de fois qu'un noeud i a utilisé la SC j
E	Nombre d'éléments utilisés pour quantifier les valeurs de c_{ij}
F_j	Index d'équité de Jain pour la SC j
K	Nombre total de trames de données transmises par simulation
$K_i(p)$	Compteur de crédit pour le flux i à l'instant p
$L_i(p)$	Longueur du paquet en tête de queue pour le flux i à l'instant p
M	Nombre de noeuds
N	Nombre de SCs à allouer
T_D	Durée de la phase de transmission des données
T_F	Durée d'une trame incluant négociation et transmission des données
T_{ij}	Durée des messages de burst en voie montante
T_N	Durée des phases de négociation
T_{DL}, T_{UL}	Resp. les durées des time slot de négociation en voies descendante et montante
$T_{timeout}$	Durée de l'expiration pour la phase de négociation
$U_i(p)$	Coût estimé pour l'utilisateur i de transmettre le p^{ime} paquet
V	Valeur pour comparer plusieurs noeuds dans l'algorithme d'équité centralisé

Table E.4: Liste des notations pour cette section que l'objectif est de :

$$\max \sum_i \sum_j a_{ij} c_{ij}$$
$$\text{t.q.:} \begin{cases} (1) & \{n_{min}, n_{max}\} \in [0; N] \\ (2) & \forall i,\, n_i = 0 \Leftrightarrow \sum_j a_{ij} = 0 \\ (3) & \forall i,\, n_i \neq 0 \Leftrightarrow \\ & n_{min} \leq \sum_j a_{ij} \leq n_{max} \\ (4) & \forall j,\; \sum_i a_{ij} \leq 1 \end{cases} \tag{E.2}$$

En d'autres termes, l'objectif est de trouver le vecteur d'allocation final qui maximise la capacité somme pour le système, tout en respectant les contraintes du problème d'optimisation.

E.8.3.3 Optimisation distribuée

Les scénarios considérés consistent dans une cellule avec un noeud central et plusieurs noeuds géographiquement distribués dans la cellule. L'idée est que les noeuds ont la responsabilité du choix de leur allocation de spectre et non plus le noeud central.

On considère que le temps est décomposé en trames de durée variable $T_F = T_N + T_D$. Chaque trame comprend une phase de négociation (durée variable T_N) et une phase de transmission des données (durée fixe T_D). La négociation consiste à chercher la meilleure allocation des SCs entre les utilisateurs, tandis que la phase de Tx des données utilise l'allocation trouvée pour transmettre. La figure E.6 présente l'enchaînement temporel des phases et inclut un exemple de négociation dans une cellule de 3 utilisateurs.

Comme décrit dans la figure E.6, la phase de négociation consiste dans une succession de :

- UL time slot: Des messages sont simultanément envoyés (vers le noeud central) par les noeuds en compétition. Chaque noeud qui le désire acquiert une SC donnée transmet un message de burst en utilisant cette SC.

- DL time slot: Ensuite, le noeud central répond en renvoyant un message à tous les noeuds de la cellule, et comprend les valeurs min et max de capacité qu'il a détecté sur chaque SC.

Ces 2 phases sont répétées jusqu'à ce que la négociation se termine. Ensuite commence la phase de Tx des données.

Notre algorithme résout les scénarios suivants : $N = M$, $N < M$ and $N > M$, ainsi que pour n'importe quel valeurs de n_{min} et n_{max}.

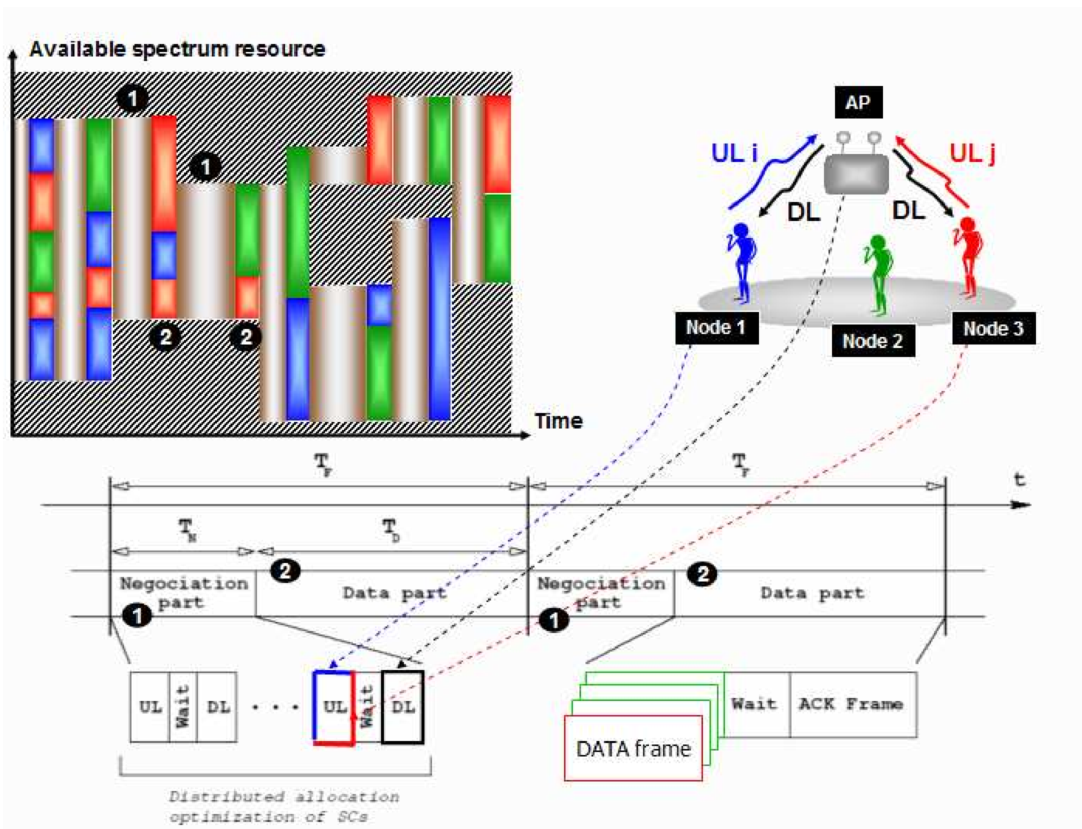


Figure E.6: Description des phases de l'algorithme d'optimisation

E.8.3.4 Description et modélisation de l'heuristique

La méta-heuristique d'intelligence en essaim a donné de très bons résultats pour optimiser des problèmes dynamiques et distribués d'allocation de tâches dans l'industrie [60], [61]. Une version améliorée (de laquelle nous sommes partis) utilise un seuil variable au cours du temps [63] qui permet de mieux s'adapter aux changements de l'environnement.

L'objectif de cette partie est de décrire comment l'algorithme parvient à trouver l'optimum. Un agent correspond à une entité logicielle qui est responsable du suivi d'une SC pour un utilisateur. Un noeud contient (1) un contrôleur de contraintes (qui vérifie que les contraintes du problème sont respectées par le noeud) et (2) N agents (un par SC). Il y a donc N agents par noeud et au total MN agents dans le réseau. Les agents interagissent de façon interne (pour respecter la contrainte 4 du problème) et externe (pour respecter les contraintes 2 et 3 du problème) au terminal, à chaque fois en échangeant et en comparant leurs valeurs de capacités. A la fin de la négociation il y a au maximum un noeud actif par SC, mais un noeud peut se spécialiser dans plusieurs SCs (au maximum n_{max}).

Les sections suivantes se concentrent sur les étapes et les opérations de la phase de négociation dans les noeuds et dans le noeud central :

- <u>Operations internes aux noeuds, suite à la réception d'un message du noeud central</u>: Algorithme pour chaque noeud i:
  - Etape 1: Pour chaque agent j, calculer les valeurs initiales de θ_{ij}.
  - Etape 2:
    - a) Déterminer la probabilité que chaque agent j soit actif (ON) au prochain TS de contention suivant l'équation :

$$p_{ij} = \frac{1}{1 + \alpha_s e^{\beta_s \theta_{ij}}} \tag{E.3}$$

      avec α_s et β_s deux paramètres contrôlant la forme de la fonction p_{ij}.
    - b) Déterminer le status de chaque agent j (ON or OFF) : si p_{ij} est supérieur ou égal à une valeur aléatoire tirée uniformément entre 0 et 1, alors l'agent est actif (ON), sinon il est inactif (OFF).
  - Etape 3: Vérifier que le nombre d'agents actifs est inférieur à n_{max} et désactiver (en choisissant les n_{max} agents avec les valeurs de p_{ij} d'être ON les plus grandes) suffisamment d'agents actifs jusqu'à ce qu'il n'y en ait moins que n_{max} d'actifs.
  - Etape 4: Transmettre un message de burst en UL vers le noeud central en utilisant seulement les SCs actives pour ce noeud i.
  - Etape 5: Attendre la réponse du noeud central, consistant à retourner les valeurs de capacité $c_{max\ j}$ et $c_{min\ j}$ pour tous les j.
  - Etape 6: Pour chaque j, calculer $\Delta\theta_{ij}$ (en fonction de $c_{max\ j}$ et c_{ij}) la variation du seuil θ_{ij} en résultat de la nouvelle valeur reçue $c_{max\ j}$ par agent j.
  - Etape 7: Pour chaque j, mettre à jour la valeur du seuil θ_{ij} :

$$\theta_{ij} \leftarrow \theta_{ij} + \Delta\theta_{ij} \tag{E.4}$$

  - Aller à l'étape 2 ... temps que la phase de négociation n'est pas terminée. La fin de la négociation arrive lorsque toutes les contraintes du problèmes sont repectées et notamment qu'il n'y a plus au maximum qu'un seul noeud par SC.

- Opérations externes aux noeuds, dont les résultats sont analysés dans le noeud central:
  - Question (1): Comment communiquer entre les noeuds ? Les noeuds ne communiquent pas directement entre eux mais ils passent par le noeud central en lui envoyant des messages simultanément sur les SCs qu'ils souhaitent acquérir.
  - Question (2): Quelle information inclure dans les messages échangés en parallèle entre les noeuds ? Les messages échangés codent la valeur de capacité du noeud sur chaque SC. Le message a une durée variable et inversement proportionnelle à la valeur de capacité correspondante.

E.8.4 Impact de la variation des paramètres de contrôle de l'algorithme

Nous avons étudié les variations de paramètres suivants, avec les conclusions en regard :

- Variation de la pente de la fonction de seuillage : plus β est grand (i.e. plus la pente de la fonction de seuillage est importante), plus la convergence est lente, pour un résultat en terme de capacité totale insensible à la variation de β.
- Variation des capacités d'apprentissage et d'oubli de l'algorithme pendant la phase de négociation : presque pas d'impact sur la capacité totale obtenue.
- Variation des capacités d'apprentissage et d'oubli de l'algorithme entre les phases de négociation : le processus d'apprentissage est bénéfique pour le système. Par contre, resetter la mémoire du système doit se faire en corrélation avec la variation des paramètres d'entrée afin de bénéficier pleinement de l'intelligence de l'algorithme. Sinon, une mémoire plus à jour a plus tendance à ralentir l'algorithme dans sa course à la solution finale.

En conclusion, cette partie a permis de mettre en évidence le comportement de l'algorithme lorsqu'il est soumis à divers changements de ces paramètres de contrôle.

E.8.5 Impact de la variation des paramètres d'entrée de l'algorithme

Les sections suivantes montrent l'impact de la variation des paramètres d'entrée de l'algorithme sur sa vitesse de convergence et la qualité de la solution finale. Nous avons étudié les scénarios suivants :

- Variation soudaine des conditions radios entre les phases de négociation : montre le potentiel de l'algorithme à s'adapter en fonction du problème.
- Variation soudaine des conditions radios pendant la phase de négociation : Accroissement important de la durée de négociation en fonction de la probabilité de changement des

canaux. En cas de changement brusque des conditions radio (sans connaissance à priori), la meilleure méthode consiste à resetter les valeurs de θ_{ij}.

- Impact des variations de M et N : Pour N donné, la capacité somme totale bénéficie d'un accroissement de M (diversité multi-utilisateur). De plus, la capacité somme totale augmente avec N.
- Impact des variations de n_{min} et n_{max} : Certaines valeurs du couple $n_{min}; n_{max}$ requièrent une plus longue négotiation que d'autres, notamment le cas 1; 1. Par contre, dès que l'on a $n_{min} = 1$ et $n_{max}! = 1$, la convergence se fait relativement rapidement.
- Impact d'une variation brusque de M: Les agents ont été capables de s'adapter à une variation soudaine de M, tout en conservant une durée de négociation insensible à la valeur de M. La capacité somme totale variait avec M (diversité multi-utilisateur).

Ces simulations ont permis de caractériser le comportement de l'algorithme pour différentes variations des paramètres d'entrée.

E.8.6 Scénarios avec 2 classes de service en compétition

Dans cette section nous avons implémenté un algorithme distribué d'équité en temps de service rajouté au-dessus de l'algorithme collaboratif d'allocation de spectre précédent. Avec cet algorithme, l'opérateur est capable de contrôler le niveau d'équité de son réseau en variant la valeur de ξ. Ensuite les utilisateurs vérifient localement qu'ils ont bien le niveau d'équité demandé par l'opérateur. Un tel algorithme est très efficace pour les prochaines générations de systèmes locaux sans fil, afin d'utiliser de façon opportuniste une partie du spectre en fonction des besoins des utilisateurs.

On a rajouté une 5ième contrainte au problème d'optimisation présenté dans la section E.6.1.1 : L'objectif du problème d'optimisation est :

$$\begin{aligned} &\max \sum_i \sum_j a_{ij} c_{ij} \\ &\text{t.q.:} \begin{cases} (1)-(4) & \text{même contraintes que dans l'équation E.5 de la section E.6.1.1} \\ (5) & Pr(|\alpha_{ik}(t_1,t_2) - \alpha_{jk}(t_1,t_2)| \geq x) \leq f(i,j,k,x) \end{cases} \end{aligned} \tag{E.5}$$

On a modifié l'algorithme d'équité en temps de service proposé dans [65] afin qu'il soit distribué et utilisable dans un contexte OFDMA. L'équité est contrôlée de façon statistique par le paramètre ξ qui modifie le coût d'utilisation d'une SC. Cet algorithme est exécuté de façon interne par les noeuds.

Nous avons utilisé cet algorithme pour assurer un accès équitable au spectre entre 2 classes de service concurrentes. Dans le scénario simulé, chaque classe de service a son besoin propre en quantité de spectre (n_i):

- Service 1: $n_{min} = 1$ and $n_{max} = 6$.

- Service 2: $n_{min} = 7$ and $n_{max} = 15$.

Nous avons comparé les résultats obtenus avec et sans équité. L'équité est ici exprimée en temps d'utilisation d'une SC donnée. Dans chaque cas on a les résultats suivants :

1. Nombre total de fois que chaque utilisateur a utilisé un groupe de i SCs pour Tx ses données.
2. Nombre total de SCs utilisées par chaque utilisateur pour Tx ses données, en agrégeant toutes les classes d'utilisation d'un groupe de i SCs. Comme exemple, un utilisateur utilisant i SCs pour Tx ses données ajoutera i à ce nombre.

Les résultats sont inclus dans les figures E.7 (sans équité) et E.8 (avec équité).

On remarque sur ces figures que sans équité les classes de service en compétition ne coexistent pas de façon équitable, alors que l'équité dans l'accès à la ressource radio est clairement rétablie avec notre algorithme d'équité distribué.

E.8.7 Conclusion

Nous avons détaillé le fonctionnement de notre algorithme basé sur la méta-heuristique d'intelligence en essaim. Les paramètres de contrôles de l'algorithme ainsi que les paramètres d'entrée du problème ont été variés pour évaluer le comportement de notre solution dans de multiples configurations. De plus, un algorithme de contrôle d'équité (en temps d'utilisation d'une ressource) distribué a été implémenté avec succès, et permet de gérer des scénarios avec plusieurs classes de service.

Une idée clé de l'approche montante est de contrôler le système de l'intérieur. La complexité du système est recrée suite aux multiples interactions entre les noeuds.

Par simulations nous avons donc pu valider les propriétés suivantes de l'algorithme :

- Robustesse face changements de l'environnement (nombre de noeuds, conditions radio).
- Flexibilité: l'algorithme peut s'adapter à de nombreuses configurations de M, N, n_{min}, n_{max} et avec plusieurs classes de services.
- Scalabilité en nombre de noeuds simultanément en compétition.
- Equité: l'algorithme gère l'équité en temps de service entre plusieurs utilisateurs.

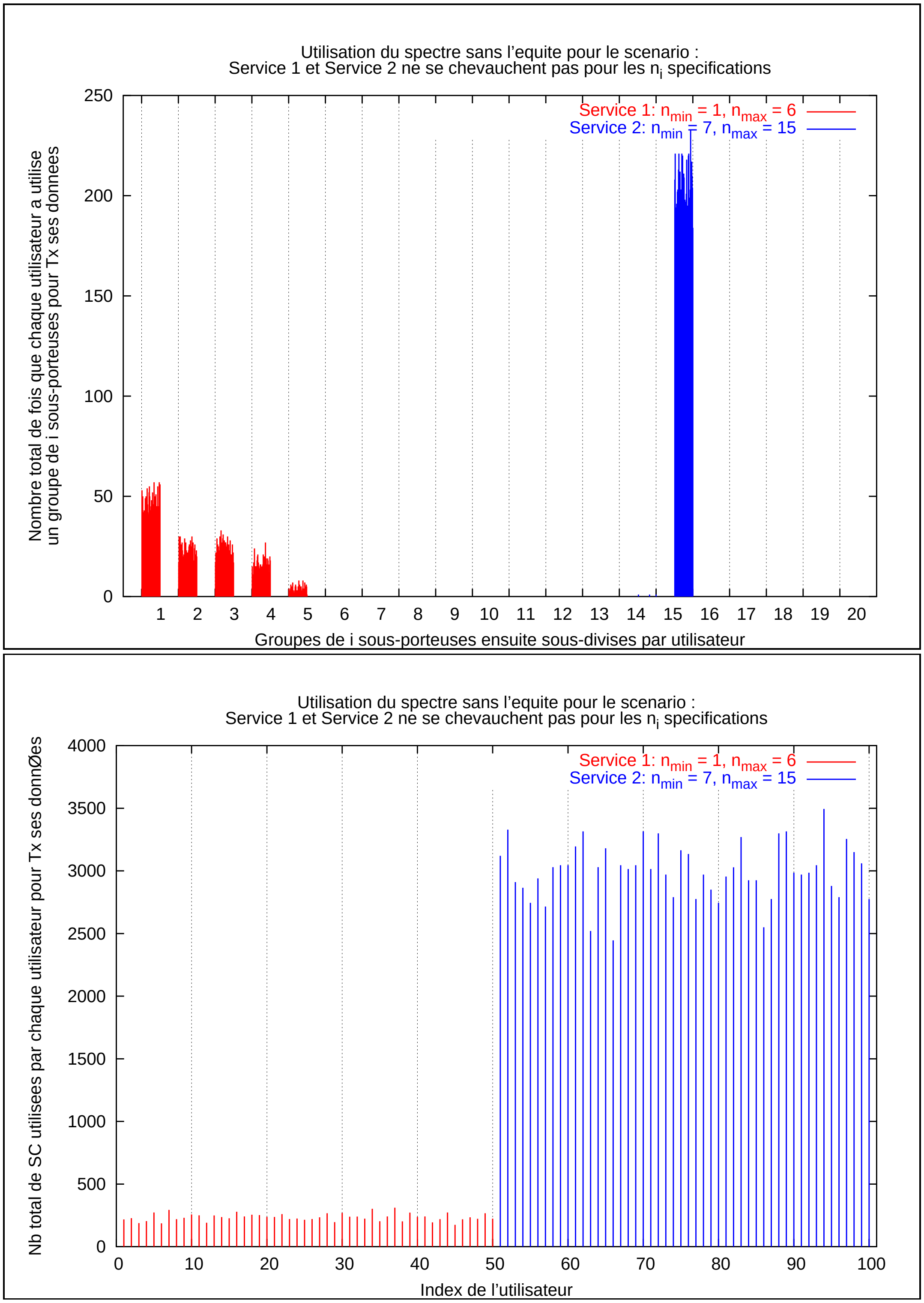


Figure E.7: Résultats sans équité

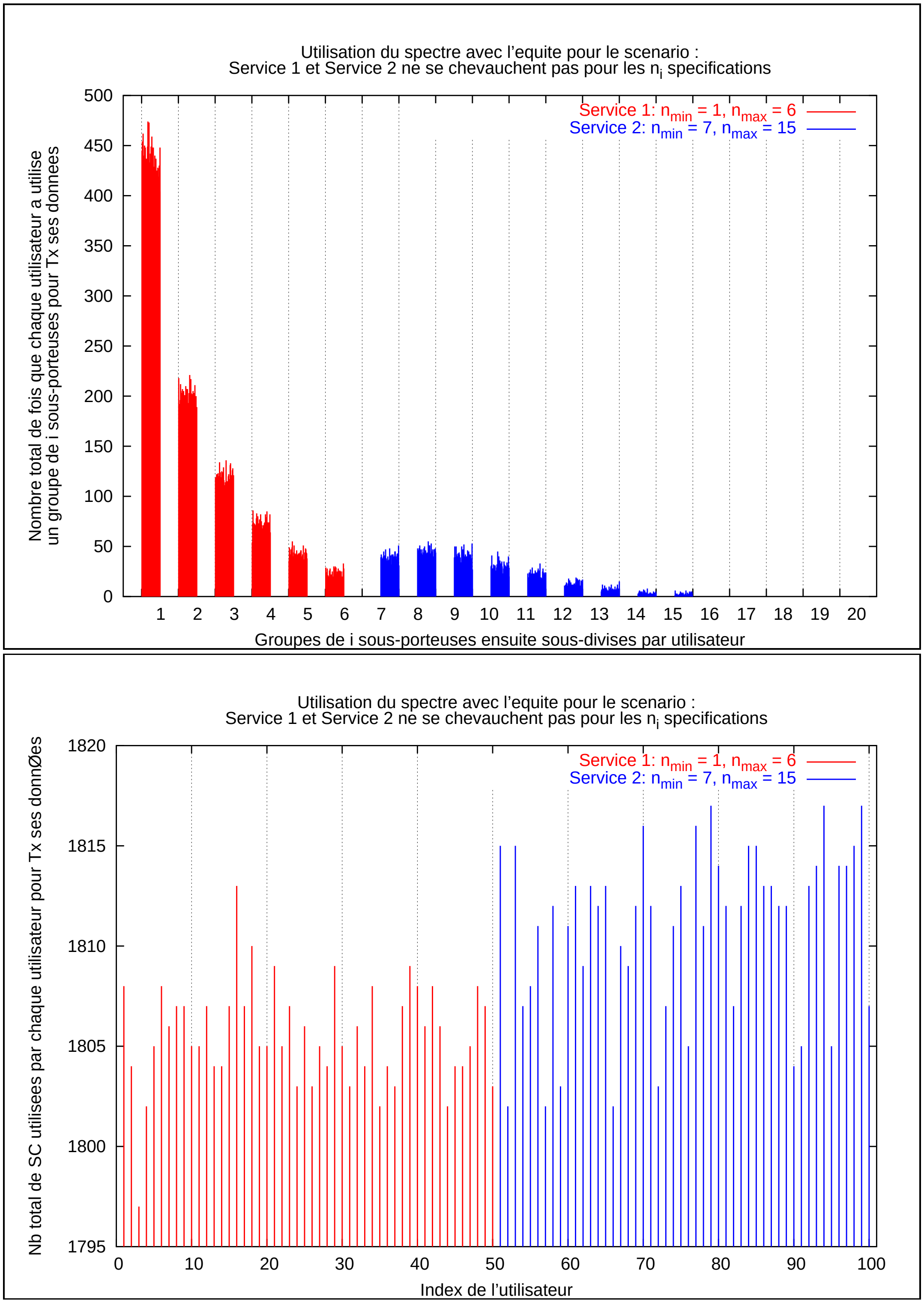


Figure E.8: Résultats avec équité

E.9 Conclusions et futurs axes de recherche

E.9.1 Conclusions

Dans la plupart des pays, les bandes de spectre sont déjà saturées. La rareté du spectre vient des méthodes d'allocation rigides employées par les régulateurs. Pour palier au nombre croissant d'utilisateurs de services de communication sans fil, il est nécessaire d'avoir recours a des algorithmes d'allocation du spectre plus dynamiques. De tels algorithmes intelligents sont capables de tirer parti des trous dans le spectre de façon à offrir un meilleur service aux utilisateurs. Les nouvelles perspectives d'allocation de spectre telles que l'échange de spectre, la mise en commun du spectre, seront utilisées avec succès par les radios cognitives qui sont des terminaux intelligents.

En utilisant la théorie des files d'attente, nous avons comparé plusieurs méthodes d'accès aux canaux radios entre elles, chacune offrant un partage différent de la ressource radio entre les utilisateurs. La méthode offrant le meilleur débit pour une cellule composée de plusieurs utilisateurs est celle qui adapte l'allocation de façon dynamique en fonction de la qualité de leurs liens.

Dans cet thèse, nous avons apporté des pistes de solutions au problème de l'allocation dynamique de spectre en utilisant une approche collaborative. Cela a donné de bons résultats et l'algorithme a les propriétés suivantes : scalabilité, robustesse, flexibilité dans la négociation.

Nous avons proposé un algorithme distribué d'allocation de sous porteuses entre plusieurs utilisateurs. Cet algorithme a également la capacité de contrôler l'équité du partage de spectre entre les utilisateurs en fonction du temps d'utilisation. Cet algorithme basé sur l'approche d'intelligence en essaim permet d'aboutir à une allocation auto-organisée du spectre, répondant le mieux aux besoins des utilisateurs.

E.9.2 Futurs axes de recherche

Il pourrait être intéressant d'étudier les conséquences sur les performances des changements dans les contraintes de notre algorithme d'optimisation.

De plus, la méta-heuristique "d'intelligence en essaim" peut être utilisée pour résoudre d'autres problèmes du monde du sans fil. Le succès des méta-heuristiques inspirées de systèmes naturels montre bien qu'il faut accroître les échanges inter-domaines pour enrichir la recherche dans son ensemble. Egalement, d'autres méthodes peuvent être envisagées telles que la théorie de jeux (très en vogue actuellement). Mais dans tous les cas nous pensons que des méthodes collaboratives distribuées sont bénéfiques pour coordonner l'amélioration des performances du système.

Enfin, de nombreux progrès vont se faire au niveau de la couche physique et qui seront ensuite exploités par les couches supérieures pour accroître le gain du système. Notamment, de nouvelles découvertes en théorie de l'information et concernant la théorie de la communication de groupes vont apparaître dans le futur et seront moteurs pour le domaine de l'allocation flexible de spectre.

Bibliography

[1] Standard IEEE 802.11. Ieee 802.11. wireless lan medium access control (mac) and physical layer (phy) specifications. Technical report, IEEE, New York, 1999.

[2] IEEE Std 802.11a-1999 (supplement to IEEE Std 802.11-1999). Part ii: Wireless lan medium access control (mac) and physical layer phy specifications: High speed physical layer in 5 ghz band. Technical report, IEEE, New York, September 1999.

[3] D. Reed. Comments on fcc et docket 02-135. Comments for FCC Spectrum Policy Task Force on Spectrum Policy, July 2001.

[4] K. Werbach. Open spectrum: The new wireless paradigm. *New America Foundation - Spectrum Policy Program*, October 2002.

[5] G. Hardin. The tragedy of the commons. *Science*, (162):1243–1248, 1968.

[6] J.M. Peha. Wireless communications and coexistence for smart environments. *IEEE Personal Communications*, October 2000.

[7] Shared Spectrum Company. Spectrum measurements campaign. *http://www.sharedspectrum.com/?section=nsf_summary.*

[8] D. Grandblaise J.C. Dunat and C. Bonnet. Efficient ofdma distributed optimization algorithm exploiting multi-user diversity. *In Proceedings of IEEE DySPAN*, pages 233–242, November 2005.

[9] S.S. Wildman C. Ting and J.M. Bauer. Government policy and the comparative merits of alternative governance regimes for wireless services. *In Proceedings of IEEE DySPAN*, pages 401–419, November 2005.

[10] W. Lehr and J. Crowcroft. Managing shared access to a spectrum commons. *In Proceedings of IEEE DySPAN*, pages 420–444, November 2005.

[11] eTForecasts company. Cellular market forcasted future trends. *http://www.etforecasts.com/products/ES_cellular.htm.*

[12] J.M. Peha. Spectrum management policy options. *IEEE Communications Surveys*, 1(1), Fourth quarter 1998.

[13] I. Katzela and M. Naghshineh. Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey. *IEEE Personal Communications*, pages 10–31, June 1996.

[14] IST FP5 DRiVE Project. http://www.ist-drive.org.

[15] IST FP5 OverDRiVE Project. http://www.ist-overdrive.org.

[16] D. Reed. Reed's open spectrum page. http://www.reed.com/, 2003.

[17] D. Reed. Why spectrum is not property - the case for an entirely new regime of wireless communications policy. http://www.reed.com/dprframeweb/dprframe.aspsection=paper&fn=openspec.html, 2001.

[18] A. Lippman and D.P. Reed. Viral communications. *Media laboratory research*, May 2003.

[19] D. Weinberger. The myth of interference. http://www.salon.com/tech/feature/2003/03/12/spectrum/, 2003.

[20] K. Werbach. Open spectrum: The new wireless paradigm. *New America Foundation - Spectrum Policy Program*, October 2002.

[21] D.P. Reed J.H. Saltzer and D.D. Clark. The end-to-end arguments in system design. *ACM Transactions in Computer Systems*, pages 277–288, 1984.

[22] J. Mitola. Cognitive radio - an integrated agent architecture for software defined radio. *PhD Thesis of the Royal Institute of Technology*, May 2000.

[23] J. Mitola. Cognitive radio for flexible mobile multimedia communications. *IEEE International Workshop on Mobile Multimedia Communications*, pages 3–10, 1999.

[24] J. Mitola and G.Q. Maguire. Cognitive radio: Making software radios more personal. *IEEE Personal Communications*, pages 13–18, August 1999.

[25] C. Evci. First report and order in et docket no. 00-47, 16 fcc rcd 17373. 2001.

[26] P.J. Kolodzy. The role of cognitive radio on novel spectrum management. *EU Workshop on Advanced wireless technologies: Implications for spectrum management*, October 2003.

[27] P. Mannion. Sharing spectrum the smarter way. *http://www.commsdesign.com/showArticle.jhtmlarticleID=18700443*, April 2004.

[28] G. Staple and K. Werbach. The end of spectrum scarcity. *IEEE Spectrum*, March 2004.

[29] F. Capar et al. Comparison of bandwidth utilization for controlled and uncontrolled channel assignment in a spectrum pooling system. *IEEE VTC'02*, March 2002.

[30] F. Capar and F. Jondral. Resource allocation in a spectrum pooling system for packet radio networks using ofdm/tdma. *IST Mobile & Wireless Telecommunications Summit 2002*, June 2002.

[31] The next generation (xg) vision - request for comments, version 1.0. *prepared by BBN Technologies, Cambridge, Ma.*, June 2003.

[32] The next generation (xg) architectural framework - request for comments, version 1.0. *prepared by BBN Technologies, Cambridge, Ma.*, June 2003.

[33] D. Schaefer. Adaptive spectrum utilization with software defined radios. *Presentation of the MITRE Corporation*, October 2002.

[34] B. Baynat. *Théorie des files d'attentes, des chaînes de Markov aux réseaux à forme produit.* Editions HERMES Science Publications, 2000.

[35] G. Mazzini F.H. Fitzek, D. Angelini and M. Zorzi. Design and implementation of an enhanced ieee 802.11 mac protocol for multihop coverage extension. *IEEE Wireless Communications*, December 2003.

[36] L. Kleinrock. Distributed systems. *Communications of the ACM, Special Issue*, 28(11), November 1985.

[37] M.W. Macy and A. Flache. Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences USA, 99 (10)*, pages 7229–7236, 2002.

[38] K.B. Letaief C.Y. Wong, R.S. Cheng and R.D. Murch. Multiuser ofdm with adaptive subcarrier, bit, and power allocation. *IEEE Journal on Selected Areas in Communications*, 17(10):1747–1758, October 1999.

[39] C.J. Kikkert E. Lawrey and. Adaptive frequency hopping for multiuser ofdm. *Second International Conference on Information, Communications and Signal Processing, ICICS 99, Singapore*, December 1999.

[40] W. Rhee and J.M. Cioffi. Increase in capacity of multi-user ofdm system using dynamic sub-channel allocation. *Proc. IEEE Vehicular Technology Conf.*, May 2000.

[41] R.W. Chang. Orthogonal frequency division multiplexing. *U. S. Patent 3 488 445, filed Nov. 14, 1966, issued, Jan. 6, 1970.*

[42] W.Y. Zou and Y. Wu. Cofdm: An overview. *IEEE Trans. Broadcast.*, 41, March 1995.

[43] P. Shelswell. The cofdm modulation system: The heart of digital audio broadcasting. *Electronic and Communication Engineering Journal*, 7:127–135, June 1995.

[44] S.B. Weinstein and P.M. Ebert. Data transmission by frequency-division multiplexing using the discrete fourier transform. *IEEE Trans. Commun. Technol.*, Com-19:628?634, October 1971.

[45] L.J. Cimini. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Trans. Commun.*, COM-33:665–675, July 1985.

[46] A. Vahlin and R.L. Frank. Use of a guard interval in ofdm on multipath channels. *IEE Elect. Lett.*, 30:2015–2016, 1994.

[47] H. Kim H. Kang and K. Hwang. Ofdm system with the guard interval and the one-tap equalizer bank for the wireless-lan. *International Conference on Consumer Electronics*, pages 20–21, 2000.

[48] J.A.C. Bingham. Multicarrier modulation for data transmission: An idea whose time has come. *IEEE Commun. Mag.*, 28:5–14, May 1990.

[49] T. Wilkinson E.A. Jones and S. Barton. Block coding scheme for reduction of peak-to-mean envelope power ratio of multicarrier transmission schemes. *IEE Elect. Lett.*, 30:2098–2099, December 1994.

[50] P. Moose. A technique for orthogonal frequency division multiplexing frequency offset correction. *IEEE Trans. Commun.*, 42:2908–2914, October 1994.

[51] P802.11y. Amendment to standard for information technology-telecommunications and information exchange between systems-local and metropolitan networks-specific requirements-part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications: 3650-3700 mhz operation in usa. Technical report, IEEE, New York, July 2005.

[52] C.L. Barrett and M.V. Marathe. Analyzing the short-term fairness of ieee 802.11 in wireless multi-hop radio networks. *In Proceedings of MASCOTS*, pages 137–144, October 2002.

[53] J. Medbo and J.-E. Berg. Simple and accurate path loss modeling at 5 ghz in indoor environments with corridors. *IEEE VTS-Fall VTC*, I:30–36, 2000.

[54] G. Bianchi. Performance analysis of the ieee 802.11 distributed coordination function. *IEEE Journal on selected Areas in Communications*, XVIII(III), March 2000.

[55] S. Choi D. Qiao and K. Shin. Goodput analysis and link adaptation for ieee 802.11a wireless lans. *IEEE Transactions on Mobile Computing*, I(IV), December 2002.

[56] H.S. Chhaya and S. Gupta. Performance modeling of asynchronous data transfer methods of ieee 802.11 mac protocol. *Wireless Networks*, 3:217–234, 1997.

[57] X. Gao T. Nandagopal, T. Kim and V. Bharghavan. Achieving mac layer fairness in wireless packet networks. *Proc. of the Mobile computing and networking conference*, pages 87–98, 2000.

[58] M. Unbehaun and M. Kamenetsky. On the deployment of picocellular wireless infrastructure. *IEEE Wireless Communications*, December 2003.

[59] R. Knopp. Achieving multiuser diversity under hard fairness constraints. *Proceedings of International Symposium on Information Theory*, page 451, 2002.

[60] G. Theraulaz E. Bonabeau and J.L. Deneubourg. Adaptive task allocation inspired by a model of division of labor in social insects. *In D. Lundh and B. Olsson, editors, Bio Computation and Emergent Computing, World Scientific*, 28:36–45, 1998.

[61] V.A. Cicirello and S.F. Smith. Improved routing wasps for distributed factory control. *In IJCAI-01 Workshop on Artificial Intelligence and Manufacturing: New AI Paradigms for Manufacturing*, August 2001.

[62] D. Morley and C. Schelberg. An analysis of a plant-specific dynamic scheduler. *In Final Report, Intelligent Dynamic Scheduling for Manufacturing Systems*, pages 115–122, 1993.

[63] V.A. Cicirello and S.F. Smith. Distributed coordination resources via wasp-like agents. *In WRAC*, January 2002.

[64] D. Gesbert and M.-S. Alouini. How much feedback is multi-user diversity really worth? *ICC*, 2004.

[65] S. Gruhl Y. Liu and E.W. Knightly. Wcfq: an opportunistic wireless scheduler with statistical fairness bounds.

[66] M. Dorigo A. Colorni and V. Maniezzo. Distributed optimization by ant colonies. *Proceedings of ECAL'91 - First European Conference on Artificial Life*, pages 134–142, 1991.

[67] M. Clerc. L'optimisation par essaim particulaire : principes, modèles et usages. *RSTI-TSI*, 21:941–964, 2002.

[68] E. Taillard L.M. Gambardella and M. Dorigo. Ant colonies for the quadratic assignment problem. *Journal of the Operational Research Society*, 50:167–176, 1999.

[69] G. Theraulaz M. Campos, E. Bonabeau and J.L. Deneubourg. Dynamic scheduling and division of labor in social insects. *In Adaptive Behavior 2000*, pages pp. 83–96, 2000.

[70] V. Cicirello and S. Smith. Wasp-like agents for distributed factory coordination. *Technical Report CMU-RI-TR-01-39, Robotics Institute, Carnegie Mellon University, Pittsburgh*, 2001.

[71] S. Nouyan. Agent-based approach to dynamic task allocation. *Proceedings of ANTS'2002*, 2463:28–39, 2002.

[72] E. Wilson. The relation between caste ratios and division of labour in the ant genus pheidole (hymenoptera : Formicidae). *Behav. Ecol. Sociobiol.*, 16:89–98, 1984.

[73] E. Bonabeau G. Theraulaz and J.L. Deneubourg. Response threshold reinforcement and division of labour in insect societies. *In Proceedings of the Royal Society of London B*, 265:327–335, 1998.

[74] D. Costa and A. Hertz. Ants can colour graphs. *Journal of the Operational Research Society*, 48:295–305, 1997.

[75] J. Bruten R. Schoonderwoerd, O. Holland and L. Rothkrantz. Ant-based load balancing in telecommunications networks. *Adaptive Behavior*, pages 169–207, 1997.

[76] O. Holland R. Schoonderwoerd and J. Bruten. Ant-like agents for load balancing in telecommunications networks. *Proceedings of Agents'97*, pages 209–216, 1997.

[77] G. Di Caro and M. Dorigo. Antnet: A mobile agents approach to adaptive routing. *Tech. Rep. IRIDIA/97-12, Université Libre de Bruxelles, Belgium*, 1997.

[78] G. Di Caro and M. Dorigo. Mobile agents for adaptive routing. *Proceedings of the 31st Hawaii International Conference on System*, pages 74–83, 1998.

[79] G. Di Caro and M. Dorigo. Antnet: Distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research (JAIR)*, 9:317–365, 1998.

[80] G. Navarro Varela and M.C. Sinclair. Ant colony optimisation for virtual-wavelength-path routing and wavelength allocation. *Proceedings of the Congress on Evolutionary Computation (CEC'99)*, July 1999.

[81] J.C. Dunat and L. Elicegui. Innovative flexible architecture for a multi-radio access technologies (rat) system level simulation tool. *Motorola S3 Symposium 2004*, July 2004.

[82] P802.11h/D2.1. Draft supplement to ieee standard 802.11. spectrum management in the 5 ghz band in europe. Technical report, IEEE, New York, September 2002.

Allocation opportuniste de spectre pour les radios cognitives

Accroissement constant du nombre d'utilisateurs, services de plus en plus évolués, sans cesse de nouveaux standards voient le jour: telles sont les grandes lignes de la situation actuelle du monde des radiocommunications sans fil. En outre, toutes les études de marché sur ce domaine prévoient une accélération de ces tendances dans les prochaines années. Cependant, le monde du sans fil doit faire face à une "pénurie de spectre", comme résultat de pratiques de régulation héritées des débuts de la radio-diffusion et perpétuées depuis, tant au niveau international que national. Cela pose un grave problème pour les futurs systèmes de communication sans fil. Des mesures effectués au niveau international révèlent une grande disparité d'utilisation entre les bandes de spectre, certaines étant inutilisées, alors que d'autres sont saturées. La "pénurie de spectre alloué" tient donc en fait à une mauvaise répartition de son utilisation. Par conséquent, une redéfinition du modèle de régulation de l'utilisation du spectre s'impose. Conscients de l'enjeu que représente le spectre radio dans l'avenir, régulateurs, opérateurs, équipementiers télécom, et économistes travaillent ensemble pour redéfinir des modèles plus adaptés. Pour accompagner cet effort, de nouveaux algorithmes d'allocation flexible du spectre sont développés, plus adaptatifs et plus intelligents. Les incroyables progrès technologiques effectués dans le domaine de la téléphonie mobile nous projettent dans un avenir où bientôt nos téléphones seront en fait des radios cognitives, capables de détecter leur environnement, d'apprendre et de s'adapter pour trouver le meilleur standard et la meilleure portion du spectre à utiliser par le service de l'utilisateur.

Dans cette thèse, nous proposons des algorithmes d'allocation flexible de spectre intelligents. Tout d'abord, nous étudions le standard IEEE 802.11a afin d'en identifier les limitations de la couche MAC. Des déploiements de plusieurs cellules et utilisateurs sont testés par simulation. Nous évaluons l'impact de la distribution des utilisateurs sur l'équité spatiale. La conclusion est que la couche MAC du IEEE 802.11a est très sensible à la distribution des utilisateurs dans le réseau, chacun ayant des performances d'accès radio conditionnées par sa position par rapport aux autres et à l'AP. Ces faiblesses peuvent être corrigées par l'utilisation d'algorithmes plus intelligents. En utilisant la théorie des files d'attentes, nous comparons plusieurs méthodes d'accès radio. Les meilleurs performances sont fournies par la méthode qui prend en compte la qualité des liens radio des utilisateurs pour adapter l'allocation de spectre. Nous proposons donc une nouvelle couche MAC pour l'"uplink" capable de contrôler au mieux, de façon distribuée et opportuniste la répartition des ressources radio entre les utilisateurs. Notre méthode propose une adaptation au monde sans fil de la méta-heuristique "d'intelligence en essaim", elle-même inspirée des insectes sociaux (termites, fourmis, guêpes). Etant donnée ses bonnes propriétés de flexibilité, adaptation face aux changements, scalabilité, notre méthode est particulièrement adaptée à des environnements dynamiques et comportant un nombre important d'utilisateurs. Nous étudions les performances de notre algorithme dans un contexte OFDMA. L'algorithme est capable de réaliser un planning autonome d'assignement des fréquences entre les utilisateurs, et d'adapter les ressources aux besoins spécifiques de chacun. Cette approche s'inscrit parfaitement dans un contexte de radios cognitives qui seraient alors capables d'adapter leur consommation de spectre au plus près des besoins des utilisateurs, diminuant ainsi le gaspillage du spectre. Les résultats obtenus avec cet algorithme démontrent le véritable potentiel de telles méthodes de collaboration entre entités distribuées pour résoudre le problème de l'allocation flexible de spectre. Le système est contrôlé de l'intérieur (depuis le bas): la complexité du système devient émergente, et résulte des multiples interactions locales simples entre les utilisateurs.