Cognitive Resource Manager Framework for Optimal Resource Allocation

Von der Fakultät für Elektrotechnik und Informationstechnik der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Ingenieurin

Marina Petrova
aus Skopje, Mazedonien

Berichter: Univ.-Prof. Dr. Petri Mähönen
Univ.-Prof. Dr.-Ing. Gerd Ascheid

Tag der mündlichen Prüfung: 8. Februar 2011

Abstract

Wireless networks are under constant pressure to provide ever higher data rates to increasing numbers of users with greater reliability. At the same time they are becoming more complex and challenging to manage. Great efforts are being done to make the wireless devices and networks adaptive and self-optimizing in order to more efficiently use the resources and deliver good quality services. High spectral efficiency, environmental adaptivity, user-awareness and energy efficiency are highly desired features in the future networks. It has also become important to support these goals at all OSI-layers in a cross-layer manner.

Making the wireless systems smarter has been a matter of research under the cognitive radio (CR) paradigm for ten years now. While CR is a very interdisciplinary and wide topic, including dynamic spectrum access and policies, flexible system architectures, learning, context awareness, cooperative networking, etc., most of the contributions so far have been limited to novel spectrum access approaches and spectrum sensing techniques. Mitola's original vision on context-sensitive smart radios was a precursor, but the current work has been still lacking precise proposal beyond high-level arguments. In this thesis we study the cognitive radios from a system point of view focusing closely on architectures, techniques and algorithms that can enable intelligent operations. We propose a modular cognitive resource manager (CRM) framework, which can facilitate a development of complex control and optimization techniques for resource management in wireless networks on diverse radio environments and problem scenarios.

This work contributes towards bringing cognitive radio a step closer to practical implementation by conducting both theoretical and experimental studies of suitable optimization methods and algorithms under the proposed CRM framework. We study in this thesis automatic and adaptive system configuration mechanisms for different resource allocation problems. As most of the problems have heavy optimization phase and often exhibit complex and non-linear parameter dependencies we have studied the use of heuristic algorithms. Genetic algorithm optimizer for PHY and MAC parameter selection has been developed and tested. For autonomous channel allocation we have studied two different classes of algorithms. An approximative coloring algorithm and a corresponding protocol were designed and successfully implemented to minimize the interference in wireless local area networks. An evolutionary game theory method based on balls and bins problem was subsequently developed to jointly address channel allocation and load balancing problems. Finally, the work in this thesis concludes by applying Minority Games to medium access control problem in order to enable self-organization without information exchange overhead.

Kurzfassung

Von drahtlosen Funknetzwerken wird fortwährend gefordert, höhere Datenraten für eine steigende Zahl von Nutzern mit größerer Zuverlässigkeit zu liefern. Dabei werden Funknetzwerke selbst immer komplexer und entwickeln sich zu einer administrativen Herausforderung. Große Anstrengungen werden unternommen, drahtlose Geräte und Netzwerke adaptiv und selbst-optimierend zu gestalten um effizienter mit Ressourcen umzugehen und Endnutzern qualitativ zufriedenstellende Dienste zu bieten. Hohe Spektral- und Energieeffizienz, Anpassungsfähigkeit an die Umgebung und die Berücksichtigung von Nutzeranforderungen sind wünschenswerte Merkmale für zukünftige Netzwerke. Zusätzlich ist es notwendig, diese Ziele grenzübergreifend auf allen OSI-Ebenen zu unterstützen.

Seit nunmehr zehn Jahren wird unter dem Paradigma des "Cognitive Radio" (CR) an Möglichkeiten zur Erhöhung der Intelligenz von drahtlosen Systemen geforscht. Obwohl CR ein hochgradig interdisziplinäres und breites Thema ist, welches unter anderem dynamische Spektrumszugriffe und -regelwerke, flexible Systemarchitekturen, Lernen, Kontextwahrnehmung und kooperative Vernetzung abdeckt, begrenzen sich die meisten Beiträge bisher auf neuartige Ansätze für den Zugriff und die sensorbasierte Erfassung des Spektrums. Mitolas ursprüngliche Vision eines kontextsensitiven intelligenten Funkgeräts war ein Wegbereiter, aber gegenwärtige Arbeiten lassen immer noch präzise Vorschläge jenseits abstrakter Diskussionen vermissen. In dieser Arbeit untersuchen wir Cognitive Radios aus der Systemsicht und konzentrieren uns auf Architekturen, spezifische Techniken und Algorithmen die intelligente Handlungen ermöglichen. Wir schlagen ein Framework für modulare Cognitive Resource Manager (CRM) vor, welches die Entwicklung komplexer Kontroll- und Optimierungstechniken für die Ressourcenverwaltung verschiedenster Drahtlosnetzwerke und Szenarien erleichtert. Diese Arbeit leistet ein Beitrag dazu, Cognitive Radios der praktischen Anwendung näherzubringen indem sowohl theoretische als auch experimentelle Studien adäquater Optimierungsmethoden und -algorithmen im Rahmen des vorgeschlagenen CRM Frameworks durchgeführt werden.

In dieser Arbeit evaluieren wir automatische und adaptive Mechanismen zur Systemkonfiguration für unterschiedliche Ressourcenzuteilungsprobleme. Da vielen der angeführten Lösungsansätzen eine aufwändige Optimierungsphase zu Grunde liegt und diese oftmals komplexe und nicht-lineare Abhängigkeiten zeigt, haben wir den Einsatz von heuristischen Algorithmen untersucht. Optimierer für PHY- und MAC-Parameter auf Basis genetischer Algorithmen wurden entwickelt und getestet. Für die autonome Zuweisung von Kanälen haben wir zwei verschiedene Klassen von Algorithmen untersucht. Ein approximierender Coloring-

Algorithmus und ein dazugehöriges Protokoll wurden entworfen und erfolgreich implementiert um die Interferenz in drahtlosen lokalen Netzwerken zu minimieren. Eine Methode der evolutionären Spieltheorie basierend auf dem Balls-and-Bins-Problem wurde im Folgenden entwickelt um kooperative Kanalallokations- und Lastverteilungsprobleme zu adressieren. Der Inhalt der Arbeit schließt mit der Anwendung von Minority Games auf Medienzugriffsprotolle um Selbstorganisation ohne den Mehraufwand des Informationsaustauschs zu ermöglichen.

Acknowledgments

While working towards this dissertation I had the opportunity to get to know and collaborate with a number of wonderful people and dedicated researchers, who in various ways helped me and supported me during this journey.

I would like to deeply thank my advisor and mentor, Prof. Petri Mähönen for granting me the privilege to join his team. Throughout the years he has been a true source of support, inspiration and enthusiasm. I am grateful for the hours of long conversations he dedicated to discuss exciting research matters as well as life experiences and lessons learned with his students. I grew to know Prof. Mähönen not only as person with great intellect and knowledge, but also as an energetic and dedicated mentor to his group.

I also wish to thank my colleague Janne Riihijärvi for the years of successful collaboration. He is a fascinating person to work with. His scientific background and ambition for ever higher achievements have impressed me over the years.

The interaction and collaborative work with a number of members of the group and visiting researcher was of great importance for shaping and development of some of the ideas. In this regard, I would like to thank Christian Sokolowski, Sebastian Albrink, Lili Wu, Natalia Olano, Guilberth Martinez Vergas, Maria Michalopoulou, Dr. Simon Fischer and Dr. Alexandre de Baynast for their support and valuable contributions.

I would also like to acknowledge the European Commission and DFG (Deutsche Forschungsgemeinschaft) for providing funding for part of the work in this thesis.

Many thanks to my boyfriend Michael for his great support, encouragement and patience, especially during the writing phase of this work.

Finally, I would like to thank my parents and my sister for their immense love and support throughout the years. A great deal of the life road I have been walking so far, could not have been possible without them being around. This dissertation is dedicated to them.

Contents

1

Introduction

This thesis contributes towards bringing cognitive radios a step closer to practical implementation by conducting extensive studies on selected optimization methods and algorithms under the proposed cognitive resource manager (CRM) framework. We investigate both theoretically and experimentally adaptive and self-optimizing system configuration mechanisms for variety of resource allocation problems encountered in wireless networks.

1.1 Motivation

Modern wireless networks are required to be robust, flexible and efficient, but yet provide high-quality and low-cost services to the users. We witness constant increase of traffic in the networks not only due to the higher number of users but also due to the emergence of new applications and services. In the attempt to find a way to handle the high demand for increased data rates and bandwidth, improving the efficiency of the wireless networks has been a highly relevant research and engineering problem for a long time now. Although a great amount of work has been already done to enable more efficient wireless networks in terms of full utilization of the available resources, e.g., deployment of multi-antenna, multi-carrier techniques and advanced power control, still better spectrum and power efficiency remain as high priority goals for the future network designs. In this context, one of the most intensively studied research questions in the last decade was and still is the issue of spectrum scarcity. Dynamic spectrum access (DSA) techniques under the umbrella of *cognitive radio* have been proposed to facilitate more efficient use of underutilized licensed spectrum, see for e.g., [1, 2] and the references therein. Great efforts have been made to come up with solutions where opportunistic secondary use of licensed spectrum are both technically and economically viable. Spectrum sensing techniques, spectrum sharing protocols and new regulations and policies have been proposed to ensure that no harmful interference is caused to the licensed users by the operation of the secondary users. One of the turning points for the dynamic spectrum access and generally for the cognitive radio community was the decision of the Federal Communication Commission (FCC) in November 2008, to allow unlicensed use of the TV frequency bands in the USA, so called TV white spaces. In order to guarantee protection of the TV transmitters and receivers the FCC mandated use of spectrum sensing as well as geolocation databases for secondary users to find out the allowable channels in a certain location. In the latest release of the rules for unlicensed use of the TV bands, however, FCC removed the mandatory sensing requirements thus making the process of free channels acquisition easier [3, 4].

Although the main drive pushing the cognitive radio research forward has been dynamic spectrum access, cognitive radio paradigm goes far beyond spectrum management. Going back to the Mitola's original idea, cognitive radio is a system, which is self-aware, can observe and plan according to the stimuli from the radio environment, learn from past actions and act accordingly [5–7]. As such cognitive radio has been used as a framework to tackle general resource management problems in today's wireless networks and further develop the idea of self-organizing and self-optimizing networks.

There are several well-known, but still open research challenges where the principles of cognitive radio can help finding out solutions. One of these challenges is *complexity*. Along with the expansion of the wireless network services and applications, wireless networks have been gaining in complexity. Managing today's heterogeneous networks is a difficult task and requires specialized tools and complex human interventions for maintenance, (re-)configuration and optimization. Additionally, radio resource management blocks in most of the cases, are of monolithic structure, so that integration of any new optimization algorithms in plug and play fashion is difficult and expensive. There are also justified doubts that the current control and resource management mechanisms could provide an efficient support for cross-layer optimization and fast adaptation. Hence, flexible, modular and highly scalable architectures are needed to build cognitive radios and networks.

The second challenge is achieving *easy and automated optimal self-configuration* of the devices/networks. The current technological trend towards software defined radio platforms gives a lot of freedom and flexibility for composing optimized PHY/MAC configurations of devices that fit to the specific channel and network conditions. Nevertheless, dealing with a large set of parameters and the interdependencies among them, which in some cases are non-linear or even unobservable, causes difficulty in finding the optimal settings for a specific user at run-time. Therefore, introducing learning into the network control and management, where the network could learn about its behaviour over time, and configure itself accordingly is likely to improve the network performance [8]. Finally, understanding better the functional and parameter dependencies the wireless systems and facilitating meaningful optimizations still remain one of the greatest research challenges: Modern devices and networks host multiple radio technologies and run multiple user applications concurrently. Different applications operate at maximum utility under different protocol configurations. Due to the inherent coupling between the protocol layers, adjusting the physical layer settings, for example, will have implications on the optimal routing, the performance of the transmission control protocol (TCP) and the overall achievable network utility. Although cross-layer issues in wireless networks have been subject to thorough studies, there has been limited progress in understanding the inter-protocol and inter-parameter dependencies and their influence on the optimal network resource allocations for different applications. Thus, layer abstractions and developing a framework that facilitates cross-layer collaboration are still needed.

Cognitive radio is an inherently interdisciplinary research field, and a plethora of approaches have been proposed in the literature on how to address, in "cognitive" fashion, the architectural and optimization challenges in the today's and future wireless networks. Considerable efforts have been invested in a design of a cognitive resource management or cognitive engine frameworks in attempt to identify the main building blocks (e.g. interfaces, schedulers,and policy engines), and the necessary algorithms and protocols, which could enable smart resource allocation and system optimization through a cognitive behaviour [9–13]. Various machine learning techniques have been investigated for design of adaptation mechanisms of cognitive radio, especially for adjusting radio parameters such as modulation order, transmit power, number of sub-carriers, carrier frequency etc. Game theoretical approaches have been also considered for modelling of resource sharing problems among cognitive radios in cooperative manner. Although some incremental steps have been made towards realization of functional cognitive radios, there are many issues such as cross-layer optimization and learning need to be further investigated to understand the real benefits. Real integrated implementations, demonstrating cognitive processes and cognitive architectures supporting theses processes are still missing. This thesis is a contribution towards adaptive resource management techniques in cognitive radios. The work in the thesis takes a systematic approach towards CR and introduces a cognitive resource manager framework (CRM) that can serve for practical realization of future cognitive radio systems. In contrast to most of the related work, both theoretical and experimental studies of optimization techniques and algorithms have been performed and solutions for problems such as channel allocation, auto-configuration of radio parameters through cross-layer optimization, and self-organization in medium access control have been presented.

1.2 Contributions of the thesis

The main contributions of this thesis can be summarized as follows.

- **Cognitive Resource Manager.** We have proposed a new architectural framework, called cognitive resource manager (CRM) that can facilitate better practical implementation of cognitive radios than the earlier proposals. The framework tries to map Mitola's abstract cognitive cycle to a software architecture and suggests necessary building blocks for realization of cognitive radio functionalities. The CRM is component based and addresses the lack of modularity and extensibility of the classical radio resource manager (RRM) and the earlier proposed cognitive engines (CE). Unique modules such as a *toolbox* of optimization and modelling techniques, *generic interfaces* for reach information exchange in the protocol stack and policy engine are part of the designed framework. A prototype implementation of CRM has been developed in the European project ARAGORN, and technology transfer projects with industrial manufacturers are on-going .

- **Study of multi-objective and PHY/MAC cross-layer optimization for cognitive radios using genetic algorithms.** We adopt an evolutionary approach for solving multi-objective optimization problems by thorough analysis of genetic algorithms (GAs). We investigate the applicability of GA as an optimization tool for finding optimal radio parameters so that a required combination of throughput, bit error rate (BER) and transmit power is achieved. As part of the experimental analysis, we developed a single-carrier dynamic channel allocation scheme using GA and tested it on a GNU Radio platform [14,15]. We furthermore designed a genetic algorithm to optimally set PHY and MAC layer parameters in OFDM multi-carrier system in a cross-layer fashion with minimal feedback information.

- **Theoretical and experimental study of frequency assignment problem by means of graph colouring and game theory.** One of the main contributions of this thesis is a study, both theoretical and experimental, on dynamic and automatic channel allocation in wireless networks using graph colouring and congestion game theory approaches. Two algorithms were proposed, namely a heuristic channel allocation based on *DSATUR* colouring algorithm and distributed load balancing channel allocation based on so called *balls and bins* game. The performance of DSATUR algorithm has been verified in the context of IEEE 802.11 (Wi-Fi) and SDR testbed. The balls and bins solution has been also implemented on USRP testbed in order to validate the performance expectations. A slightly modified version of coloring algorithm as been already adopted by a manufacturer into its products.

- **A novel self-organizing resource allocation scheme with minimal feedback using minority game.** We have applied minority game (MG) to resource sharing in wireless network to show that it can be achieved in a self-organizing fashion without exhaustive information exchange or coordination between the users. We have designed a MG resource allocation scheme to schedule user transmissions and combined this model with collision avoidance multiple access (CSMA) protocol in order to minimize the number of collisions in dense networks. Through our analysis we showed that different dynamical cooperation models could be applied to cognitive radios where strict cooperation is not necessary and information exchange is minimal.

Most of the contributions of these thesis have been already published in reviewed conferences and journals. The reader is referred to [10,16–31] for publications related to the work in this thesis.

1.3 Thesis outline

The thesis is organized in seven chapters. In Chapter 1 we gave a motivation for the work and summarized the main contributions of the thesis. Chapter 2 introduces software defined radio (SDR) and cognitive radio (CR) paradigms. Furthermore it discusses the architecture and the operational principles of cognitive radio introduced by Mitola. The chapter also introduces shortly the concept of dynamic spectrum access for more efficient spectrum sharing and opportunistic spectrum usage.

Chapters 3,4,5 and 6 contain the main work done in this thesis. In Chapter 3 we propose a cognitive resource manager (CRM) framework and describe in detail its building blocks and functionalities. The framework can facilitate development of control and optimization techniques for resource management and sharing in wireless networks using the principles of cognitive radio. Chapter 4 studies genetic algorithms as tools for solving multi-objective optimization problems for cognitive radio. The particular problem addressed is finding out optimal transmission radio parameters and MAC parameters in a cross-layer fashion so that maximal throughput and reliability are achieved in a changing radio environment. Chapter 5 proposes and studies two different channel allocation approaches. The first scheme is based on a graph coloring. The second one is a distributed load balancing channel allocation scheme that uses congestion game called balls and bins. Both simulation and experimental analysis of the proposed algorithms are provided. Chapter 6 presents a novel self-organizing resource allocation scheme with minimal feedback information based on minority game (MG). First a comprehensive background in minority games is given followed by description and simulation analysis of self-organizing user medium access scheme. Finally, Chapter 7 concludes the thesis and makes recommendations for the future work.

2

From Software Defined Radio to Cognitive Radio

The cognitive radio paradigm was originally introduced by Mitola and Maquire in [5]. The main idea behind it was to design a fully reconfigurable radio that could automatically set its transmission or reception parameters in accordance to the current network state and user requirements. Cognitive radio (CR) was considered as a natural follower of the software defined radio (SDR), which would be able to learn, decide and act based on the environmental perceptions and context.

In this chapter we give an overview of the cognitive radio technology and introduce the cognitive network paradigm. At the very beginning we discuss the basic concept of software defined radio and its hardware realizations available today. Then, we give an insight of the cognitive radio itself: the architecture and the operational principles introduced by Mitola. Furthermore, we address the concepts of spectrum agile radio and dynamic spectrum access, which are opening new possibilities for more efficient spectrum sharing and opportunistic spectrum usage. Finally we argue that the cognitive radio is a promising platform that could enable optimization both on the terminal and network level and could lead to a cooperative federation of CRs into a Cognitive Wireless Network (CWN).

2.1 Software defined radio

2.1.1 The basic concept

The concept of "software radio" (SR) ,also known as "software defined radio" (SDR), has been around for some time, having initially been discussed in context of military communications. The term *software radio* was introduced by Mitola in 1991 [32] referring to a radio, which can be easily reconfigured and reprogrammed. The flexibility to change the radio configuration allows for the same device to perform a variety of different functions at different times.

In practice the term "software radio" refers to a software implementation of the radio functionalities, which enables the user terminal to dynamically adapt to the radio environment in time. In comparison to the traditional radio designs where the radio architecture is optimized for specific frequencies and applications and is tightly coupled with the hardware, the software radio substitutes software for hardware processing and facilitates a transition from dedicated to general-

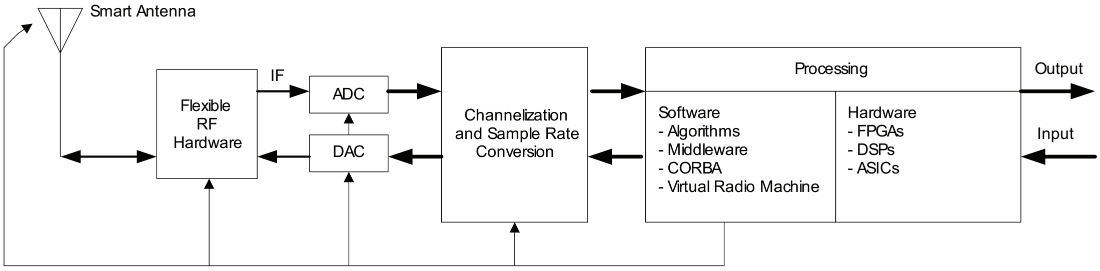

Figure 2.1: Model of a Software Radio (adapted from [33]).

purpose hardware. This means that the radio interface functionalities of the transmitter (Tx) and the receiver (Rx), so far implemented by dedicated hardware, could be implemented in software. Dedicated hardware would be replaced with digital signal processors (DSPs), which will execute the necessary software.

The most obvious benefit from the software radio is that instead of having to build extra circuitry to handle different types of radio signals, one can just load an appropriate software modules. At one instant a device could be an AM radio, at the next a wireless data transceiver and then perhaps a high definition TV (HDTV) set. This flexibility of software could be leveraged to do things that are difficult, if not impossible, with traditional radio setups.

2.1.2 Architecture principles

So far, there is no sharp definition that precisely determines the level of reconfigurability the radio has to exhibit in order to become a software radio. Nevertheless, the main characteristic of the software radio is that a vast number of waveforms and different modulations, error correction codes, encryption processes, etc., are implemented in software. This enables high degree of reconfigurability so that the same hardware platform can be used for different applications.

In Figure 2.1 a basic model of a software radio is shown. Every SDR transceiver should be preferably equipped with smart antenna. In general smart antennas provide improved quality of the wireless communication by mitigating the fading through diversity and beamforming and minimizing the interference through spatial filtering.

One of the key design goals for the software radio is to be able to digitize the analog signal as early as possible in the receiving path and convert it in the analog domain as late as possible in the transmitting path. In most of today's implementations the analogue/digital (A/D) conversion is done in the intermediate frequency (IF) band. There are two main advantages of doing digitalization in the IF bands. Firstly, current A/D converters can achieve enough speed and resolution at IF frequencies. Secondly, this design requires less computational resources.

The ultimate goal for the software radio designers is to move the A/D conversion just after the antenna in the receiver chain. This would open the possibility to perform all the signal processing in software so that the same piece of equipment could be used for any new frequency, standard or application with just simple software upgrades. However, due to several technical limitations of the currently available converters, the A/D conversion at the RF is not a reality yet.

Present A/D converters are limited in speed and resolution at high frequencies. Moreover, when A/D converters are placed right after the antenna, sampling is done over signals with very different strengths: the dynamic range of the signals may vary from μvolts to several volts. Current A/D resolutions are not able to cover such dynamic ranges. Nevertheless, significant research efforts are taking place to overcome these problems.

Speed and power consumption are also a trade-off in A/D converter designs. Fast A/D converters exhibit higher power consumption than slower ones. If the power consumption is very high, the A/D converter could dissipate too much heat and overheat the device. This issue is particularly critical in mobile devices, where cooling systems cannot be installed and the battery life is an extremely limiting factor.

In order to make the radio digitalization possible the future converters would need wide signal bandwidths, high sampling rates and high dynamic range; an operating bandwidth of several GHz to enable conversion of a signal over a varying frequency bands; a large spurious-free dynamic range to provide recovery of small-scale signals in high-interference environments; and last but not least they have to fulfil the previously mentioned criteria at a reasonable power consumption and cost.

As presented in the Figure 2.1 after the A/D conversion is performed, channelization and sample rate conversion is carried out. All the signal processing that follows later is done in software on a high speed digital signal processors (DSPs) or general purpose processors (GPPs), alternatively field programmable gate arrays (FPGAs) or application specific integrated circuits (ASICs) can be used. Each of these specific hardware designs offers different levels of flexibility, modularity and scalability required in the software radio. The different approaches have also highly different power consumption characteristics. While a DSP is a microprocessor-based structure and supports programming of variety of functions, the ASIC implements the system circuitry in silicon and can be used only in applications it has been designed for. The ASIC is the most specialized hardware, optimized in terms of power consumption and speed. An FPGA offers a lot more flexibility than an ASIC but less than a DSP. General purpose processor are as fast or faster than DSPs but power consumption may be significantly higher. They are especially suited for handling non-signal processing tasks such as packet level protocol stack processes. One big advantage is that GPPs are highly versatile and can implement a large number of applications.

When the system performance demands exceed the hardware capacity, a combination of the above mentioned platforms might be used. Very popular approaches are e.g., to use DSPs as cores inside ASICs to provide some flexibility to the ASIC or using FPGA to enhance the performance of a DSP and so on.

The choice which hardware platform to use for the software radio heavily depends on the requirements for speed, flexibility, power consumption, cost and many others. Numerous platforms for digital signal processing exist today and new alternatives will come available in the future but the fundamental trade-offs can not be avoided. However, one should bear in mind that the price for flexibility and the use of common purpose electronics and computational components is the increased complexity and power consumption.

In order to reach their full potential, it is necessary that the software radios are seamlessly integrated in the existing heterogeneous network structures. This is quite a challenging task taking into account the capability of the software radio to reconfigure and change its functionality in real-time. In this context the software radios should not only be able to cope with different radio environments but also deal with different application suites across different networks as the user migrates. All these requirements pose strong demands to the software architecture of the software radio. In the ongoing research and implementation the use of object brokers has been considered to be an appropriate solution to integrate the software radios into a network. In particular CORBA (Common Object Request Broker Architecture) has been one of the solutions often adopted [34].

2.1.3 The software radio evolution

Since the proliferation of the software radio paradigm there has been a substantial amount of work towards implementation and prototyping of this emerging technology. It is very clear that the flexibility the software radio offers to the communication system poses great implementation challenges. For example designing a RF front-end which can handle a broad range of carrier frequencies and modulation types is an extremely difficult task. In fact achieving a dynamic range which is a measure of the highest- and lowest-level signal that can be detected by the radio is a key issue in the software radio design [33].

The precursors of the SDR date more than ten years before the term was coined by Mitola. Developing reconfigurable radio architectures was stimulated by the necessity of the U.S military to have *multi-band multi-mode* radio (MBMMR), which can operate across different frequency bands using variety of protocols. One of the first such radios built was the US Air Force's Integrated Communications Navigation and Identification Avionics (ICNIA) system, developed in the late 1970's [35]. The system used a DSP-based modem to operate multi-function multi-band airborne radios in the 30 MHz to 1.6 GHz band. ICNIA's technology has been later the foundation for many other military radios. In late 1980s, GEC-Marconi Electronic Systems Corporation developed the first Programmable Digital Radio (PDR) described in the literature [36].

Obviously, one of the best known projects to build a software defined radio for military applications is SPEAKeasy. It was a joint Department of Defence and industry program initiated to develop a software programmable radio operating in the frequency range from 2 MHz to 2 GHz. The driving vision in the project was to develop an open and modular platform which could employ number of waveforms selected from memory, or downloaded from disk, and could be reprogrammed over the air [37,38]. The system built during two phases of the project served as a solid base in the following software radio programs such as the Joint Tactical Radio System (JTRS) [39], trying to solve the problem of integration of multiple systems into a single network.

Apart of the military projects in the USA, several university groups and companies have designed and built their own experimental SDR platforms. We will name here just a few examples: Spectrum Signal Processing Inc. is a company that is specialized in SDRs. They provide a wide range of SDR systems and baseband processing boards that utilize a combination of PowerPC, DSP and FPGA signal processing devices [40]. Motorola has been building their 4G base station using SDR technology and has been one of the first manufacturers deploying SDR in its products since 2001 [41]. Vanu is other significant player on the market that produces software for SDRs. They have designed their own software radio architecture which enables base stations to simultaneously operate as GSM, CDMA, Integrated Digital Enhanced Network (iDEN), WCDMA, and beyond.

One of the most popular and widely used SDR platforms in the research community is the Universal Software Radio Peripheral (USRP board family) designed by Ettus Research [15][1]. It is an integrated board, which incorporates AD/DA converters, a number of RF front-ends covering different frequency bands, and an FPGA which does a part of computationally most expensive pre-processing of the input signal (see Figure 2.3). The low-cost and relatively high speed made the USRP board the best choice for a GNU Radio user to implement some real time applications. It is also very often known simply as a GNU radio platform as it is mainly used in combination with the open-source GNU radio software [14]. The GNU radio software itself is built in a modular fashion. It provides a library of signal processing blocks and the framework to tie these blocks together. The programming environment is a combination of C++ and Python The project is open to anybody who is willing to contribute in the further development.

Many research groups have been also actively working on building their own software defined radio platforms. The KUAR platform has been designed to be a low-cost experimental platform targeted at the frequency range 5.25 to 5.85 GHz with a tunable bandwidth of 30 MHz [42]. The platform includes an embedded 1.4 GHz general purpose processor, Xilinx Virtex2 FPGA and supports gigabit Ethernet and PCI-express connections back to a host computer. This allows for all, or almost all processing, to be implemented on the platform, minimizing the host-interface communications requirements. The platform was designed to be battery powered thus allowing for untethered operation. KUAR utilizes a modified form of the GNU Radio software framework to complete the hardware platform.

[1]Etus Research LLC has been recently acquired by National Instruments Corporation.

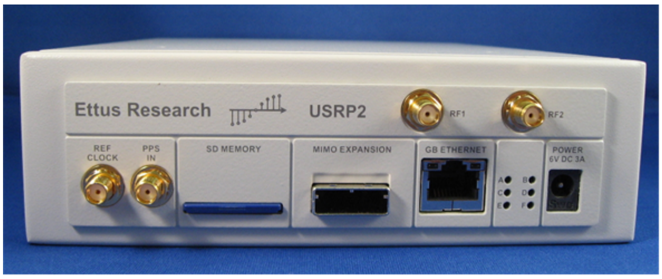

Figure 2.2: USRP2 SDR platform.

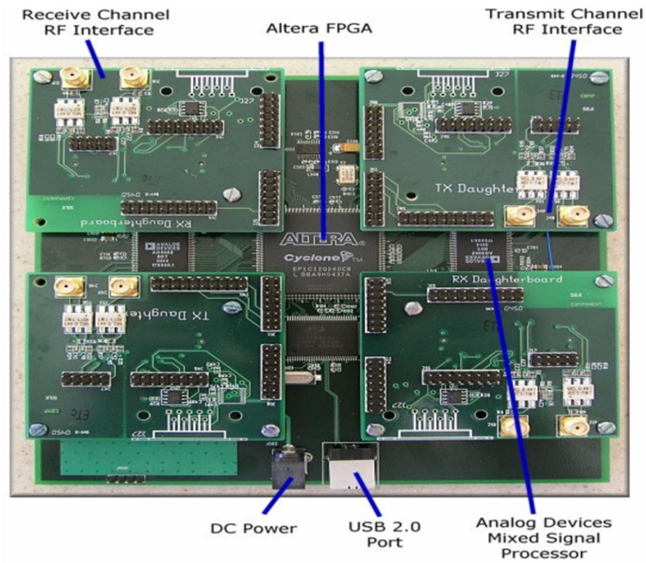

Figure 2.3: USRP motherboard with four daughterboards.

At present Wireless Open-Access Research Platform (WARP) developed at Rice University is one of the most used reconfigurable research platform to prototype high-speed wireless communication algorithms and systems (see Figure 2.4). Due to its programmability and flexibility it allows for easy implementation of various physical and network layer protocols. The latest design is built around a Xilinx Virtex-4 FPGA as a primary processor on the main board. Additionally there is a PowerPC processor embedded in the FPGA, which provides an embedded programming environment for MAC and network layer designs. Maximum four radio daughterboards can be attached to the main board so that up to 4×4 multiple-input multiple-output (MIMO) system can be supported. Furthermore, WARP is a dual-band platform and can operate in 2400-2500 MHz and 4900-5875 MHz. More details about the design and applications of the WARP board are available in [43].

Similar but more powerful and expensive reconfigurable platform called Berkeley Emulation Engine (BEE2) has been developed at Berkeley Wireless Research Center [44]. The engine consists of five Xilinx Vertex-II VP70 FPGAs connected

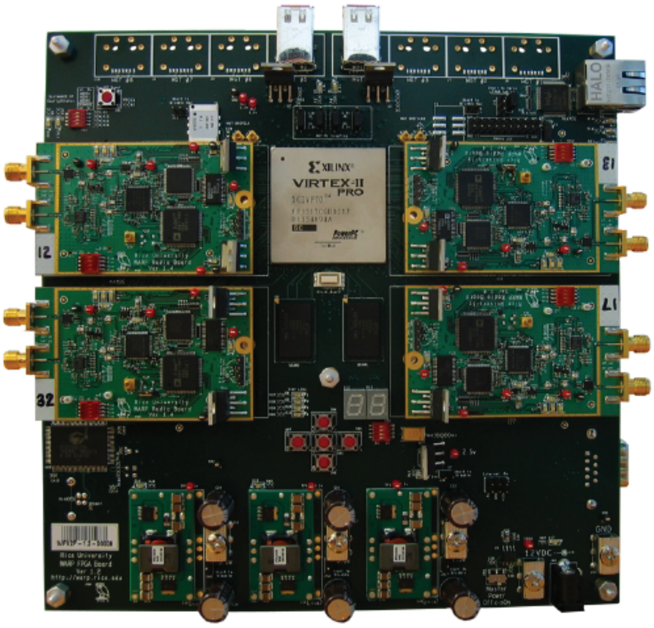

Figure 2.4: WARP FPGA and radio boards (adopted from [43]).

with high speed internal links, which provide a possibility to execute in parallel intensive signal processing algorithms. One of the five FPGAs called a control FPGA executes Linux OS on its embedded PowePC to control the peripherals. To interface the BEE2 with radios a total of 8-10 Gbps full-duplex links are available per platform. Due to its modular design and scalability BEE2 is applicable to a wide range of high-performance applications such as real-time radio telescope signal processing, cognitive radios, computer architecture emulation and so on. As previously mentioned due to its high cost it has been mainly used for different projects in-house and it has not reached to popularity of the USRP and WARP boards.

Out of these platforms we have mainly used USRP2s for the research done in this thesis. As the USRP platform is extremely well presented in the literature, we omit detailed description in order to keep the thesis compact. It should be noted that the WARP platform became available only quite late during the thesis work, and thus USRP was our main experimental platform.

2.2 Cognitive radio

2.2.1 *View points and definitions*

In 1999 Mitola and Maquire took the definition of the SDR one step further by introducing the term cognitive radio (CR) [5, 6]. The vision behind the CR was to make the future radios self-aware, user-aware and RF-aware by being able to learn from their surrounding. This multiple awareness means that the radio is able to reason about a set of RF bands, air interfaces, protocols and spatial and temporal patterns of spectrum usage called *radio etiquette* in order to satisfy the user requirements and achieve better user experience. For expressing the knowledge the cognitive radio has about itself and the environment around it, and organizing this knowledge, an axiomatic language framework called Radio Knowledge Representation Language (RKRL) has been designed by Mitola. The framework also provides a reference ontology of radio protocols, air interfaces, networks, and user communications states and it is believed to facilitate a rapid prototyping of the cognitive radio. All the specifics and the design principles of the RKRL are given in Mitola's Ph.D. thesis [6].

This early and rather futuristic view of cognitive radio by Mitola motivated the research community to start introducing more flexibility into the radio and apply learning mechanisms inspired from the artificial intelligence (AI) community in order to make radios adjustable to the changing wireless environment. Very soon, many definitions of cognitive radio were introduced as a result of the variety of problems that were tackled under the name of cognitive radio and the individual understanding what the cognitive radio in reality might be. In 2005, Simon Haykin presented an excellent follow-up discussion on the fundamental tasks of cognitive radio [7]. For most of the community working in this exciting research area the radio able to do Dynamic Spectrum Access (DSA) is considered to be cognitive. This definition is somewhat narrow, and it specifically takes into account only one aspect of the Mitola's original definition, namely RF- awareness and agility. However, in the literature very often radio systems that can do context based routing, automated channel allocation, adaptive modulation or in general can exhibit any kind of self-configuration have been also defined as cognitive radios, by more system oriented researchers. In order to avoid the confusion and bring some cohesion in the current state-of-the-art we distinguish between two types of cognitive radios, namely CR-I which is specifically DSA capable and CR-II which in general follows the idea of having a fully configurable radio system which can learn and decide based on the environmental stimuli and *intelligently* optimize its performance. In the following sections we are going to discuss both CR types and give an overview of state-of-the-art. In the rest of the thesis we are going to refer to both types with the term cognitive radio interchangeably.

2.2.2 The cognitive cycle

The idea of cognitive cycle was first described by Mitola in [5]. It reflects the overall cognitive process in cognitive radios as illustrated in Figure 2.5. The cognitive process starts with *observing* the outside world. Generally this could mean sensing the environment in many ways, namely sensing the spectrum in order to find spectrum opportunities, finding out n-hop neighbors, getting channel state and network state information, location information, etc. The observation process includes measuring environmental and system variables, reading the current radio and protocol level settings as well as getting information from other

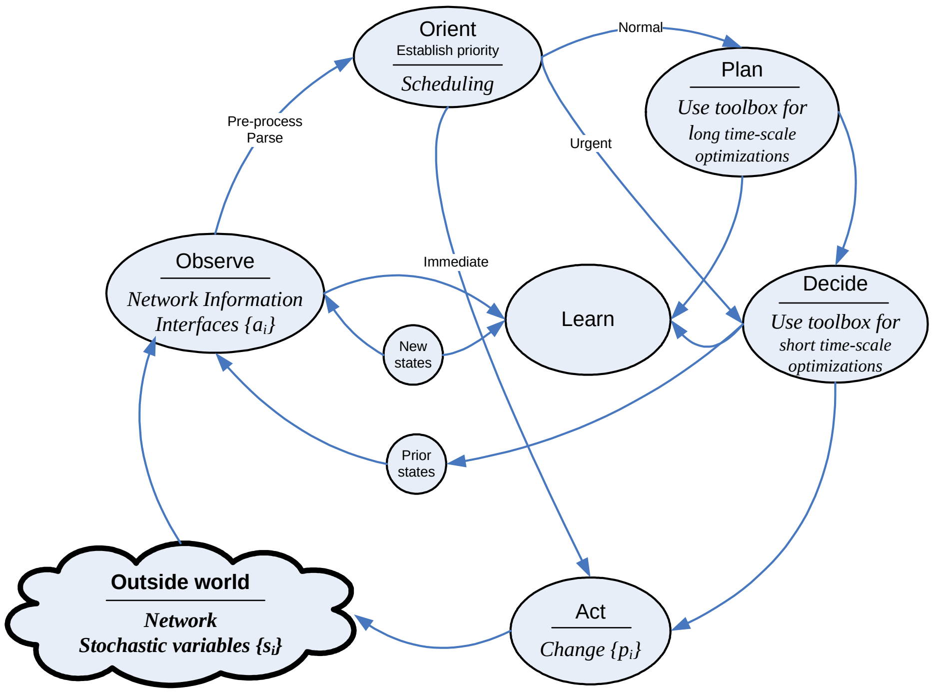

Figure 2.5: Cognitive cycle.

sources (cooperative users, databases, etc.) via a control channel. The knowledge gathered in the observation state can be used both in the process of *learning* and *orientation*. During the orientation the cognitive radio schedules its actions and sets priorities. If there is need for an *immediate action*, the CR will do so. For example if the radio "knows" that it will be in a tunnel for the next 10 minutes and the battery of the hand-held device is not going to last for a long time, it can act to shut down the wireless network interfaces or even the device itself to prevent it from constant trying to get a connection and drain the battery power. There can be also situations where the CR should urgently go from orientation into a *decision state*. Alternatively, for actions that are not time-critical the radio could "afford" careful *planning* and evaluation before making a decision and acting. *Learning* takes a central place in the cognitive cycle. It is a process of extracting the useful information from the observations, passed decisions and actions and also being able to recognize the changes in the environment and adopt accordingly. Both supervised and unsupervised learning can be used depending on the application [45]. *Acting* is the culmination in the cognitive cycle. It results in optimal resource allocation and parameter settings such that the performance of the radio device or the network is optimized under certain conditions and preferences.

The cognitive cycle as defined above is very general and broad, and encompasses different degrees of cognition. Introducing a cognition in the radio at such scale is a very challenging problem. Nevertheless, at a beginning, cognition could be implemented in a a small set of tasks. For example, the cognitive cycle or part of it could be build around functionalities such as: spectrum management, channel allocation, power control and so on.

2.2.3 Dynamic spectrum access

Dynamic spectrum access emerged as a potential killer application for cognitive radios and it is still one of the major forces driving the CR research forward. The basic idea is to realize opportunistic use of the under-utilized frequency bands in a dynamic fashion, by exploiting the spectrum agility and other cognitive features of the radio [1, 2].

The need to relax the traditional spectrum licensing has been motivated by the increased demand on radio frequencies for new high-speed wireless services over the years. There is a strong belief that the present spectrum licensing is not efficient and flexible enough [46–48]. Having most of the spectrum allocated to dedicated users (operators, radio and TV broadcasters, military, etc.), it is becoming hard to find free bands for deployment of new communication services. Moreover the exclusive right to use certain blocks of the spectrum the license-holders have, eliminates the possibility to reuse the spectrum or parts of it when they are inactive. Such a spectrum scarcity exists especially in the frequency bands below 3 GHz, which are actually bands with the most commercially appealing propagation characteristics [49].

Number of university groups, industry labs and regulators performed measurement campaigns to explore the utilization of the spectrum in practice [50–61]. The reader is specifically referred to a recent thesis by Wellens and references therein to gain the state of the state-of-the-art view on spectrum measurements [62]. A common finding of these extensive spectrum measurements was that a significant amount of licensed spectrum was unused. Spectrum underutilization claims stimulated intensive activities in finding out better technical and regulatory solutions for spectrum management. It is almost a decade now since the first ideas on radical spectrum reforms have been proposed. During this time a large number of dynamic spectrum access approaches have been suggested and discussed. These fall under three main access models as shown in Figure 2.6.

Dynamic exclusive use refers to a more flexible licensing policy. Spectrum bands remain licensed for a exclusive use but spectrum trading and dynamic spectrum allocation are additionally foreseen. In the first case the licensees have the right to sell or lease unused spectrum. The second approach aims at assigning exclusive right to use the spectrum in a dynamic fashion taking into account the spatial and temporal traffic statistics of the services. The economical modeling of spectrum sharing has been extensively studied in the last couple of years. For studies on spectrum trading, auctioning and pricing the reader should refer to [63–69].

Open sharing model also known as *spectrum commons model* allows for public use of spectrum with equal rights towards anyone. The ISM (Industrial, Scientific

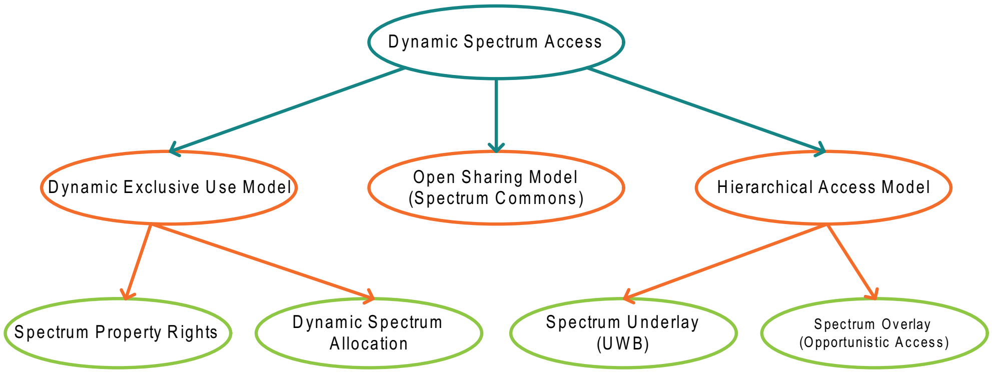

Figure 2.6: Dynamic spectrum access models: adapted from [2].

and Medical) band is a well know example of this spectrum model. Several wireless standards including Wi-Fi, Bluetooth and ZigBee operate successfully in this band. As a result of the positive experience with the ISM band, the proponents of this model have been proposing opening more spectrum to the "common" pool.

In the *hierarchical access model* two types of spectrum users are involved, namely primary and secondary users. This model opens the licensed spectrum to unlicensed secondary users while putting restriction on the maximal perceived interference by the primary users. Under this model two different approaches have been proposed, namely spectrum underlay and spectrum overlay. Spectrum underlay refers to a spectrum management principle by which signals with a very low spectral power density can coexist with the primary users of the frequency bands. For example, ultra wide band (UWB) and other spread spectrum technologies can use a spectrum underlay. Assuming that the primary users transmit constantly, with a power density many times higher than the underlay users, detection and exploitation of white spaces are not needed in this approach. The overlay access, also known as *opportunistic spectrum access* takes advantage of the silent periods of the primaries in time and space. This model does not necessarily limit the transmission power of the secondary users but puts restriction on where and when they may transmit. However, still one of the biggest challenges for the secondary users is to reliably detect and exploit a spectrum opportunity without harming the primary system operation.

The transition of DSA from concept to reality has made a big step forward in the last several years. Not only the research and system prototyping have been progressing but also standardization activities have been increasing in volume. One of the main drivers of making dynamic spectrum access commercially viable has been the Federal Communication Commission (FCC) approval of unlicensed use of television white spaces in 2008 [3, 4]. The regulatory developments in the TV bands have been closely related to the IEEE 802.22 Working Group on Wireless Regional Area Networks (WRANs) [70], which is the first IEEE standard on cognitive radio to exploit the underutilized TV channels in the 700 MHz band. The primer application of the proposed standard is to provide broadband services in rural areas, where cable broadband access is not available. A tremendous amount of work has been done in the working group to define the main components of the WRAN system including physical layer, medium, access control, spectrum sensing and detection techniques, and others. The rules adopted in FCC allow various wireless devices to operate in the TV bands. The strict requirement is, however, to provide a sufficient protection to the primary users in terms of interference levels, which can be done by accessing a database with registered primary users or sensing the environment.

Cognitive radio has been also a focus of several other standardization bodies. One of the most well-known ongoing activities is IEEE Standards Coordination Committee 41 (SCC 41). SCC 41 was born out of the IEEE P1900 Standards Committee and its main objective is to support a development of standards on new technologies for next generation radios and spectrum management. Since early 2008 also the European Telecommunication Standards Institute (ETSI) tries to standardize results from European research in software defined radio and cognitive radio. Finally the Wireless Innovation Forum (former SDR Forum) is a venue were mostly industry representatives have the opportunity to network and cooperate in the area of SDR, CR and DSA technologies.

As indicated in the introduction, the main focus of this thesis is on optimization and self-organization capabilities of cognitive radios. Although spectrum allocation will be considered as one of our optimization problems from the secondary users point of view, specific DSA problems in the context of primary-secondary spectrum sharing is out of the scope of this thesis. In that sense this thesis is closer to Mitola's original work and definition of cognitive radio. Due to this we limit our discussion on dynamic spectrum access to a level of comprehensive overview for sake of completeness of the thesis.

2.3 Towards cognitive wireless networks

If radios become cognitive, connecting those to form a network will require an "intelligent" coordination and resource management, and obviously new network architectures. Even without having fully operational cognitive radios, the networks have become more complex and heterogeneous. The current networking technologies are quite limiting and do not allow for easy adaptation of the protocol stacks, behaviors of the networks, policies, cross-layer interactions, etc. This inflexibility eventually leads to suboptimal end-to-end network performance. In order to handle the increasing complexity of the data networks several research groups have introduced "cognition" into networks as a possible avenue to go. In that regard several definitions have emerged in the literature on what a cognitive network (CN) or cognitive wireless network (CWN) could be.

Following the model of cognitive radio, a cognitive network is a network composed of elements that, through learning and reasoning, dynamically adapt to varying network conditions in order to optimize end-to-end performance. In a cognitive network, decisions are made to meet the requirements of the network as a whole, rather than the individual network components. In his thesis, Mitola briefly discusses how the CRs could interact on a system level forming a cognitive network [6]. Several other authors have brought up the idea of cognitive networks in their work. For example, in his seminal paper Clark *et al.* proposed an approach to future Internet design based on tools from AI and cognitive systems and introduced a *knowledge plane* within the network that could build and maintain high-level models of what the network should do to provide optimal perforce and services [8]. Furthermore, Mähönen [71] disused cognitive networks with respect to next generation wireless systems, pointing out various challenges that need to be addressed in order to achieve context-sensitive and auto-configurable networks. A rather visionary but detailed view on which capabilities the cognitive network should have and how it should function is given in [72]. The authors define cognitive networks as "*future networking technology that will allow data networks operating in complex, heterogeneous, noisy and dynamic environments to reclaim stability by learning and adapting their behaviors to meet top-level end-to-end goals*". In their later paper the same authors advanced the idea of cognitive network by introducing a software framework for cognitive networks [73].

The above work takes a wide scope approach in defining cognitive networks. There are also more limited views on the level of sophistication the CN could exhibit. Especially for the DSA community a cognitive network is a network of CRs with a capability to perform cooperative opportunistic spectrum sharing. After almost a decade from the first idea papers on cognitive networks the research is without doubt advancing and different components enablers have been already prototyped and tested. Typical examples are designs of generic interfaces, radio environmental maps (REMs), different software frameworks for testing radio reconfigurability and cross-layer optimization, etc. A great part of the work in this thesis is closely related to design issues of cognitive radio networks. We will in detail discuss our cognitive resource manager framework (CRM) and several optimization approaches that can be useful in the future wireless networks.

2.4 Summary

Cognitive radio and cognitive network paradigms are definitely driving the design of the forthcoming generation of communication systems. In this chapter we introduced the concepts of software define radio, cognitive radio and cognitive networks. In order to give a complete picture of the field we have discussed both the latest developments in the area of dynamic spectrum access and the recent efforts towards embedding intelligence and context-awareness on a system level as enablers for cognitive networks. The rest of the thesis will discuss our own contribution toward making cognitive system a reality.

3

Cognitive Resource Management

As discussed earlier, cognitive radios (CR) and cognitive networks (CN) have been proposed as key technologies to address intelligence, self-configuration and efficient resource allocation in the next generation wireless networks. The fact that a large number of blocks from the radio protocol stack can be implemented in software, has recently allowed many radio configuration parameters to become easily available for adaptive fine-tuning. This flexibility, however, has turned the problem of finding optimal radio settings into a complex combinatorial game. Metaheuristic algorithms inspired from artificial intelligence and machine learning have been proposed for automated radio configuration and optimization. In order to be able to introduce cognition (i.e context-awareness and learning) and optimal self-management of resources in the future communications systems not only algorithmic solutions are needed but also new software architectures, which will support intelligent processes and modularity. The classical monolithic radio resource management (RRM) approaches that we know, e.g., from the cellular systems, cannot answer the requirements of the future cognitive systems. Thus flexible, extensible and easily implementable designs are desired. In this chapter we introduce such a cognitive system framework, which we call CRM (Cognitive Resource Manager). The CRM design goal has been to enable efficient implementation of the cognitive cycle and introduce more adaptive and flexible use of the network stack. In this chapter we describe its main building blocks and functionalities and report on possible applications of our CRM framework as an enabler of cross-layer optimization in the future cognitive wireless networks. The implementation details are left out of this thesis, but are available in research reports and articles.

3.1 Resource sharing and management in wireless networks

A Radio Resource Manager (RRM) is an integral part of every modern radio systems today. RRM modules in WCDMA/UMTS cellular networks provide very advanced adaptivity and control capabilities. However, those RRMs tend to be very complex and monolithic regardless of the use of object orientation. Moreover, due to the highly complex software architecture it is increasingly difficult to understand the interactions and dependencies between different modules. Therefore there is a justified doubt, if the present day RRMs or combination of those could simply handle for example, very fast adaptation, context sensitivity or cross-layer optimization in the future cognitive wireless networks straightforwardly. Overall the very scalability of these architectures is questioned. In particular, providing support for cross-layer algorithms which adapt to changes in wireless link quality, radio interface, node density, traffic loads and network topology is extremely challenging and requires advanced control and management structure. One approach to overcome these limitations would be to follow the SDR design principles and provide modular, extensible and easier to implement framework instead of building highly integrated RRMs that are technology and standard specific.

The need to reconsider modular and clean architectures for resource management is not only technical, but also economical. Well-designed architectures are easier, faster and cheaper to extend with different radio resource management modules. They can also support better analysis and debugging of system software. Thus the work done in this thesis and projects related to it has been also strongly supported by industrial cooperators.

The approach taken in WCDMA/UMTS and other cellular networks has been very successful. Base stations are responsible for communication with mobile stations within cells of fixed or variable size. A mobile station, when entering the network, is assigned certain amount of radio resources (in terms of transmit power used, code lengths etc.) by a centralized RRM, which is optimizing the use of the radio resources usually for a number of cells simultaneously. The focus of these RRM designs is at present purely on the PHY and MAC layers of a single technology, and all the optimization decisions are taken by fixed algorithms that are almost completely static software solutions that are either proprietary technologies or often bounded by telecommunications standards. If the algorithms present in the system perform poorly, a system upgrade or intervention by a human administrator is necessary; there is no provision of learning capability in the system. In principle this kind of centralized algorithmic framework can be extended to the multi-mode terminal case as well. An example of such an extension is the Multiradio Radio Resource Management (MRRM) introduced in [74], also known as Joint Radio Resource Management (JRRM) across different radio access technologies. However, this methodology is not without drawbacks. First, it requires almost complete redesign of the presently used RRM frameworks, and necessary extensions cannot be simply "plugged in". Substantially more signaling traffic is necessary to carry all the information required in the MRRM functions to the resource managers. Additional difficulties are complicated and highly non-linear dependencies between the radio resources of different technologies, especially those sharing the 2.4 GHz ISM-band. The need for a priori knowledge on the interactions between the radio resources of different technologies also makes these types of monolithic systems unable to include new standards at runtime, unless complete knowledge on the interactions between the legacy systems and new technology is available. An interesting and novel SDR based approach has been taken for example by Vanu [75], which is, as discussed in the previous chapter, one of the first manufacturers producing SDR base stations. This solution is increasing modularity and provides possibility for dynamic change of the set of technology standards and/or the amount of the carriers spectrum allocated to each of them. While these examples indicate that some progress has been made in development of simple cognitive radio hardware and software at the physical layer, the questions of how to provide flexibility in the upper levels and organize cognitive radios in a network still remain open.

Since the cognitive radio paradigm is expected to drive the next generation radio devices and networks, new architectural approaches for control and management as well as new protocol and algorithm designs are necessary. In this chapter we will present a resource management framework for cognitive radios and networks, which we call CRM framework [10, 16, 18]. The CRM framework was essentially designed upon the following principals:

- **Extensibility and flexibility**: Extensible and flexible architecture is vital for designing a generic system that can support various optimization problems and learning mechanisms. A modular design where new components, such as optimization modules, radio device drivers, flexible protocols and so on can be added, removed or modified with ease, is desirable.
- **Cross-layer interactions and easy flow of information**: Well-defined APIs, which can expose information from different layers of the protocols stack in a generic way and be able to configure system parameters will facilitate cross-layer information flow and resource management.
- **Portability**: It is important and necessary that the CRM framework is portable across heterogeneous devices and operating systems.
- **Distributed system support**: The CRM should be also capable of managing the resources that are distributed over the network. Thus support of coordination between resource managers at each local node is necessary to achieve configuration and optimization on a network level.
- **Reasonable complexity**: The complexity of the CRM framework should be relatively low and should justify the deployment of such a system on the resource limited wireless nodes.

3.2 The framework

Cognitive resource manager is a framework, which can facilitate development of control and optimization techniques for resource management and sharing in wireless networks using the principles of cognitive radio. The cognitive component in the system comes from the capability to support adaptivity that is based on rich context information, learning and reasoning about the changing radio conditions. In fact, the "learning from experience" process can generate valuable knowledge that can be used in the process of optimization to improve the system performance. It is important to notice that the CRM framework enables the use of cognitive cycle and machine learning, but it is not inherently based on artificial intelligence.

The CRM is designed to be a multi-functional entity, which can use feedback, learning and environmental knowledge to make better decisions in order to achieve end-to-end goals. For example, one of the many functionalities of CRM is to operate as a “connection manager” deciding upon the frequency channels as well as the type of communication technology to be used (IEEE 802.11, Bluetooth, UMTS, etc.) in case a variety of interfaces and networks are available. Different types of services (voice-calls, audio- and or video-streaming, etc.) can definitely benefit in quality if their different connection requirements are taken into account and accordingly the most appropriate (in terms of delay, bandwidth, bit rate, etc.) communication link is chosen.

The framework is modular and consists of a large set of components working seamlessly together. Its structure is similar to the design of many micro-kernel based operating systems or distributed systems in general. This approach was chosen as the best fit for our requirements. Figure 3.1 presents a high level architecture of the framework with all its building blocks and interfaces.

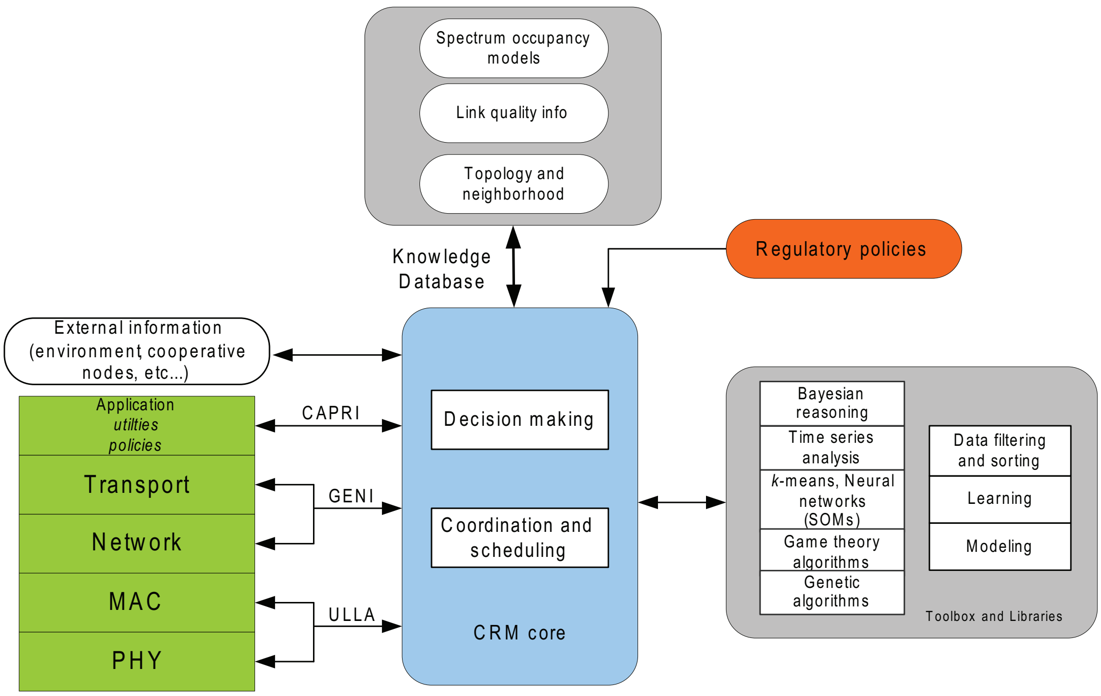

Figure 3.1: CRM framework high-level architecture.

A central component of the framework is a CRM core that functions as a kernel and scheduler of the whole cognitive system. It enables the construction and management of the components that carry out optimization control loops, and their interaction with the environment and the protocol stack. The protocol stack is treated as a separate entity. Nevertheless a connection to the CRM core is realized through well defined interfaces as depicted in Figure 3.1. This design allows still independent development of radio protocol stacks and lowers the complexity of the core CRM. Moreover, taking into account the legacy systems and the investments done toward existing protocol stacks, it is not realistic to assume a clean-slate protocol stack design and implementation. Thus we have chose to support virtually any protocol stack that can be made to interact with our control interfaces.

The CRM core is tightly coupled with a toolbox of optimization and learning techniques. These can be used to perform local (on the node or terminal) and/or global (network) optimization based on information from the protocol stack, environmental readings or outputs from the past learning experience of the system. The optimizers and the machine learning modules are also separated from he CRM core just like in case of the OSI protocol stack. This allows for modular and distributed design as well as later extensibility. This is one of the major differences to other cognitive engine architectures that have been proposed at the same time-scale.

The key differences in this concept compared to the classical RRM are the intelligence and learning capabilities, completely modular design of the CRM, as well as the possibility to perform dynamic decisions and distributed cross-layer optimization. Due to micro-kernel type of architecture the system can be reconfigured flexibly while running. Apart of learning the system can also use other mathematical tools for modelling. Building models can help the radio to predict the (radio) environment in the near future and chose an optimal state of operation by tuning its parameters. For example, the cognitive radio could build spectrum occupancy models from the sensed data using time-series analysis. Time-series are powerful tools for finding out the periodicity spectrum opportunities. Such a spectrum occupancy model will provide the cognitive radio with a priori knowledge of the spectrum allocation in the near future (an hour or several hours in advance), which will allow the CR to chose its frequency accordingly. It is important to note that the toolbox itself is not limited to enable optimizing techniques mentioned above. It is envisioned that further methods could be added in a plug-and-play fashion.

As shown in Figure 3.1 the decision process in the CRM is heavily dependent on the information and triggers it gathers from the relevant OSI-layers, on the one hand, and the set of available optimization and reasoning techniques on the other hand. Based on this knowledge the CRM can take appropriate actions that could, e.g., adapt MAC parameter settings or start certain tasks such as a reassignment of frequency channels. While the classical cross-layer optimization takes into account only the information exchanged between the neighboring layers in the protocol stack, the CRM framework offers more open optimization approach. The relevant data from the separate layers can be collected into one place and exposed to all other layers. All the knowledge (information form the layers of the protocols stack or environmental data from measurements) is exposed to the CRM core via a set of well defined interfaces. Three different types of interfaces named Unified Link-Layer API (ULLA), Generic Network Interface (GENI) and CAPRI (Common Application Interface) have been defined as a part of the framework. As suggested by the names of the interfaces, ULLA is intended for interaction with the link/physical layer, GENI for monitoring and control of the transport and network layers and CAPRI for exposing application and user layer requirements. All three interfaces can abstract the different technologies and protocols, making technology-independent cross-layer optimization possible. Furthermore, they are orthogonal to the normal protocol stack interfaces, which means that for the CRM they are only management interfaces and have nothing to do with the data or control messages passing through the protocol stack. The interfaces will be described more in detail later in this chapter.

In order to provide more flexibility in dealing with constraints in the decision process, a policy language is needed to specify policies between the CRM and the interfaces. Policies are an integral part of CR systems especially in the case of DSA. Policies can be also used to define optimization constraints and strategies. We will talk about policy specifications later in this chapter.

As discussed above, self-configuration and adaptation in CRM are carried out in a cognitive fashion using both up-to-date and historical information. The information collected from each interface can be stored in a knowledge database for processing or later use. The knowledge base, also know as a REM (Radio Environment Map) in the CR community could contain for example, locations and activities of the CRs in a certain area, network topology, link quality information, etc. The knowledge from the data base can be exchanged among cooperative CRMs to enrich the learning and optimization process.

3.3 Interfaces

One of the main functionalities of the CRM framework is to facilitate context-rich and context-aware resource management and system optimization. In the future, dynamic spectrum management and other advanced radio resource optimization mechanisms will require access to more information than currently available. Thus a well-defined abstractions and common interfaces to enable easy flow of information are vital components. Furthermore standard generic API definitions covering methods and attributes related to capability, measurement and configurations of the systems could be very helpful in managing of multi-standard platforms. Many APIs and interfaces available today are generally closed and proprietary limiting the opportunity to deploy and exploit different management functionality. Even within a single operating system, different APIs exist with widely different approaches. Proprietary APIs lack harmonization and standardized generic APIs provide a number of advantages, in particular when wireless systems are managed or controlled by multiple stakeholders. Hence, we argue that the future cognitive radios and networks will benefit form a a generic

API design, which is technology agnostic, light-weight and "future proof". Of course, the level of abstraction of such interfaces will need to take into account the current and future wireless technologies, business models and stakeholder requirements. In the following we will present three generic APIs designed as an integral part of our CRM framework. However, the development of interfaces has been mainly carried out in two collaborative research projects GOLLUM [76] and ARAGORN [77]. Moreover, both ULLA and GENI have been already prototyped on several hardware platforms, whereas CAPRI is currently in a prototyping phase[1]. A free source code of ULLA is also available and can be found on Sourceforge webpage [78].

3.3.1 ULLA

The Unified Link-Layer API or ULLA is an interface that was developed to retrieve data link layer and physical layer information independent of the underlying technology (WLAN, Bluetooth, GSM, LTE, etc). Applications using ULLA do not have to be aware of the communication technology underneath since the interface offers monitoring and controlling for all technologies. The platform independence also gives us the freedom to easily extend the API for the new upcoming standards. However, the API can also provide technology- and implementation-specific details. Detailed description and the performance analysis of the API can be found in [78, 79].

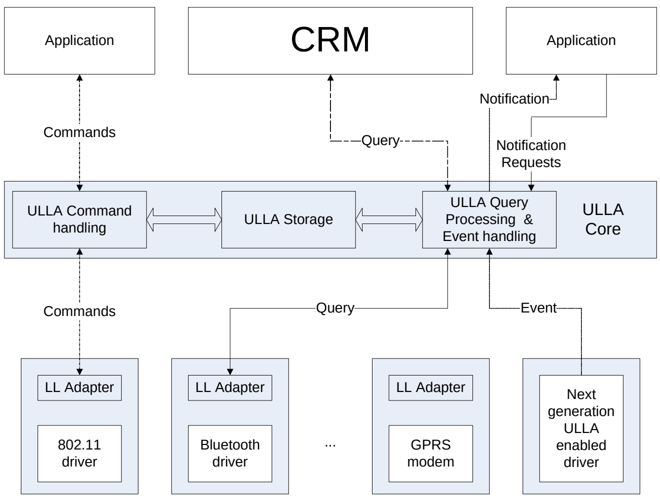

Figure 3.2: ULLA architecture.

A simplified ULLA architecture is shown in Figure 3.2. It is composed of three main parts: applications using ULLA, which are also known as *link users*, *ULLA*

[1]As of writing this thesis ULLA has been submitted in part towards standardization in IEEE SCC41/P1900.4 by two industrial cooperation partners

core and *link providers*, which are abstractions of the network interface cards in the device. It is important to note that a link user must not be an application in a OSI protocol stack sense *per se*. Link users can be applications running on the lower layers that will use link information such as link-aware vertical handover or link-aware routing protocols. Our CRM core is a kind of application that uses the link-layer information for performing knowledge-based optimizations.

The ULLA core is the central part of the API. Its main task is to process the requests and the updates from the link users and link providers respectively. The interaction between the link providers and the ULLA core and the link users and the ULLA core is done through two independent interfaces, namely **link user interface** and **link provider interface**. Through the link user interface the applications can issue **commands** to the link providers to, for example, scan the available links or request specific information (attribute values) from the link providers using **queries**. A ULLA Query Language, which is a subset of the SQL, was developed to specify the queries. For example, an application can request for all links that have latency smaller than 150 ms with the query: `SELECT linkId FROM ullaLink WHERE txLatency < 150`. Furthermore applications can subscribe for **notifications** if they want to be informed regularly about the updated values of the attributes or when the values reach certain threshold. The link provider interface tasks correspond to the ones of the link user interface. It is used for asking for updates in attribute values, for issuing commands, and for registering and de-registering links and link providers. The example below shows how Link User can request a notification when attributes of a link changes. The request might be for a single or periodic notifications. The attribute of interest here is the `rxQuality`. When the notification is triggered the ULLA Core will call a notification handler. An `ullaResult` is passed as parameter to the handler and the Link User parses the result.

```
// a notification handler
void notificationHandler(IN RnId_t rnId, IN ullaResult_t res,
void* privdata){
ULLA_INT_t vInteger;
// The query used for the notification only returns 1 integer field
ullaResultIntValue(res,1,&vInteger);
printf("NOTIFICATION: The rxQuality = %d\n",vInteger);
// Free the result
ullaResultFree(res);
}

// a single event notification
RnDescr_t rnDescr;
RnId_t rnId;
rnDescr.count=1; // fire once
rnDescr.period=0; // the notification is event driven
rnDescr.privdata=NULL; // no data is going to be passed to the handler
// Give a notification when the rxQuality of link 1 goes below 15%
rnDescr.query="SELECT rxQuality FROM ullaLink WHERE id=1 AND rxQuality<15;";
```

```
ullaRequestNotification(&rnDescr,notificationHandler,&rnId,0);
```

Getting multiple events for the same query can be done by setting `count` to the number of events the Link User wants to receive. The period is the time in milliseconds between the notifications. When `count` is set to zero the notification will be fired until it is cancelled.

```
// a periodic notification
RnDescr_t rnDescr;
RnId_t rnId;
rnDescr.count=0; // periodic notification until it is cancelled
rnDescr.period=1000; // fire every second
rnDescr.privdata=NULL; // no data is going to be passed to the handler
// Give a notification when the rxQuality of link 1 goes below 15%
rnDescr.query="SELECT rxQuality FROM ullaLink WHERE id=1;";
ullaRequestNotification(&rnDescr,notificationHandler,&rnId,0);
// The notification will be cancelled when it is not needed
ullaCancelNotification(rnId);
```

We see ULLA as an important enabling technology for the CRM. Using ULLA, the CRM can query particular parameters or information from the lower layers such as bit rate or latency from particular links. On the other hand the CRM can also subscribe to ULLA for periodic or event-based notifications. The periodic notifications will provide the CRM with enough historical information of a certain parameter which will be needed in the learning process or later analysis (e.g time-series analysis of the quality of a certain link). Event-based notifications can be useful for fast interventions. Upon receiving event-based notification, CRM can modify parameters in the protocol stack in order to keep the guaranteed performance or to provide the best possible one. A typical example is the TCP-over-wireless problem, where sudden fading can trigger CRM to regulate the TCP congestion control.

3.3.2 GENI

GENI is an API that provides access to the transport and network layers. The purpose of the GENI interface is to support interaction between the CRM and the network stack. It provides the necessary platform independent abstraction for network, network interfaces and connections using generic API function calls. There are several functional requirements related to GENI, which need to be fulfilled in order for the interface to be generic. First, GENI should be usable with multiple types of transport and network protocols. Second, it should be platform independent. Finally, similarly to ULLA, GENI should allow for asynchronous notifications to be registered. In principle GENI design is based on ULLA architecture. Through GENI the CRM core can access and control the network stack by setting policies that determine the use of different networks or by configuring the network interface profiles. The implementation approach of GENI is similar to that of ULLA so we omit details in this dissertation.

3.3.3 CAPRI

The Common Application Requirement Interface (CAPRI) is an API through which applications can register their preferences and requirements in terms of network constraints and required performances. These requirements are then going to be used by the CRM optimization and decision processes to configure the system in such a manner that the application/user goals are achieved.

In order to be able to quantify the performance of the application and reason about the application "preferences and satisfaction" we adopt the concept of utility maximization. We define the utility as a scalar function on network connection attributes such as throughput, delay, jitter, packet loss rate, etc., that quantifies the "goodness" of the connection for the application. We use slightly generalized form of the network utility maximization problem of Kelly [80, 81]. We furthermore extend the classical concept of *p*olicies to allow applications to introduce additional constraints via specifying policies on network use. As an example, the application could indicate that CRM should try to maximize throughput for its connection (expressed via the utility function), but only using free access technologies, such as open Wi-Fi hotspots (expressed via policy).

Using functional calls CAPRI can register a utility function for a certain application, add a new policy or revoke a previously introduced policies. These functional calls are specified below

$$\text{register\_ utility}(\langle \textit{process\_id} \rangle,\ \langle \textit{utility\_spec} \rangle),$$

$$\text{add\_policy}(\langle \textit{process\_id} \rangle, \langle \textit{policy\_spec} \rangle),$$

$$\text{revoke\_policy}(\langle \textit{process\_id} \rangle, \langle \textit{policy\_name} \rangle).$$

All functions return a value indicating either success or an error code. The first argument, $\langle \textit{process\_id} \rangle$, is in each case used to specify the thread or application the utility and policy specifications apply to. Based on the discussed above, we can summarize that CAPRI allows applications to impose constraints on the optimization process via policies in terms of network attributes accessible via ULLA and GENI. This means that the main entities to be specified through the API are utilities and policies. In the present CAPRI design a separation between the utility and policy specification is kept, although both specifications can be integrated.

In the following we will give a brief overview on the utility and policy specifications, i.e., $\langle \textit{utility\_spec} \rangle$ and $\langle \textit{policy\_spec} \rangle$, respectively.

3.3.4 Utility specification

Let us consider a cognitive network, which has the following type of optimization problem to solve

$$\text{maximise} \sum\nolimits_i U_i(\boldsymbol{a}_i) \tag{3.1}$$

subject to

$$P_i(\boldsymbol{a}_i, \boldsymbol{p}_i) = 1, \tag{3.2}$$

where $\boldsymbol{a}_i$ is the vector of network attributes describing a connection used by application i, U_i is the *utility function* of the application, $\boldsymbol{p}_i$ is the set of configurable parameters related to the connections used by the application, and P_i is the boolean *policy enforcement function* for the application. For example, VoIP client and a web-browser benefit from throughput in a different manner, which can be shown by the utility functions in Figure 3.3.

The parameters $\boldsymbol{p}_i$ correspond to all the “knobs” in a cognitive radio that can be tuned to maximize the application satisfaction. Typical components of $\boldsymbol{p}_i$ are: link selection, packet size, the radio transceiver configuration, a type of TCP or MAC protocols, etc. Finally, the function P_i excludes some of the allowable $\boldsymbol{p}_i$ tailoring the parameter space of the problem according to application policies. A simple example of an application policy may be to only use free Wi-Fi hotspots for downloading email attachments.

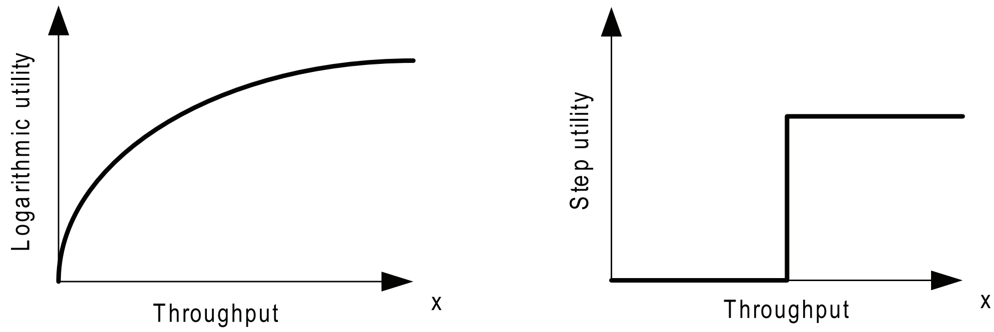

FIGURE 3.3: Examples of one-attribute utility functions. The figure on the left corresponds to a utility function of an elastic bandwidth-sharing application (such as file transfer or web-browsing), whereas the right figure illustrates a typical utility function of VoIP client.

The mathematical formulation of the policies above allows us to also aggregate multiple policies together by taking the logical AND of the values of the individual sub-policies as shown bellow

$$P_i(\boldsymbol{a}_i, \boldsymbol{p}_i) \equiv P_{i,1}(\boldsymbol{a}_i, \boldsymbol{p}_i) \wedge \cdots \wedge P_{i,n}(\boldsymbol{a}_i, \boldsymbol{p}_i), \tag{3.3}$$

This has a high practical value as policies usually originate from a number of sources, including regulators, operators and terminal manufacturers, and same mechanisms can be used in the CRM to aggregate all of these.

For specifying utilities we use a simple dedicated language typical for computer algebra systems. The language used consists of integers and real numbers (in the commonly used dotted decimal or scientific notations), usual basic arithmetic operations (+ for addition, - for subtraction/negation, * for multiplication, / for division and for exponentiation) together with parentheses. Expressions log(a) and exp(a) are included, and evaluate to numerical approximations of the natural logarithm and exponential function with argument a. Additionally, the Iverson bracket notation is supported facilitating the expression of simple conditions, step functions etc. In this notation literal form [⟨*condition*⟩] is used, evaluating to one if the ⟨*condition*⟩ is satisfied, and to zero otherwise. For expressing the conditions the usual expressions =, ! =, <, <=, > and >= are used. Finally, alphanumerical identifiers are used for the arguments of the utility function, corresponding to the various measurable attributes of the connection. At present these are "t" for throughput, "d" for delay, "plr" for packet loss rate, "ber" for bit error rate and "j" for jitter. In this language the utility function version of the good grade VoIP connection requirement expressed above in policy form would simply be

`"[t > 87.2kbps]*[d < 150ms]*[plr < 0.03]"`.

Apart of a specific language for describing utility functions a number of *helper functions* were developed to assist developers in constructing common examples of utility functions easily, without need to understand their detailed use in network optimization. An example of such a helper function is the following

$$\text{step\_utility}(\langle attribute\rangle, \langle threshold\rangle)$$

This helper function constructs a utility function specification consisting of a step function at ⟨*threshold*⟩ for ⟨*attribute*⟩. More details on the utility functions specification language and the introduced helpers can be found in [19, 20].

3.3.5 Policy specification

While constraints on spectrum access are clearly necessary, they are not sufficient on the overall system level. We argue that application and user constraints could be specified and used in the optimization as means to achieve the desired goals. For instance, the users might want to impose constraints on the radio access technologies used for different applications, as in our earlier example on users wanting to use only free Wi-Fi hotspots for downloading email attachments. If the terminal or the network is to perform access selection for the user this should be formalized as a constraint on configurations allowable for the applications involved. In [19] and [20] we discuss the issue on extending the current spectrum-related policy frameworks for cognitive radios to include both useful constraints and goals for reconfiguration and optimization. Our solutions build on the existing work in the policy community on developing frameworks for expressing and reasoning about spectrum policies. We propose to extend the scope of the policy-related constraints by adding technical and economic ones. This naturally requires also extending the policy language used. In our work we chose to build our policy specification using CoRaL policy language [82]. CoRaL is a new language specially designed for formulating machine-readable policies and efficient policy reasoning as a part of the XG policy framework [83, 84]. Although the language primarily is constructed to express policies for opportunistic spectrum access, it is easily extensible and allows for new types of policies to be encoded. We extended CoRaL by introducing new *ontologies* for constraints related to higher levels of the protocol stack, reconfiguration and optimization processes. For example, ontologies for link layer attributes and parameters, and end-to-end connections were defined. No change to the actual language specification was necessary. In the following we will give several examples that illustrate how the new ontologies can be used in user-specified policies.

The example below illustrates a scenario involving costs of using different radio access technologies. For high-bandwidth but delay tolerant applications the user might want to impose a policy of using free connections only. This user restriction can be passed to the CRM through CAPRI with the following policy specification:

policy *free_connections_only* **is**
use *connection_attributes*;

disallow if
    connection.costPerUnit != 0;
end

The second example illustrates how constraints on performance can be expressed using the extended CoRaL policies. Through such policy specifications the applications can profile, e.g., the minimum requirements needed form a connection so that a certain QoS is achieved. This example addresses the quality a connection has to have for a good grade VoIP service.

policy *VoIP_good_grade* **is**
use *connection_attributes*;

disallow if
    connection.throughput $<$ 87.2 kbps **or**
    connection.delay $>$ 150 ms **or**
    connection.lossrate $>$ 0.03;
end

The structure of the policy is similar to the first example. Obviously, much more complex conditions could be specified as well. Essentially any common QoS specification can be expressed in the *connection_attributes* ontology, and using additional utility function specifications increases flexibility and richness of the available information significantly still.

CoRaL policies can be further extended by introducing operator and service provider specific ontologies. These ontologies will consist of expressions of business relations, of payments either made or offered in spectrum auctions, of interference relations between systems and so on. Examples of such policy specifications might include allowing cheap enough roaming provided that the home network of a terminal is not available and a contract exists between the two networks. Such extensions will make it possible to define very rich interactions between the different stakeholders in a dynamic spectrum access scenario. The design of these policies as well as the implementation and enforcement mechanisms are left for future work and are out of the scope of this thesis.

3.4 Toolboxes

Decision-making is one of the most challenging and complex process in a cognitive radio. It includes a large number of actions such as efficient data mining, analysis, modelling etc. The toolboxes and libraries in our framework incorporate all the optimization modules in addition to data filtering, sorting and modelling, and the learning mechanisms that might be needed in a certain cognitive system. Depending on the given probelm(s) to be solved, the CRM core can load instances of one or more modules. For instance the data filtering and sorting module is loaded whenever the CRM core receives data that require filtering and classification. The quality and accuracy of the data are crucial in the decision process. Techniques such as Bayesian reasoning and statistical learning theory can be used to deal uncertainty and ensure the reliability of the data and inference.

For modelling different aspects of the radio environment, diverse mathematical tools are also incorporated into the toolbox. Models are very useful in predicting the environment state in the near future and taking faster optimization decisions. For instance, the CRM can build a spectrum occupancy models from the sensed data using time-series analysis. Time-series analysis is another powerful method to characterize the activity pattern in a given frequency based on the historical data. Geo-location and generally topology models of the network have also proven to be useful in the optimization. We have earlier studied that the distribution of the users in a network have a great impact of the throughput [85, 86]. Thus, a tool such as topology engine, which can use ready made models or generate topology models of a given environment can be part of the CRM toolbox.

Learning mechanisms are also part of the toolbox and could be used to assist the future decision by learning from the prior decisions and system behaviors. Learning could potentially enable high degree of adaptivity for the system. Learning mechanism such as reinforcement learning could enable the system to interact with environment and configure radio and protocol parameters based on a feedback information. Finally, different optimization modules could be also plugged in the toolbox and activated by the CRM core. Typically optimization problems include optimal management of spectrum resources, adaptation of radio and link parameters in order to achieve the required QoS from the applications running, etc.

In principle, the toolboxes discussed above represent a small selection of mathematical tools and algorithms that can be useful for achieving adaptation, learning or optimization. In fact, the modular architecture of CRM will allow the system to integrate any type of modules in a plug-and-play fashion. The only requirement for seamless integration is having interfaces that are compatible with the framework.

3.5 Related work

Several academic groups have been working, on high-level cognitive radio architectures, which partially capture the Mitola's initial vision of cognitive radio. Most active in this regard have been research groups at Virginia Tech, Kansas University and Trinity College Dublin. Nearly in the same time frame, independent from our work, several cognitive architecture proposals, with various degrees of maturity, emerged in the literature. In [73, 87] the authors described a three-layer software framework that can be used to implement a cognitive network. Furthermore the authors introduce a concept of software adaptable network (SAN), which provides the actual physical control of the system and action space for the cognitive processes. In principle SAN partially is similar to our CRM in the functionalities it provides. For example, the notion of extensible and flexible SAN API, which can modify the so called cognitive elements through a cognitive process is closely related to our generic APIs (ULLA, GENI and CAPRI). Similarly to the CAPRI approach, Thomas *et al.* proposed a cognitive specification language that maps end-to-end QoS-like requirements of the users and applications to underlying mechanisms. However, to the best of our knowledge, the proposed framework is purely conceptual and no progress has been done so far in terms of implementation. About a year earlier, Rieser introduced the concept of cognitive engine (CE) [9,88]. Motivated by the reconfigurability opportunities in the SDR platforms the idea behind the CE was to create a software architecture for intelligent control and tuning of radio parameters. The proposed CE uses genetic algorithms to find the optimal combination of PHY and MAC settings and was tested on a programmable radios. The latest status on the CE development and prototyping has been recently summarized in [12].

A cognitive engine implementation for OFDM multicarrier transceivers has been also developed by Newman *et al.* at Kansas University [13]. In comparison to the work at Virginia Tech the authors here provide also some numerical analysis of the dependencies between the environmental parameters and the transmission parameters of the radio. This work has been a natural extension of the earlier proposed flexible software defined radio development platform called KUAR [42].

A complementary concept to the cognitive engine, called a reconfigurable node has been proposed by Sutton *et al.* in [89]. The authors present a node architecture for cognitive networks, designed to enable network-wide observation and adaptation. The reconfigurable node architecture consists of a layer and a radio component. The layer component shares some functionalities with our CRM, namely it provides the interface towards the protocol stack and enables inter-layer communication and reconfiguration. The radio component is simply a flexible software architecture running on a general-purpose processor. So far the platform has been manly used for dynamic spectrum access experimentation [90].

Several other groups have also shown interest in frameworks and architectures for cognitive networks. In [91], for example, Raychaudhuri *et al.* present an architectural framework called CogNet for integration of cognitive networks into the future Internet. Under the same name Manoj *et al.* proposed recently a conceptual cognitive knowledge architecture that exploits the context in a system-wide manner [92]. Mainly following the operators and manufacturers point of view the European research projects E^2R II and E^3 [93, 94] have come up with an architecture for seamless reconfiguration of a network in order to allow for universal end-to-end connectivity. This architecture, which is mostly on a level of a reference model, is focused on operation and maintenance of 4G cellular and wireless networks.

At this point we would like to stress out that our CRM framework was one of the first cognitive resource management architectures on a system level that suggested all the necessary components to implement the cognitive cycle. Lately the architectural work in context of cognitive radios and networks have been focusing on different components with specific tasks. Typical examples are the research efforts on designing policy engines and radio environmental maps (REMs). For more information the reader is referred to [20,86,95–97] and the references therein.

3.5.1 Similarities with the traditional AI cognitive architectures

Obviously, designing advanced cognitive architectures has been a subject of extensive research in Artificial Intelligence (AI). One of the major challenges there has been to come up with a framework that allows computer systems for automated problem definition, problem resolution and resource usage, similar to the human brain. So far, a range of cognitive architectures have been proposed and developed following various paradigms and methodologies (e.g. cognitivist, emergent and hybrid approaches) [98]. The term cognitive architecture originated from the seminal work of Newell et al. in attempt to come up with a unified theory of cognition [99, 100]. Cognitive architecture represents, in fact, a collective definition of cognition covering a number of cognitive features such as, memory, problem solving, decision making, learning, etc. From a cognitivist paradigm point of view, the cognitive architecture models the aspects of cognition that are constant over time and independent from the tasks the system performs. In this respect, the cognitive architecture defines the way in which the cognitive system or cognitive agent manages the computational resources. Furthermore, the architecture specifies the knowledge representations and the memory used to store those, the processes initiated based on the knowledge as well as the learning mechanisms that acquire that knowledge. The emergent approaches define a structure where the mechanisms for perception, action, adaptation, anticipation and motivation can be embedded. This allows the cognitive agent through self-organizing to construct its reality while operating in the environment.

Our CRM framework can be easily mapped to several well-known cognitive architectures. For example two different cognitivists architectures, namely SOAR and ICARUS have a great potential as candidates for modelling the complete cycle of cognitive radio [101–103]. Additionally, the design principles of an emergent type of architecture called subsumption architecture can be also used in building a cognitive radio and cognitive wireless network [104]. In the following we will give a short description of the three cognitive architectures so that the reader can easily infer the similarities between CRM for cognitive radios and mainstream cognitive architectures for AI systems.

SOAR originates from the classical artificial intelligence and is one of the first cognitive architectures in the literature. It practically implements the conceptual mode of human intelligence presented by Newell. The main objective of the architecture is to mimic a wide range of capabilities of an intelligent agent through learning from experience. Its core elements are problem spaces that represent tasks by states, operators, and goals. Problem solving is performed by a decision cycle that processes different memory types for reaching a goal that is defined as a final state in some problem space. Two integrated components allow enhanced intelligent problem solving: so-called impasse handling which is a procedure to automatically handle non-resolvable problems, and chunking that allows learning in order to improve the system's problem solving capabilities. A simplified block architecture of SOAR is given in Figure 3.4. In its essence SOAR is a production (rule-based) system that operates in cycles, namely production cycle and decision cycle. By production system we mean a system that operates on `if...then` condition-action principals. The memory structure in SOAR includes long-term memory (LTM), a working memory (WM), which is practically considered to be a short-term memory and a preference memory. The inputs gathered from perception and monitoring of the external environment are stored in the working memory. The working memory is manipulated by the so-called decision cycle as the central problem solving mechanism of the architecture. The general domain knowledge of the system is stored in the long-term memory. There are three different classes of long-term memory, depending of the type of knowledge being stored: procedural, semantic and episodic.

ICARUS is a cognitive architecture for physical agents, which uses stored concepts and skills, organized in hierarchy, to recognize familiar situations and control its behavior. It belongs to the group of cognitive architectures that exploit symbolic representations of knowledge, use pattern matching to select the appropriate knowledge elements and employ incremental learning. Hence, ICARUS shares some characteristics with SOAR, but also differs from it in several aspects. For example, ICARUS goes beyond the concept of single representation

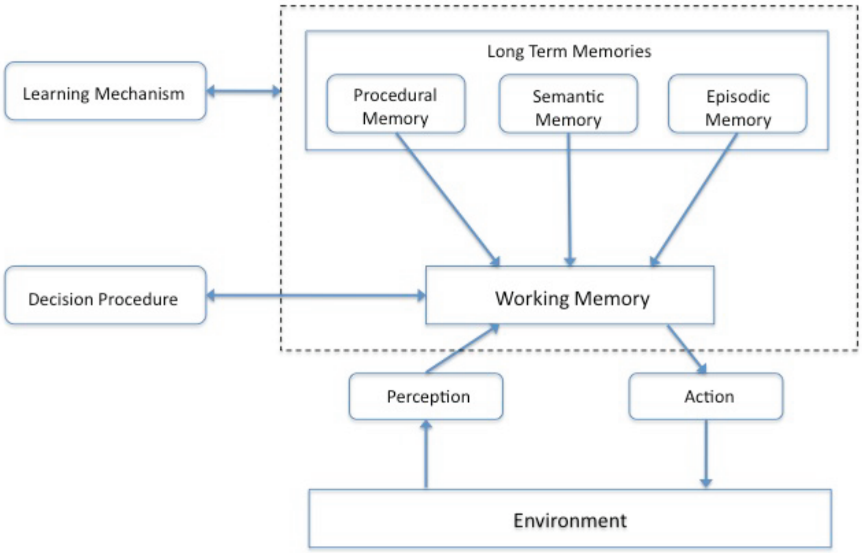

Figure 3.4: The SOAR cognitive architecture (adapted from Lehman et al. [105].)

of long-term knowledge and introduces two separate long term memory types, namely long-term conceptual memory and long-term skill memory. The first one stores knowledge about general classes of objects and relations whereas the latter one knowledge about ways to achieve goals. Furthermore, ICARUS adopts hierarchical structure of the long-term memory and provides a support for it on an architectural level. ICARUS also distinguishes between short-term conceptual and short-term skill memory. Each short-term memory element is a specific instance of some long-term structure, in contrast to the SOAR architecture where no mapping between the production rules in LTM and the WM exists. Figure 3.5 depicts the ICARUS architecture.

Subsumption architecture is behavior-based architecture developed inside the robotics community. It was introduced by Brooks as a bottom-up approach to deal with problems in the traditional artificial intelligence such as extensibility, robustness and managing and achieving multiple goals [104]. The subsumption architecture is organized in layers of different behavior-based modules. Behavior generally refers to the actions or reactions of an object, usually in relation to the environment or surrounding world of stimuli. All the layers can access the sensor's data and multiple behaviors can operate in parallel. The lower layers allow the agent to promptly adapt to the quick changes in the environment and the higher layers are responsible for the overall goals of the system. That said

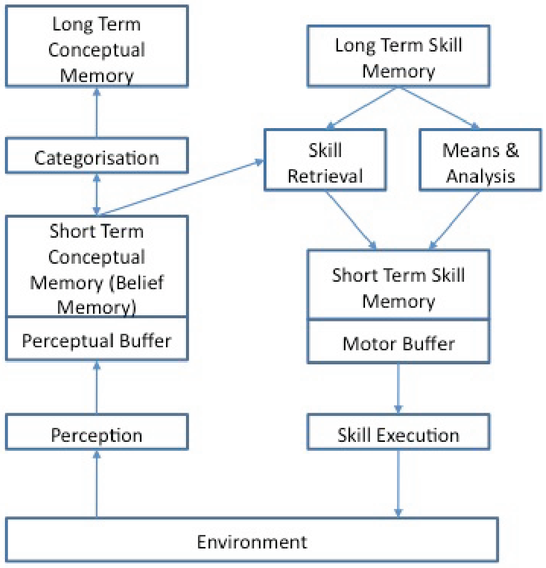

FIGURE 3.5: ICARUS block architecture (adapted from Chong et al. [106].)

the behaviors coming from the lower layers are reactions to the environment, whereas the higher lever behaviors are driving the system towards achieving the primary goals. In fact each layer realizes a sub-goal of the more complex overall goal. The higher-level behaviors of the controller may take advantage of complex optimization and learning functions such as partial plan generation and time series learning, while the lower layers provide a tight coupling between the sensory input data and the actuation, and often employ reactive learning algorithms with little to no state such as self-organizing maps, decision trees or hard codes input to output mappings. The system can increase its competence by simply adding a new level without changing any of the existing ones. One of the main characteristics of the subsumption architecture is that it is free of central control. No dynamic communication or global data exchange takes place among the layers. The communication between two layers is one directional, resulting in a minimal interaction among the layers in the system. This makes the system reactive and actions occur as a consequence of the changes in the environment. Finally, there is no explicit representation of knowledge in the subsumption architecture.

Although the subsumption architecture is suitable for many control systems, it has a number of drawbacks when it comes to practical implementation of cognitive radios. The complexity of the system can easily increase and the communication between the layers can become quickly unmanageable by adding more layers into the structure. Moreover it is not certain if the hierarchical control loops are the most suitable approach to achieve flexibility in the decision-making in cognitive radios. The lack of explicit knowledge representation is certainly problematic as it is crucial for formalizing the logical rules, states and actions in the CRs. Both SOAR and ICARUS are quite suitable as reference models for implementing our CRM. Especially the presence of short and long memory is of great importance for cognitive radios to facilitate online and off-line learning and structure the gained knowledge. However, it is good to note that it is not a necessary requirement to strictly follow any of the above discussed cognitive architecture in order to realize CRM. One can also implement the CRM as an independent kernel code, following the approach how modern operating systems kernels are built. The attractiveness in mapping CRM to ICARUS or SOAR architecture is that those provide a natural mechanisms to have short- and long- term memory, and provide a clean reference for reasoning. More thorough analysis on the practical implementation issues of the proposed CRM framework are out of the scope of this thesis and have been left for future work.

3.6 Summary

In this chapter we have presented a new architectural framework, which can be used for implementation designs of cognitive radios and cognitive networks. In the proposed cognitive resource manager framework we map Mitola's abstract suggestions on a cognitive radio cycle to a component based system, which can facilitate context-aware and cross layer optimization, environmental awareness and intelligent and efficient resource management. Although the presented CRM framework has some similarities previously suggested approaches, its scope and design is at the same time original and complementary. In particular we have addressed the lack of modularity and extensibility of the classical RRMs and earlier cognitive engines. We have also proposed a generic design of APIs to facilitate easy flow of information and context-based optimization. A toolbox of intelligent algorithms for learning, optimization and adaptation, which can be loaded and unloaded in a plug-and-play fashion has been also introduced as an integral part of the framework. In the rest of the thesis we will focus on possible algorithmic solutions for a small set of resource allocation problems, which will become a part of the CRM toolbox.

4

Genetic Algorithm based Optimization

The optimization problems we face in wireless communication systems are mostly multi-objective. Finding solutions for such problems is challenging as the optimization goals tend to conflict with each other. Hence, to obtain the optimal solutions trade-offs between the conflicting objectives have to be made. Today, there are in principle two methodologies to approach multi-objective optimization problems (MOOPs), namely classical methods and evolutionary algorithms. Classical optimization methods convert multi-objective optimization problems in a single-objective problem using a preference-based strategy. Changing the preference vector can lead to computing a different trade-off solution in the next iteration. Classical methods can at best find one solution in one simulation run and therefore are inconvenient for solving multi-objective optimization problems, where a set of optimal trade-off solutions are preferred. In contrast to the classical approaches, the evolutionary algorithms (EAs) use a population of solutions in each iteration, instead of a single solution. The ability of EAs to find a set of optimal solutions makes them attractive methods for solving MOOP.

In this chapter we study evolutionary optimization approach for cognitive radios by adopting genetic algorithm (GA) as basis for our solution. We investigate the applicability of GA for multi-objective optimization where the target is to find optimal radio parameters such that desirable throughput, bit error rate (BER) and power consumption are achieved. Furthermore, we propose a genetic algorithm based approach to optimally set PHY and MAC layer parameters in OFDM multi-carrier system in a cross-layer fashion with minimal feedback information. The results show that our GA-based optimization can provide true benefits in the future cognitive wireless systems.

4.1 Multi-objective optimization

In the simplest case the optimization in the wireless communication systems is oriented towards maximization of the throughput or minimization of the latency or the bit error rate. However, as the number of applications running on the wireless systems and the variety of the services offered to the end user are constantly increasing, the optimization of the system performance becomes challenging as multiple goals have to be taken into account. Very often several objectives from different applications, running simultaneously, are to be optimized as fairly as possible. Just for illustration, minimizing the packet loss, maximizing the throughput, maximizing the spectral efficiency and minimizing the power consumption is a very likely set of objectives desired to be achieved at one time instance in the system. In this case there is a need for a multi-objective optimization (MOO) [107]. Most of the time it is a challenging procedure as some of the objectives may be linked to each other, as shown later in this section. For example it is unrealistic for the system to optimize the throughput while minimizing the bite error rate (BER) at the same time without any compromises. A conflict occurs in the physical layer due to the modulation order which is affecting each objective in an opposite direction [21].

In general, the multi-objective optimization is concerned with the minimization or maximization of a vector of objectives, also known as a *fitness* or *utility function*

$$\vec{y} = \langle f_1(\vec{x}), f_2(\vec{x}), ..., f_m(\vec{x}) \rangle, \tag{4.1}$$

subject to

$$\vec{x} = \langle x_1, x_2, ..., x_n \rangle \in \mathcal{X}, \tag{4.2}$$
$$\vec{y} = \langle y_1, y_2, ..., y_m \rangle \in \mathcal{Y} \tag{4.3}$$

where x is a set of decision variables, y is a set of objectives and $\mathcal{X}$ and $\mathcal{Y}$ are the parameter and objective space respectively.

Wherever multiple objectives or decision variables are treated it is usual that only some values of each are feasible in practice. Determining the optimal set of decision variables for a single objective often leads to an non-optimal set with respect to other objective. In most situations, where values interact, it is quite impossible to optimize all the objectives at the same time, so we must adopt various trade-offs and compromises. One of the problems that arises in this context is that often the relative importance of these objectives is not generally known until the best capabilities of the system are determined and trade-offs between the objectives fully understood. As the number of objectives increases, trade-offs are likely to become complex and less easily quantified. Thus it is very important to clearly formulate and express the preferences throughout the optimization cycle.

Due to the trade-offs among the objectives, there exist no single optimal solution but a set of solutions that lie on a so called *Pareto-optimal front* [108,109].

Definition 4.1. *A vector of decision variables $\vec{x^*} \in \mathcal{X}$ is Pareto optimal if there does not exist another $\vec{x} \in \mathcal{X}$ such that $f_i(\vec{x}) \leq f_i(\vec{x^*})$ for all i=1,...,m and $f_j(\vec{x}) < f_j(\vec{x^*})$ for at least one j.*

Definition 4.2 (Non-dominated set). *Among a set of solutions $\mathcal{P}$, the non-dominated set of solutions $\mathcal{P}'$ are those that are not dominated by any member of the set $\mathcal{P}$.*

The vectors $\vec{x^*}$ corresponding to the solutions on the Pareto front are globally optimal and are result of the trade-offs between the multiple objectives. Further improvements are not possible due to the correlation between the multiple objectives. Pareto-optimal solutions are also known as *efficient*, *non-dominated*, or *non-inferior*. All the other solutions are said to be *dominated* by the optimal ones and can be discarded.

In the following two figures Pareto fronts for two different multi-objective optimization problems are given. In Figure 4.1 an example of a Pareto-optimal front is given for a fitness function that tends to maximize the throughput and minimize the BER in a radio system for different power levels. Optimizing these two objectives at the same time creates a conflict for any transmission power as shown in the figure.

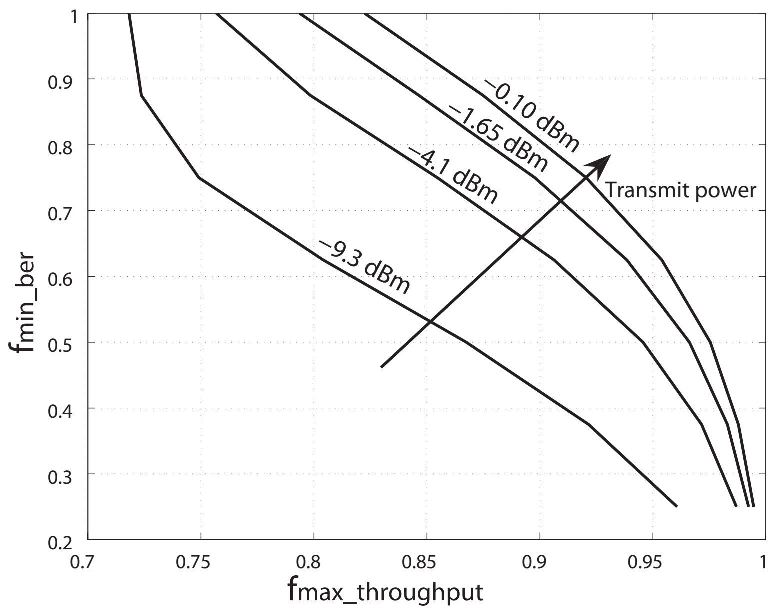

FIGURE 4.1: Pareto optimal front. Both objectives $f_{max\_throughput}$ and f_{min_ber} conflict with each other any transmit power.

In Figure 4.2 the Pareto front trade-offs between the minimum BER, $f_{min\_ber}(x)$, and minimum latency, $f_{min\_latency}(x)$, for several error correcting coding rates are given. The x-axis of the graph represents the fitness function used to minimize the latency for several coding rates from $R = 7/8$ to $R = 1/2$, while the y-axis the fitness function for minimizing the probability of error. In this example we chose a multi-carrier system with $N_c = 64$ sub-carriers and $SNR = 3$ dB. It can be clearly seen that as the fitness value of $f_{min\_latency}(x)$ drops, the fitness value of $f_{min\_ber}(x)$ increases. These curves present the key constraints for carrying out multi-objective optimization. For a certain coding rate these curves represent

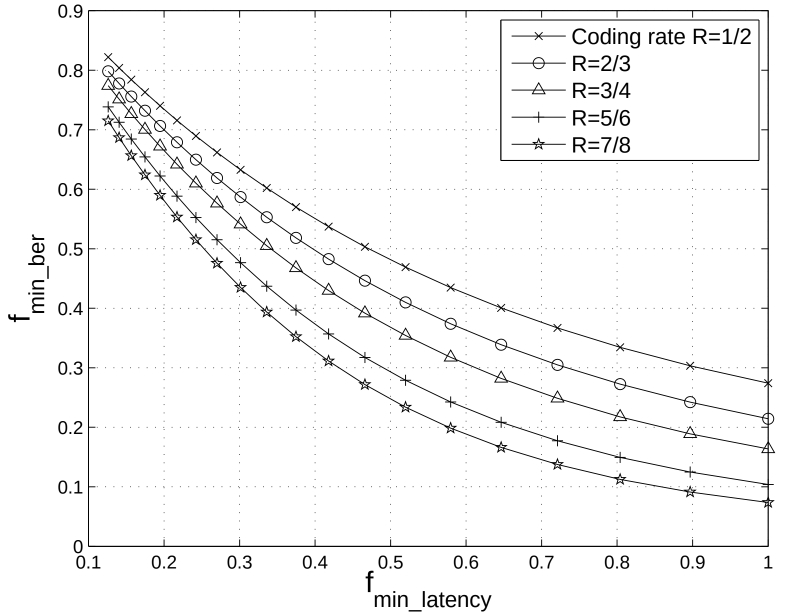

FIGURE 4.2: An example of Pareto front for trade-offs between the minimum BER $f_{min\_ber}(x)$ and minimum latency $f_{min\_latency}(x)$ for several error correcting coding rates.

the Pareto front as no parameter set on the curves can be set so that the fitness value in respect to both objectives $f_{min\_ber}(x)$ and $f_{min\_latency}(x)$ improves.

There are several ways to address multi-objective optimization problems mainly grouped in classical and evolutionary methods. Under classical methods, here we understand optimization and search algorithms that use a point-to-point technique to approach the optimum and find a single solution in each iteration. Classical methods have been around for long time and a large number of algorithms have been developed and successfully used so far [107]. Most popular classical algorithms include: weighted sum approach, ϵ-constraint method, goal programming methods, etc. In order to find the Pareto-optimal solution, the classical methods usually convert the multi-objective optimization problem into a single-objective optimization problem by using variety of user-defined conversion methods. For example, the weighted sum approach minimizes a weighted sum of multiple objectives, the ϵ-constraint method optimizes one objective function and uses all other objectives as constraints, the goal programming minimizes a weighted sum of deviations of the objectives from the user-defined targets, etc. The solution gained as a result of the optimization is expected to be a Pareto solution, but it is specific to the parameters used in the conversion process. Thus if multiple Pareto-optimal solutions are to be find the classical algorithms require some problem knowledge, i.e., suitable weights or ϵ or target values. In order to find another Pareto-optimal solution, a new single-objective optimization problem with different parameters has to be solved. If the MOOP is nonconvex some of the classical methods, such as the weighted sum approach, are not able to find the Parteo-optimal region. Apart of their shortcomings, classical methods also have a number of advantages. For some of the algorithms there exist proves that the optimal solution of the converted single-optimization problem is one of the Pareto-optimal solutions. Another attractive feature of these methods is that they are often simple and easy to implement.

To approach to the optimal solution, these traditional algorithms use direct search or gradient-based methods. In direct search only the objective function and the given constraints are use steer the search, whereas in gradient-based methods the first and second order derivatives of the objective function and/or the constraints guide the search. Therefore the latter methods can quickly converge to the optimum, but are proven to be inefficient for discontinuous problems. Several drawbacks are common for both types of search methods:

- Most algorithms tend to be stuck to a local suboptimal solution.
- The convergence to an optimal solution is dependent on the initial solution.
- Difficulty in handling problems with discrete search space.
- The algorithms are not suited for parallel processing.

The discussion above outlining the weaknesses of the classical optimization methods indicates that these methods, without adjustments and modifications may not be suited for practical engineering optimization problems. In this context, evolutionary algorithms have been successfully applied to solve multi-objective optimization problems in the last twenty years [110]. Evolutionary algorithms are particularly attractive for solving MOOs because they deal simultaneously with a set of possible optimal solutions. This allows for finding the Pareto-optimal set in a single run of the algorithm instead of running series of in the case of the conventional approaches mentioned above. We apply evolutionary approach to solve multi-objective optimization problems in context of cognitive radios by adopting genetic algorithm (GA). In the rest of this chapter we will provide an insight of the GAs and emphasize the main results and findings of using GA in multi-objective optimization for selected problems in cognitive wireless systems.

4.2 Genetic algorithms

Genetic algorithms (GAs) [111,112] are proven to be successful machine learning methods for solving complex optimization problems. Their main strength is the capability to efficiently find the global maximum/minimum in large search spaces using the principles of the evolutionary processes in the nature.

GAs find the optimal solution of a given problem by searching a population of candidate solutions. Each member of the population is composed of set of parameters and can be represented as a bit string. In the evolutionary theory, each string is named a *chromosome* and the parameters are regarded as a *genes* of the chromosome. The first step in executing a genetic algorithm is the generation of the random population of N bit strings[1] which represent the set of possible problem solutions. The next step is the process of evaluation of each string. The degree of "goodness" of a bit string for solving a particular problem is determined by the value of a so called *fitness function* $f(x)$. The fitness function defines the criterion for ranking and probabilistic selection of the members in the next generation. The selection process is such that the better bit strings are given more chances to reproduce and survive than those which offer poor solutions to the fitness function. There exist number of ways to carry out the selection. Some common methods are *tournament selection*, *proportionate selection* and *ranking selection* [113]. For example, in the tournaments selection, two solutions play a tournament and as a result the better solution gets a place in the mating pool. With the proportionate selection solutions are assigned copies proportionally to their fitness values. If the average fitness value of the population is f_{avg}, a solution with fitness f_i will get f_i/f_{avg} copies, etc. After the selection process has been done, an offspring can be created through a *crossover* as in Figure 4.3. The crossover occurs between a pair of randomly selected bit strings with a probability p_c. After the crossover a mutation of the bits in the bit strings can happen with a probability p_m. This is the mechanism that avoids the convergence of the algorithm towards a local optimum. By altering random bits within a chromosome, it allows the GA to explore the search space apart from areas that can be reached by mating. One will usually not apply mutation on the best performing chromosome(s) of a generation to preserve the currently best results. This is achieved by a pre-defined elitism rate, r_{elite}, which determines how many chromosomes will be passed to the next generation without being changed. As long as the mutation is finished, the GA has completed one interaction. The algorithm usually runs till it converges to a stable solution. In Figure 4.4 a flow chart of the operational cycle of a binary GA is given.

Genetic algorithms are able to explore the search space in several locations at the same time. Furthermore, GA search can move much more faster compared to the neural networks backpropagation algorithm, for example, replacing the parent chromosome by an offspring that may be radically different from the parent [115]. This is exactly what makes GAs such robust and powerful optimization tools. They are good in finding optimal (or close to optimal) solutions and it is less likely that they fall into local minima. GA based algorithms are particularly well suited to explore optimization possibilities, if the parameter space is large and distributed decisions need be made. For some problems they operate less than a

[1] We are referring here to binary GA. There exist also continuous genetic algorithms that use the same principles as the binary ones but work with continuous variables.

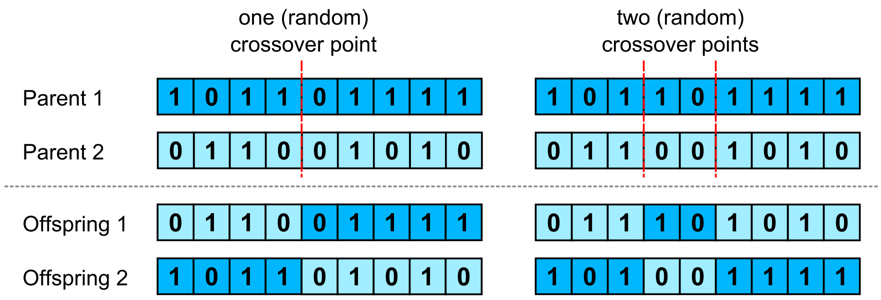

Figure 4.3: Crossover and generation of new offspring.

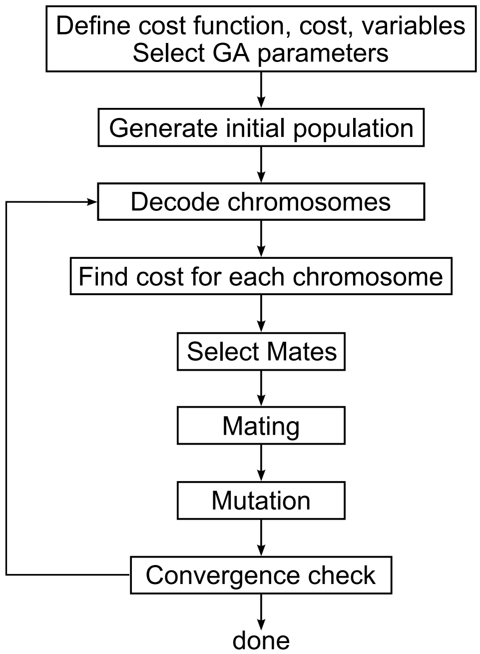

Figure 4.4: Flow chart of the binary GA [114].

real optimizer and more like a data mining tool. With this we mean that if the utility function is not complex and the number of parameters is relatively low then GAs find solutions, but the actual on-line parameter search and setting could be implemented much more efficiently by using, e.g., decision trees or adaptive fuzzy logic based methods.

We use genetic algorithms to evolve radio and MAC parameters with a goal of optimizing the cognitive radio performance. One of the earliest works on using GAs in the cognitive radio context and wireless networks in general has been presented by Rieser [88]. However, our work goes beyond optimizing only PHY layer parameters of the radio and considers MAC layer parameters, i.e., cross-layer optimization. The use of GAs in cross-layer optimization, where the parameter space includes variety of data from all OSI-layers and network topology, showed to be an extremely useful method. We argue that if the CR can afford the latency and computational power for GA optimization, it is one of the most versatile tools to include in the toolbox of CRM architecture that was presented in Chapter 3.

In order to carry out a multi-parameter optimization the genetic algorithm will need a well defined list of parameters that can be read and/or set through ULLA-type of interfaces and can straightforwardly define the fitness function as discussed in the previous section. Some examples of these parameters include: signal-to-noise ratio (SNR), carrier frequency, packet size or fragmentation size, packet retry limit, transmission power (P_{tx}), contention window size, etc. A set of these parameters will be applied to the fitness function which will give an estimate how well the parameter set meets the optimization objective. The parameter set will be further on manipulated with the GA until the optimal solution is found. In the rest of the chapter we will present several optimization problems in terms of cognitive radio, which we addressed by employing genetic algorithm.

4.3 GA optimization test cases

In this thesis we consider three objectives that represent common wireless communication optimization goals:

1. *Minimizing the symbol error rate* increases the reliability of the radio transmission. This objective can formulated as
$$f_{min\_BER} = 1 - (1 - P_e)^N, \tag{4.4}$$
where N is the packet length and P_e is probability of transmission errors.

2. *Maximizing the throughput* is also an important objective and in this case it is given by
$$f_{max\_tp} = \frac{k}{k_{max}}, \tag{4.5}$$
where k is the number of bits per symbol and k_{max} the maximum order. In our implementation, we choose $k_{max} = 4$.

3. *Minimizing the transmission power* aims to minimize power consumption and interference with other transmissions. In the case of portable devices with limited battery power this issue is of paramount importance. Additionally, transmission power has to be also set low in order to limit the interference with the other transmissions. The objective function can be stated as
$$f_{min\_power} = 1 - \frac{P_{Tx}}{P_{Tx\_max}}, \tag{4.6}$$
where P_{Tx} is the transmission power and P_{Tx_max} is the maximum available transmission power.

Determining the optimal set of decision variables for a single objective, e.g., minimizing power, often results in a non-optimal set with respect to the other objectives. Consequently, if we take all of the aforementioned objectives as inputs to a multi-objective optimization, they need to be ranked with regard to their importance. In our work we use a weighted sum approach to construct the aggregate fitness function. This approach has been presented in [116] as an attempt to maximize the sum of the positively normalized, weighted single objective fitness scores for each parameter set. In [13, 117] Rondeau *et al.* and Newman *et al.* use also this approach to autonomously adapt a cognitive radio. Aggregating our three objectives, the scalar fitness function holds:

$$f = w_1\, f_{min\_BER} \;+\; w_2\, f_{max\_tp} \;+\; w_3\, f_{min\_power}, \tag{4.7}$$

$$w_{BER} + w_{tp} + w_{power} = 1$$

This method suits the cognitive radio concept well since any solution with respect to the user's QoS requirements can be reached simply by changing the weights in Equation (4.7). Although simple and straightforward, the weighted sum approach is not without disadvantages. In a situation where some of the objectives are to be minimized and some need to be maximized, all objectives should be converted in one type. Furthermore, problems arise on issues such as how to normalize, prioritize and weight the contributions of the various objectives in order to come to a suitable measure. In the following we will present a simulation study on the convergence properties of GA used to optimize radio parameters such as modulation order and transmission power level of an OFDM (Orthogonal Frequency Division Multiplex) transceiver. Additionally, we will show experimental results of using GA optimizer for dynamic channel selection in a time-varying radio environment.

4.3.1 Optimization of the radio parameters of OFDM transceiver

Let us consider an OFDM system with the parameter as shown in Table 4.2. We want to configure the modulation type and power level per subcarrier so that Equation 4.7 is optimized. Please note that instead of BER in the first objective function, here we are using SNR as the PHY-layer in our simulation model has no *a priori* knowledge about the data that it is going to receive and has only limited capabilities in detecting demodulation and decoding errors. Hence, it is difficult to calculate the exact BER. Instead, we use an estimated SNR as input for the fitness function. The estimation is done in the demodulator by re-modulating the demodulated signal. The difference between the received signal and the remodulated signal serves as the estimation of the SNR. When the BER is below a certain threshold, the transmission is considered error-free. This threshold is usually set to 10^{-6}, which means that in one million transmitted bits a maximum of one wrongly detected bit is allowed. Table 4.1 shows at which SNR values different modulation orders hit a specific target error rate TER. We use these thresholds as normalization factors in the fitness function.

TABLE 4.1: SNR thresholds for different target bit error rates (TER).

Description	TER $= 10^{-5}$	TER $= 10^{-6}$
BPSK	8.00 dB	9.15 dB
QPSK	10.43 dB	11.80 dB
16QAM	16.85 dB	18.30 dB
64QAM	22.60 dB	23.55 dB

$$f_{\text{SNR}} = \frac{\text{SNR}_{\text{estimated}}}{\text{SNR}_{\text{threshold}}(k)}. \tag{4.8}$$

The set of possible modulations has four elements as shown in Table 4.2, whereas the set of possible power levels for each subcarrier has 16 normalized values. These two transmission parameters can be straightforwardly mapped to a 6-bit chromosome so that all the possible combinations of the parameter values are encoded as shown in Figure 4.5. We denote the population size with N_{PS}. When the last chromosome of the current generation is received at the receiver and all scores are assigned, the GA starts selecting and mating the chromosomes for the next generation. As a part of the selection process, the chromosomes are sorted according to their scores and the ones with a score worse than the pre-defined threshold, s_{th}, will be discarded. To keep the best performing chromosomes for the next generation, we copy the top N_{elite} chromosomes to the next generation. The *elitism rate*, r_{elite}, is set to 20 % , which means that the top 20% will manage in the next generation.

The rest of the chromosomes for the next generation are generated by mating the survivors, i.e. the chromosomes above the score threshold.

$$N_{\text{elite}} = N_{\text{PS}} \cdot r_{\text{elite}} \tag{4.9}$$

Mating is performed by tournament selection. For both parents, we randomly select three competitors out of the survivors and the one with the highest score becomes a parent. As the chromosomes are already sorted, this is implemented by a random selection of three numbers out of the number of survivors. The smallest number among the three is the winner, because the chromosome score decreases with increasing the order number. In order to ensure diversity among the offsprings, we make sure that a combination of two chromosomes A and B has not been already selected to produce offspring. At very low SNR values, it might happen, that the number of survivors is too small to fill up the next generation. In this case, we mate all chromosomes with each other and for the missing places in the population, we randomly create new ones. After having created all the chromosomes for the next generation, we mutate N_{m} bits within a chromosome according to the mutation rate r_{m} except for the elite chromosomes.

We run extensive simulations to determine the convergence properties of the GA in optimization of OFDM transceivers. Table 4.2 shows the OFDM system parameters used in our simulations. We used Matlab/Simulink to carry out the simulations. The rest of the simulation parameters that characterize the system model are given in Appendix A.

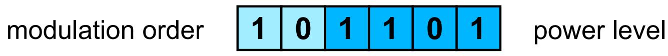

FIGURE 4.5: Chromosome layout per subcarrier.

TABLE 4.2: OFDM system parameters.

Parameter	Value
FFT size	128
Number of data subcarriers	100
Lower guard band (LGB)	12
Upper guard band (UGB)	11
Number of pilot tones	4
Possible modulations	BPSK, QPSK, 16QAM and 64QAM

The following plots show simulation results, which are based on the default parameters over an AWGN channel[2]. The best performing chromosome of a population is plotted against the average of all chromosomes. Green colored markers depict a bandwidth of 5 MHz, while blue represents 20 MHz wide OFDM symbols. Figure 4.6 illustrates the evolution of the fitness score. The left part of the figure shows the scores after five iteration steps. One can clearly see the advantage of the best chromosome over the population average. The bandwidth does not have an influence on the achieved score. The right half of the plot shows the performance of the GA over the iterations carried out at $SNR = 11$dB. There are very large score increases after the first few iterations. Afterwards the score increases relatively slowly and starts to saturate. The simulations showed that carrying out larger number of iterations will not improve the result drastically and GA convergence very fast to a close to optimum state.

Figure 4.7 shows the evolution of the modulation order used by the subcarriers. The markers represent the sum of subcarriers on a particular modulation order. As the number of subcarriers is 100, at each iteration step, the different sums represented by the markers should sum up to 100. The resulting evolution of the modulation order is a consequence of w_{tp}, the weight of the throughput objective. Hence, evolution is directed towards high modulation orders in order

[2] The default parameters are indicated in Table A.3 in the Appendix.

to increase the throughput. In the right part of Figure 4.7, at 35 dB, this becomes very obvious: Right after the first iteration 64QAM is in favour of the GA and yet after the second iteration almost all subcarriers use 64QAM. For low $SNR = 11\,dB$, depicted in the left half of the figure, the GA converges to lower modulation orders, as w_{SNR} lowers the per-subcarrier-scores for 64QAM.

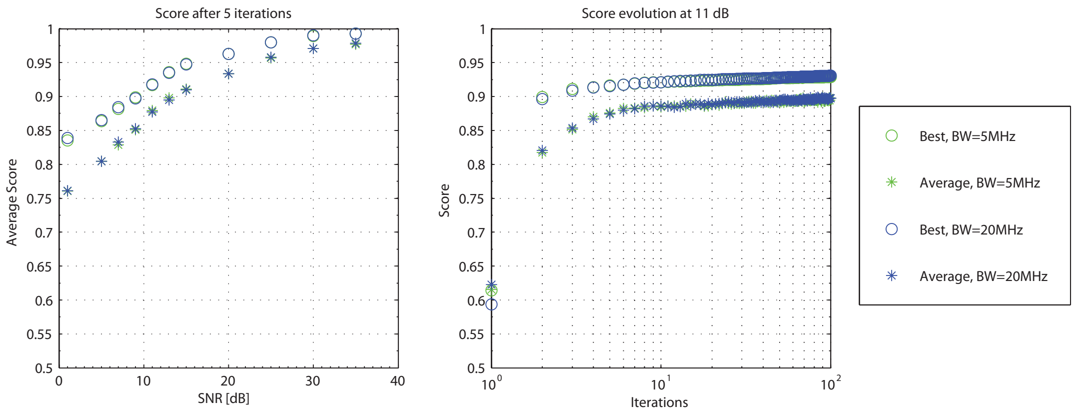

FIGURE 4.6: Evolution of the average fitness score. Increase with SNR (left) and with performed iterations (right).

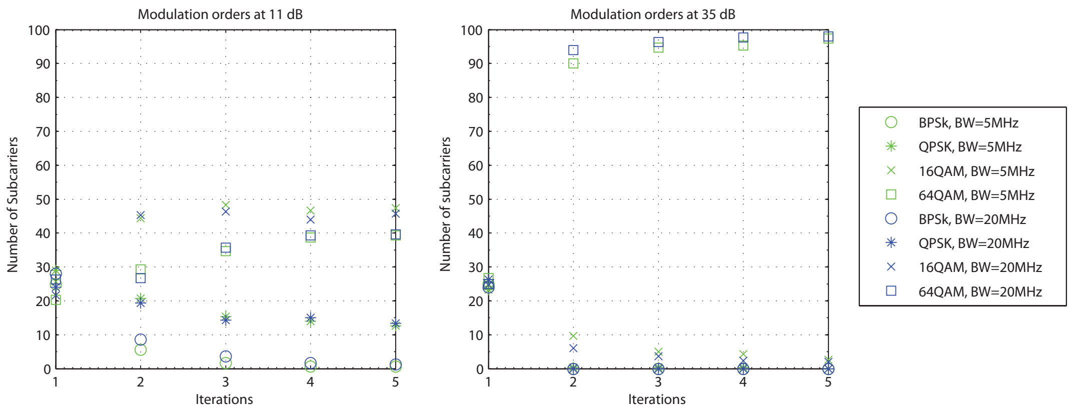

FIGURE 4.7: Evolution of modulation order at 11 dB (left) and 35 dB (right).

Generally, a high power consumption lowers the fitness score of a subcarrier. But at the same time, increasing the power level could cause a slight raise in the current SNR value so that a transmission at higher modulation order is possible. If the values of the weights are set as defined in Table A.2, i.e., $w_{\text{SNR}} = 80\,\%$, $w_{\text{tp}} = 15\,\%$ and $w_{\text{power}} = 5\,\%$, the optimization targets will strongly drive towards a solution where higher modulation order wins against lower power consumption. This can be observed in Figure 4.8, which depicts the average power level of a chromosome. In the left part, at relative low SNR, the GA increases the power level of the subcarriers in favour of the higher modulation orders. In the same way, in the right part of the figure, at very good conditions, it reduces the power consumption for a further increase of the fitness score.

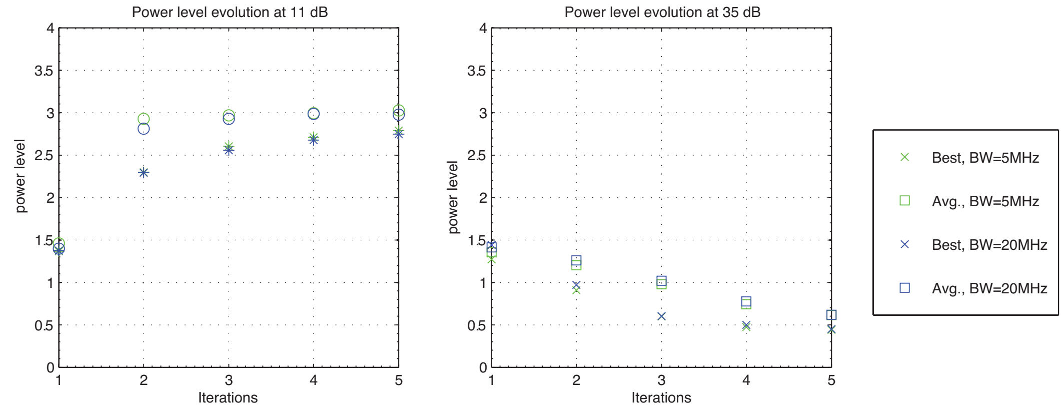

FIGURE 4.8: Evolution of the average power level at 11 dB (left) and 35 dB (right).

The BER performance of the default settings is shown in Figure 4.9. One can see that the best chromosome outperforms the population average, regardless of the bandwidth. As a reference, the solid lines in the plot represent the theoretical performance of the four modulation orders over an AWGN channel. The high BER results, especially at high SNR values, are due to the low power levels selected by the GA, while the theoretical values assume a transmit power of 1 W.

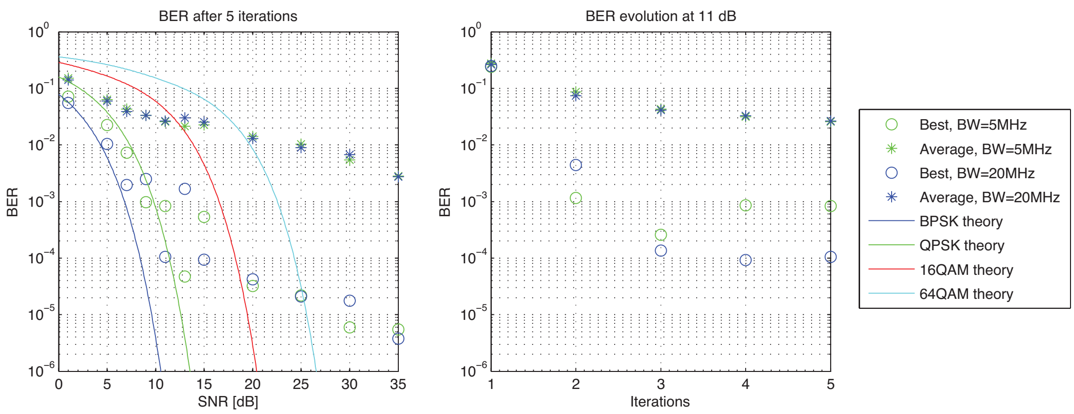

FIGURE 4.9: BER after five iterations (left) and evolution of BER (right).

In addition to the performance analysis of GA as an multi-objective optimization tool for automated parameter setting in our OFDM system, we also analyzed the influence of the GA parameters such as mutation rate, population size, number of iterations and score threshold on the solutions the GA algorithm delivers.

According the extensive simulation analysis performed in various radio environments, i.e., AWGN, Flat fading and Rice channels and different values of SNR we can conclude the following:

1. The influence of different mutation rates hardly exceeds 5 %. The mutation rate has a higher effect at lower SNR values than at good channel conditions. For all channels and bandwidths, the higher mutation rates perform better, but even for 50 % mutation rate, this is only a small advantage (see Figure 4.10). Hence, it is not worth spending additional computational power on randomly selecting bits for mutation and performing bit flipping.

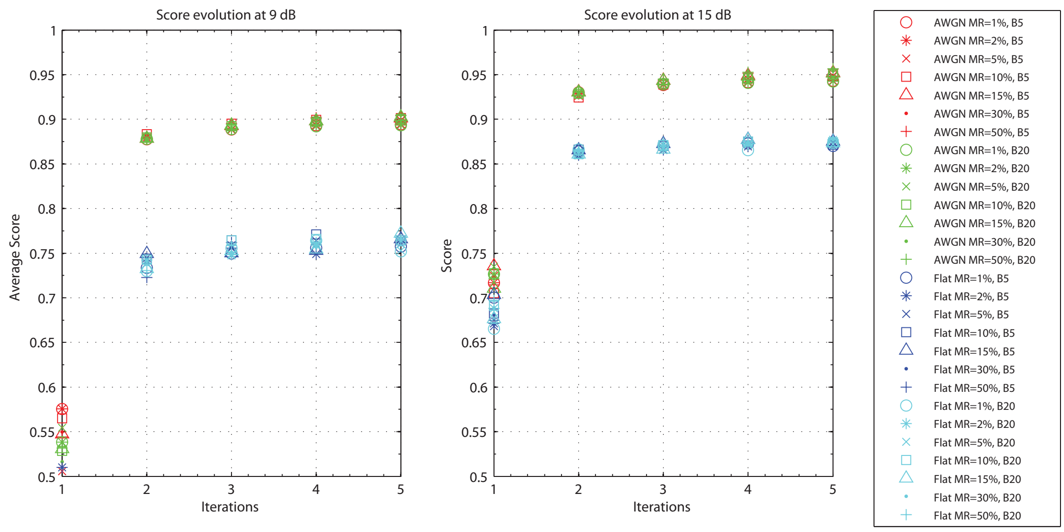

Figure 4.10: Mutation rate sweep at 9 dB (left) and 15 dB (right) over AWGN and Flat Fading channels.

2. As each evolutionary step increases the fitness score, the longer a GA runs, the better its result gets. However, our results show that the biggest advantages are at SNR levels below 15 dB. The advantages of 10 or 15 iterations over the default five iterations are at about three to four percent. On the other hand, these improvements cost a lot of transmission capacity. With our standard population size of ten chromosomes, we have to transmit 50 chromosomes more if we run 10 iterations, and 100 chromosomes more for 15 iterations. This capacity can be better used for data transmissions as the performance increase is insignificant in this case.

3. Chromosomes, performing worse than the score threshold are not considered for mating, and are discarded during the selection process of the GA. However leaving out the score threshold can improve the fitness score of the best chromosome by up to 10 %. These advantages are achieved at low SNRs where commonly most of the chromosomes within a population reach only

low fitness scores. With score threshold enabled, these chromosomes would be directly discarded and many of the population elements of the next generation have to be created randomly. By lowering or completely disabling the score threshold, more chromosomes remain in the pool for mating and can pass their genetic information to the next generation. Passing the genetic information in the next generation performs obviously better than randomly generated offsprings. This is illustrated in Figure 4.11 for a Rayleigh channel.

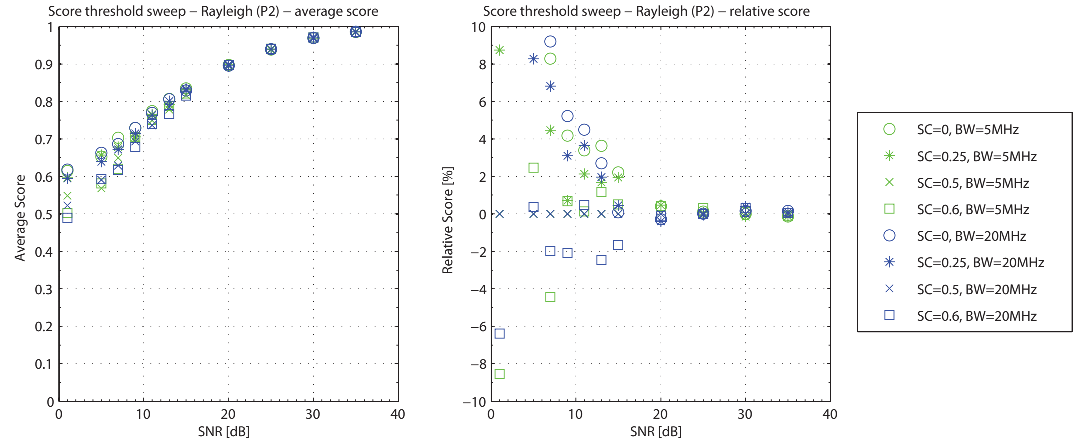

FIGURE 4.11: Score threshold sweep: Rayleigh Fading channel, path length = 50 m

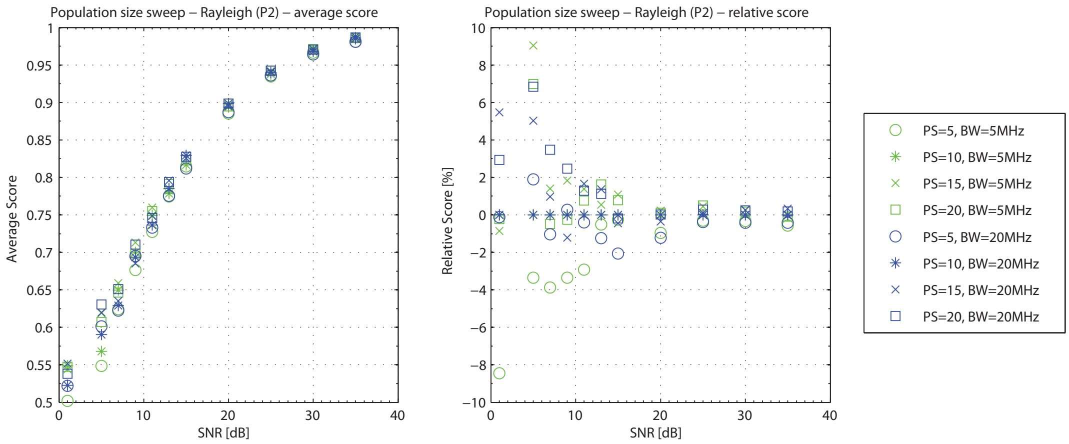

FIGURE 4.12: Population size sweep: Rayleigh Fading channel, path length = 50 m

4. Similarly to the number of GA iterations, the fitness score scales with the population size. Due to the great variety of genes in large populations, it

is more likely that one of the chromosomes carries already a close to optimum solution. This is illustrated in Figure 4.12 for a Rayleigh Fading channel. However, slight increase of performance as a result of larger populations does not legitimate the delay it takes to transmit all those additional chromosomes.

It should be noted that although GAs can converge quite fast to different equilibrium solutions those are not (usually) suitable for very rapid adaptations such as optimizations agains fast fading. In the following section we will extend the analysis of deploying genetic algorithms as an optimization tool to a cross-layer optimization case.

4.4 Cross-layer optimization approach using GA

As discussed in the previous chapter one of the key ideas behind the CRM framework is to facilitate interplay across layers and enable and coordinate cross-layer network optimizations. In this section we will discuss genetic algorithm based method for performing cross-layer optimization including PHY and MAC layer. The particular work has been earlier published inl [21]. We focus, without loss of generality, on wireless OFDM system with MAC layer that is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol as for example in IEEE 802.11 standard [118] but our approach can be also extended to other wireless systems. We propose an Automatic Repeat reQuest (ARQ)-based protocol for cognitive radio system that controls the transmission QoS in terms of delay, throughput, packet loss rate and transmission power consumption along the lines of [119]. Particularly, we show that all optimal transmission parameters can be determined through our GA implementation with the acknowledgement signalling (ACK or NAK) of the prior transmitted packets as the only external input. No additional network state information such as the propagation channel transfer function or the number of active users in the network, is needed at the transmitter. Moreover no transmission model is required for the optimization process. This result is of great importance since most of current transmission protocols use ACK signalling. No additional overhead is required by our approach. Assuming channel assignments have been determined for each user, our algorithm selects the best possible quality of service level as the availability of resources varies. Finally, our GA implementation optimizes at once the parameters of the physical and link layers in a cross-layer fashion.

4.4.1 The parameters

We are going to distinguish two types of parameters, namely *environmental* and *control* parameters. Environmental parameters characterize the surrounding radio environment and are useful input for the cognitive radio to make configuration and optimization decisions. Typical environmental parameters include BER, SNR, noise level, etc. Conventionally, these parameters are measured or estimated at the receiver side and using different feedback mechanisms are brought to the transmitter. Obviously, this generates additional communication overhead, which penalizes the goodput of the system. In our genetic algorithm we make use of the ACK signal as the only environmental parameter from the receiver side. Since most of the communication protocols include ACK control signaling, our approach is compatible with most of standards and does not require any further modification. In case there is no ACK signal in the protocol, our approach can alternatively use an estimate of the transmission packet error probability which can be provided for instance at the output of the error-correcting decoder at the physical layer. In our protocol, we assume the knowledge of the following set of parameters at the transmitter:

1. ACK signal (positive ACK or negative NAK),

2. Number of occurrences that a packet transmission has been successful. With τ we denote the ratio between this number divided by the total transmission time in term of time slot; the parameter τ can be evaluated with a basic counter which is incremented each time a positive ACK is received. In order to accurately evaluate τ for the current network state, we assume that the counter is periodically reset to zero,

3. Number of occurrences that a packet transmission has been unsuccessful. We denote τ' as the ratio between this number divided by the total transmission time in term of time slot. Note that τ' can be determined directly from τ.

If additional information about the current network status, e.g., the propagation channel impulse response(s) or the current number of active users in the network is available at the transmitter, GA can employ this information to ameliorate the optimization process of the transmission parameters.

Control parameters can be tuned in the radio device in an optimal way once an optimal decision has been carried out by the GA. These are used as a decision variables in the utility function in the optimization process. Defining a complete list of decision variables to generate a generic fitness function usable by all radios is a challenging task. We selected a set of decision variables at the physical and MAC layers, which can be adjusted in order of milliseconds or seconds to adapt to the changing channel environment. for most of the cognitive radios. The decision parameters used as outputs in our GA implementation are shown in Table 4.3.

4.4.2 Multiple-objective fitness functions

We define four objectives for our system: reliability, defined through the packet error rate, power consumption, throughput and delay bound. Below, we will derive the corresponding objective functions used by GA in order to lead the

Table 4.3: List of the transmission parameters in our GA implementation at the physical layer (uncoded OFDM transmission), the link layer (CSMA/CA with exponential backoff) and the network layer.

Link layer	Packet size L Minimum contention window size $\mathrm{CW}_{\min}$ Maximum backoff stage m
Physical Layer	Transmission power per subcarrier P_i, $i = 1, \dots, N_c$ Modulation order per subcarrier M_i, $i = 1, \dots, N_c$

system to an optimal state. In order to facilitate the selection of the weights of the objective functions, we normalize each objective function score to the range $[0, 1]$. The four objective functions in our GA implementation are respectively:

1) **Minimize Power Consumption**, i.e., decrease the amount of transmission power:

$$f^{(\mathrm{PHY})}_{\mathrm{min\_power}} = 1 - \sum_{i=1}^{N_c} \frac{P_i}{N_c P_{\max}} \tag{4.10a}$$

where N_c is the number of subcarriers in our OFDM system, $P_i,\ i = 1, \dots, N_c$ is the transmission power on subcarrier i and $P_{\max}$ is the maximum possible transmission power for a single subcarrier. P_i can take values within $[P_{\min}, P_{\max}]$ but can also be zero, i.e., no transmission on subcarrier i if the channel fading coefficient in this band is smaller than a pre-determined threshold as in [120]. The function $f_{\mathrm{min\_power}}$ is equal to 1 when no signal is transmitted over any subcarrier and equal to 0 when all subcarriers are transmitting with maximum power $P_{\max}$. At the link layer, the power consumption depends on the protocol. For CSMA/CA protocol with exponential backoff, the total transmission power also includes the power used in packet retransmission. The fitness function at MAC layer can be expressed as follows:

$$f^{(\mathrm{MAC})}_{\mathrm{min\_power}} = 1 - (1 + \tau'/\tau) \cdot \sum_{i=1}^{N_c} \frac{P_i}{N_c P_{\max}}, \tag{4.10b}$$

where τ and τ' are number of occurrences that a packet transmission has been successful and unsuccessful respectively (see Section 4.4.1). Strictly speaking, $f^{(\mathrm{MAC})}_{\mathrm{min\_power}}$ can be negative for large transmission power and high retransmission rate. However, $f^{(\mathrm{MAC})}_{\mathrm{min\_power}}$ is monotonically decreasing function with respect to τ and as we will observe in our simulations, in Section 4.4.5, using Equation (4.10b) in our GA implementation does not deteriorate the performance. Whereas Equation (4.10a) or Equation (4.10b) penalizes system states with higher power consumption, it might be not enough to guarantee that the current power consumption is equal to or lower than a certain threshold P^*. In order to ensure this, we modify Equation (4.10b) as follows:

$$f^{(\mathrm{MAC})}_{\mathrm{min\_power}} = \begin{cases} 0, & \text{if } \sum_i P_i > P^*, \\ 1, & \text{if } \sum_i P_i \leq P^*. \end{cases} \tag{4.10c}$$

2) **Maximize Throughput**, i.e., increase the overall data throughput transmitted by the radio. At the physical layer, the throughout per user T can be expressed in number of bits per symbol period as $T = \sum_{i=1}^{N_c} \log_2(M_i)/N_c$, where $M_i,\ i = 1, \ldots, N_c$ is the number of bits per symbol emitted on subcarrier i, $M_{\max}$ is the maximum modulation order with typical values 64 or 256 in wireless networks. M_i can take values from 1 to $M_{\max}$ with a special case occurring when subcarrier i is shut down. $M_i = 1$ means that the rate $\log_2(M_i)$ is equal to zero; no information is transmitted. In this particular case, the corresponding transmission power P_i is set to zero. Clearly, we have: $0 \leq T \leq \log_2(M_{\max})$, where the value $\log_2(M_{\max})$ is achieved when all subcarriers are loaded with symbols modulated with the largest available modulation order. Therefore, the objective function for the throughput is simply:

$$f^{(\mathrm{PHY})}_{\mathrm{max\_throughput}} = \frac{1}{N_c \log_2(M_{\max})} \cdot \sum_{i=1}^{N_c} \log_2(M_i). \tag{4.11a}$$

The function $f^{(\mathrm{PHY})}_{\mathrm{max\_throughput}}$ is equal to 1 when all subcarriers transmit with largest modulation order and equal to 0 when all subcarriers are switched off.

At the link layer, the saturation throughput T can be expressed as in [121]

$$T = \frac{\tau \cdot P \sum_{i=1}^{N_c} \log_2(M_i)/N_c}{(1 - \tau - \tau')\sigma + \tau T_s + \tau' T_c},$$

where P is a packet duration and σ denotes a slot duration. We adopt for T_s and T_c the same definitions as in [121], i.e., T_s is the duration between the end of a packet transmission and the reception of the corresponding ACK signal, and T_c is the maximum delay after each packet transmission before to declare that the packet is lost.

An upper bound on the throughput T occurs if the highest modulation order $M_{\max}$ is used for all subcarriers and if all packet transmissions are successful which yields $T_{\max} = (P \cdot \log_2 M_{\max})/T_s$. Therefore, the objective function for the CSMA/CA throughput can be expressed as the ratio between T and $T_{\max}$ in the following way:

$$f^{(\mathrm{MAC})}_{\mathrm{max\_throughput}} = \frac{\tau \cdot T_s \cdot \sum_{i=1}^{N_c} \log_2(M_i)}{[(1 - \tau - \tau')\sigma + \tau T_s + \tau' T_c]\, N_c \log_2(M_{\max})}. \tag{4.11b}$$

Whereas Equation (4.11a) or Equation (4.11b) penalizes system with lower throughput, it might be not enough to guarantee that the current throughput exceeds a certain threshold T^*. As for the transmission power, we therefore modify Equation (4.11b) as follows:

$$f^{(\mathrm{MAC})}_{\mathrm{max\_throughput}} = \begin{cases} 0 & \text{if } T < T^*, \\ 1 & \text{if } T \geq T^*. \end{cases} \tag{4.11c}$$

3) **Minimize Bit/Packet-Error-Rate**, i.e., improve the reliability of the transmission. One possible objective function for characterizing the reliability of the system is:

$$f_{\mathrm{min\_ber}} = 1 - \log(0.5)/\log(\overline{P}_e), \tag{4.12a}$$

where $\overline{P}_e$ is the average bit-wise probability of error per subcarrier. This objective function which has been initially proposed by [13], has two drawbacks in our context. First, the receiver estimates the probability of error of the transmission and forwards a quantized version of it to the transmitter. This additional overhead should be included into the protocol and requires modification to IEEE 802.11 standard. Second, it does not fit well with the usual QoS requirement. QoS usually requires a maximum tolerated bit or packet error probability. Above this threshold, the communication is disrupted. Here we propose two new objective functions for the reliability of the transmission: the first function ensures that packet error probability is equal to or lower than a target PER denoted as PER^*, i.e.,

$$f_{\mathrm{min\_ber}} = \log(\max(\mathrm{PER}^*, \mathrm{PER}))/\log(\mathrm{PER}^*). \tag{4.12b}$$

This objective function penalizes only the sets of decision variables that yield to $\mathrm{PER} > \mathrm{PER}^*$; otherwise, $f_{\mathrm{min\_ber}} = 1$ as long as $\mathrm{PER} \leq \mathrm{PER}^*$. In other words, any set which satisfies $\mathrm{PER} \leq \mathrm{PER}^*$ would be optimal from the PER minimization viewpoint independently if $\mathrm{PER} = \mathrm{PER}^*$ or $\mathrm{PER} = \mathrm{PER}^*/1000$. The second objective function that we propose here is a binary version of Equation (4.12b), i.e.,

$$f_{\mathrm{min\_ber}} = \begin{cases} 0 & \text{if } \mathrm{PER} > \mathrm{PER}^*, \\ 1 & \text{if } \mathrm{PER} \leq \mathrm{PER}^*. \end{cases} \tag{4.12c}$$

Similarly to Equation (4.12b), Equation (4.12c) also ensures that packet error probability is equal to or lower than PER^*. In addition, it requires at the transmitter the knowledge that the packet has been successfully transmitted or has been lost (collision with other users or transmission error due to the transmission channel distortion). Therefore, the value of Equation (4.12c) can be estimated from the acknowledgement signalling value only.

4) **Minimize Transmission Delay**, i.e., decreasing the time interval between two successful packet transmissions. The objective function for characterizing the delay bound of the system is

$$f^{(\mathrm{PHY})}_{\mathrm{min\_delay}} = \frac{L \cdot \log_2(M_{\mathrm{max}})}{L_{\mathrm{min}} \sum_{i=1}^{N_c} \log_2(M_i)}, \tag{4.13a}$$

where L and $L_{\min}$ are the current and minimum packet lengths, respectively. At the link layer, packet retransmissions have to be taken into account. The objective function becomes

$$f^{(\mathrm{MAC})}_{\mathrm{min\_delay}} = \frac{L \cdot \log_2(M_{\max})\lceil 1/\tau \rceil}{L_{\min} \sum_{i=1}^{N_c} \log_2(M_i)}. \tag{4.13b}$$

4.4.3 The weighted sum approach

Similarly to our previous simulations, here we use a simple weighted sum approach to compose the optimization utility function. The weighted sum approach attempts to maximize the sum of the positively normalized, weighted, single objective scores of the parameter set solution $x = [P_1, P_2, \ldots, P_{N_c}, M_1, M_2, \ldots, M_{N_c}, CW_{\min}, CW_{\max}, L]$:

$$\begin{aligned} f(\mathbf{x}) &= w_1 f_{\mathrm{min\_power}}(\mathbf{x}) + w_2 f_{\mathrm{max\_throughput}}(\mathbf{x}) \\ &+ w_3 f_{\mathrm{min\_ber}}(\mathbf{x}) + w_4 f_{\mathrm{min\_delay}}(\mathbf{x}). \end{aligned} \tag{4.14}$$

This method suits the cognitive radio scenario well since it provides a convenient process for applying weights to the objectives. When the weighting for each objective is constant, the search direction of the evolutionary algorithm is fixed. This is the intended property when trying to find a single optimal solution for a given environment. Changing the objective direction of the fitness function requires only a simple change of the weighting vector. The problem is that the direction is not necessarily known in advance for a given QoS. For example, assume that PER$\approx 10^{-1}$ with current settings and that the target PER$^* = 10^{-3}$. In order to satisfy the PER requirement, it seems obvious to increase the weight related to Equation (4.12) but the problem is to find the incremental value. In addition, if the weight related to the maximization of the throughput is too dominant, the QoS in term of PER is satisfied but at the expense of the other objectives. In this example, the throughput will be too low or the power consumption too high. A basic strategy would be to update the weights iteratively until a solution close to the requirements is reached. However, the convergence to the optimal set of weights may be (very) slow and the approach thus can become highly inefficient.

The strategy that we adopt in our GA implementation exploits the discrepancy of the solutions in Equation (4.10c), Euation (4.11c) and Equation (4.12c). Indeed, those objective functions may take only binary values, 0 or 1, so whatever the weights are, the overall fitness function score in Equation (4.14) is very likely low if one or several objective function scores are equal to null. In Section 4.4.5, we validate this approach by means of simulations.

4.4.4 Genetic algorithm with acknowledgement signalling

The optimization problem defined in Equation (4.14) involves non-linear functions. Additionally, this implies that it is not possible to treat each parameter as an independent variable which can be solved in isolation from the other variables. There are interactions such that the combined effects of the parameters must be considered in order to maximize or minimize the solution set. Genetic algorithm is suitable to solve that kind of optimization problem [122]. We assume that the variables representing the set of parameters $\{P_1, P_2, \ldots, P_{N_c}, M_1, M_2, \ldots, M_{N_c}, CW_{\min}, CW_{\max}, L\}$ can be typifying by bit strings. This means that the variables are quantized *a priori* and that the range of the quantization corresponds to some power of 2.

The first step in GA is to generate a *single* initial random bit string representing a possible solution $\mathbf{x} = [P_1, P_2, \ldots, P_{N_c}, M_1, M_2, \ldots, M_{N_c}, CW_{\min}, CW_{\max}, L]$ to the optimization problem (4.14). A first payload packet is transmitted with respect to these parameters. After receiving in return the acknowledge signal (positive ACK or negative NAK), the string is then evaluated and assigned the fitness value $f(\mathbf{x})$ given by Equation (4.14). If optimization at the physical layer is considered, the objective functions to evaluate Equation (4.14) are Equation (4.10a) or Equation (4.10c), Equation (4.11a) or Equation (4.11c) and Equation (4.12c) depending on the QoS requirements. If cross-layer optimization is considered, then Equation (4.14) is evaluated with Equation (4.10b) or Equation (4.10c), Equation (4.11b) or Equation (4.11c) and Equation (4.12c). The principle of the optimization procedure employing genetic algorithm is depicted in Figure 4.13.

A new random bit string representing another possible solution to the optimization problem in Equation (4.14) is used for the second packet transmission. Based on the value of the ACK signal for this packet, this string is evaluated and assigned the fitness value $f(\mathbf{x})$ given by Equation (4.14). The population after two packet transmissions is 2. For the next packet transmission, this process repeats. Hence, the population grows linearly with the number of transmitted packets independently if the transmission fails or succeeds until it reaches a maximal value N. Then, selection is applied to the current population of N strings to create an intermediate population. Then recombination and mutation are applied to the intermediate population to create the next population also of N strings. The process of going from the current population to the next population constitutes one generation in the execution of a genetic algorithm and is performed after each new set of N transmitted packets. Selection process that will more closely match the expected fitness values is "remainder stochastic sampling". As discussed earlier, there are several ways to do selection. An efficient implementation described in [122] uses a method known as "Stochastic Universal Sampling". Assume that the population is laid out in random order as in a pie graph where each individual is assigned space on the pie graph in proportion to fitness. Next an outer roulette wheel is placed around the pie with N equally spaced pointers. A single spin of the roulette wheel will now simultaneously pick all N members of the intermediate population. The resulting selection is also unbiased. After selection has been carried out the construction of the intermediate population is complete and recombination can occur. This can be viewed as creating the next

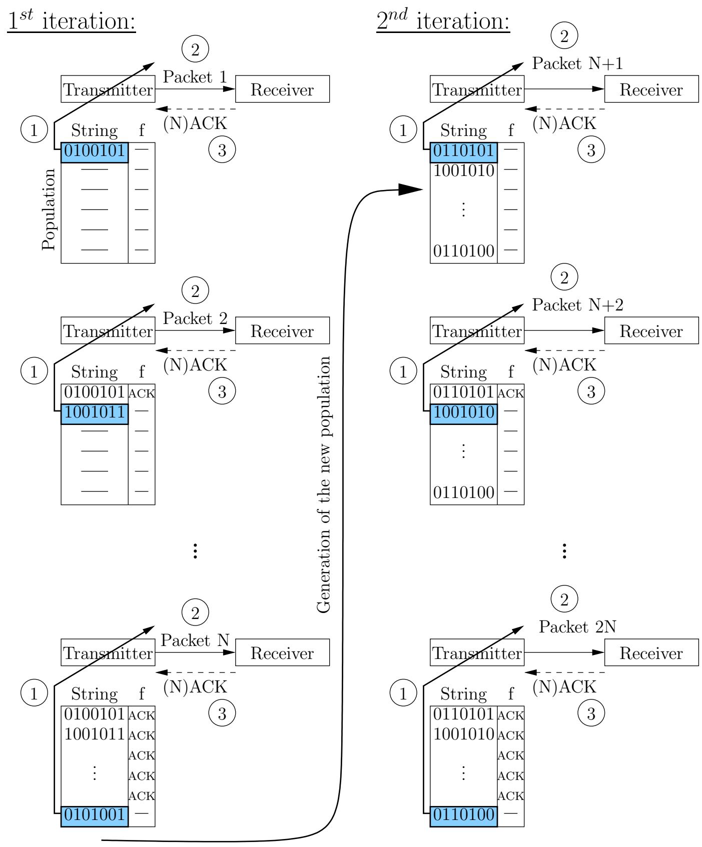

FIGURE 4.13: GA Principle based on acknowledgement signalling. At the first iteration, (1): generate a random string and update the transmission parameters accordingly; (2): transmit packet 1; (3): update the fitness function f based on the received ACK; (1): generate a new random string and update the transmission parameters accordingly; (2): transmit packet 2; (3): update the fitness function f based on the received ACK. Continue the same procedure until transmission of packet N. Then apply GA to generate a new population (a better one). This completes the first iteration. At the second iteration, (1): use the new string 1 to updating the transmission parameters; (2): transmit packet $N + 1$; (3): update the fitness function f based on the received ACK. Repeat the procedure for each next set of N packets.

population from the intermediate population. Elitism of 10% is considered. For generating the other strings of the new population, crossover with single recombination point is applied to randomly paired strings with probability $p_c = 0.6$.

After recombination, we apply a mutation operator. For each bit in the population, mutate, i.e., flip the bit `x` to `1-x` with probability $p_m = 1 - 1.8^{-\frac{1}{N}}$, where N is the size of the population as proposed in [123]. After the process of recombination and mutation is complete for the selected N strings, the new population is re-evaluated through the transmission of the N next packets. The process of evaluation, selection, recombination and mutation forms one iteration in the execution of a genetic algorithm. We iterate until convergence to a stable solution for the set of parameters $\mathbf{x}$.

4.4.5 Simulation results

In this section, we characterize the performance of the proposed Genetic Algorithm for ARQ-based link adaptation for multicarrier transmission in various scenarios. In all cases, we simulate a multicarrier system with $N_c = 64$ subcarriers using Matlab simulator. Sufficient cyclic prefix is assumed. Each subcarrier is assigned a random attenuation value $|H_i|^2$, $i = 1, 2, \ldots, N_c$ with chi-square distribution. Hence, the signal-to-noise ratio (SNR) varies independently from one subcarrier to another and induces a need for the power and rate adaptation for each individual subcarrier. The channel was assumed to be "block-invariant", implying that the transmission channel impulse response remains constant or undergoes only minor changes over several consecutive packet transmissions. We assume regular Quadrature Amplitude Modulation (QAM) signaling (4-QAM, 16-QAM and 64-QAM) but our approach can readily be extended to arbitrary modulations. We also permit to switch off some subcarriers if the fading is too deep for the corresponding bands. The transmission power P_i can take 16 values ranged uniformly from 0.1 mW to 2.56 mW. These are example values, of course, and do not represent any limitation for our GA based approach. At the link layer, adaptive contention window size is considered as suggested in [121]. We assume no RTS/CTS mechanism. The minimum contention window size $CW_{\min}$ can take four possible values comprised between 4 and 32. The maximal contention window size $CW_{\max}$ can take 8 values comprised between 32 and 4096. We also consider 8 different packet sizes, L from 18 bytes to 2304 bytes.

Overall, with 16 possible values for the transmission power, 4 possible modulation indexes, this gives 16×4 possible values for each subcarrier. With 4 (respectively 8) possible minimal (respectively maximal) contention window sizes, 8 different packet sizes, and 64 subcarriers, this gives a total search space of $64 \times 16 \times 4 \times 4 \times 8 \times 8 = 1,048,576$.

4.4.6 Scenario 1: ARQ-based Discrete Waterfilling Algorithm

In the first example, we focus on the transmission parameters optimization for the physical layer. Whereas next examples will demonstrate the importance of cross-layer optimization, this example permits us to compare performance of our ARQ-based genetic algorithm against the performance obtained with optimal bit-loading algorithm. GA is compared to the bit loading algorithm proposed in [124], which is near-optimal at moderate computational complexity. It serves us as a benchmark for this example but also for all the other examples of this section. Additionally, we compare our GA algorithm against the solution provided in [13], which is also based on genetic algorithm. The main difference resides in that the fitness function given by Equation (4.14) is evaluated by using Equation (4.12a) in [13] instead of Equation (4.12c) in our case. Also the weights are different. The only way to meet the QoS requirement, say target PER* using Equation (4.12a) is to adapt the weights of the objective functions.

In [13] the authors proposed several sets of weights that advantage either the system reliability or the power consumption or the throughput. In our simulation, we choose their most advantageous weight set, i.e, the set that satisfies the QoS constraint while providing the highest throughput, also known as a multimedia mode ($w_1 = 0.05, w_2 = 0.8, w_3 = 0.15$). The final comparison in this example is carried out with the adaptive modulation scheme used in IEEE 802.11 standard, which we call a conventional algorithm. In this scheme, the modulation order is identical for all subcarriers. The highest order is chosen such that it satisfies the average target PER*.

Figure 4.14 shows the throughput performance achieved by the four considered algorithms as function of SNR. In this example, target PER* is equal to 10^{-3}.

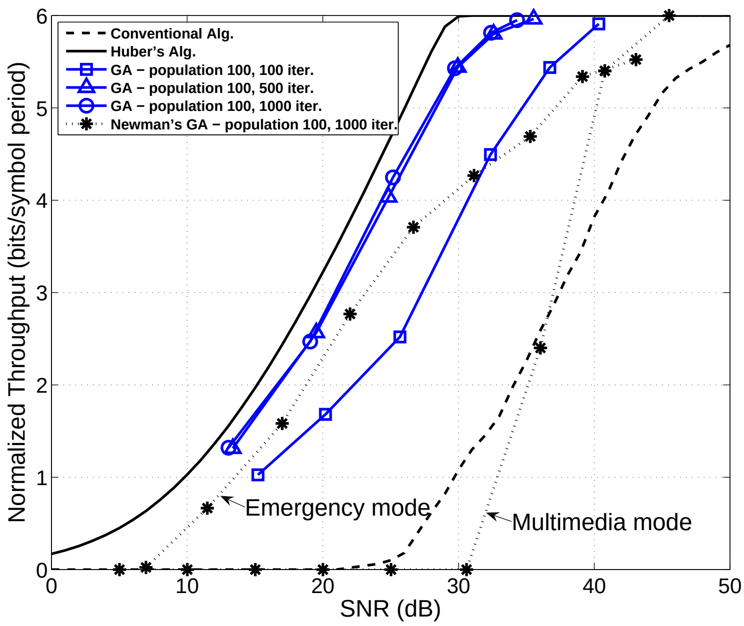

FIGURE 4.14: Joint Power Optimization / Bit-loading Algorithm for OFDM system. Target PER* $= 10^{-3}$. Conventional algorithm performs adaptive rate as in the IEEE 802.11 standard; Algorithm proposed in [124] is a near optimal power allocation/bit-loading algorithm with computational complexity order of $N_c \log N_c$; we use this algorithm as a benchmark for our GA-based algorithm; Newman'algorithm denotes the GA-based bit-loading algorithm proposed by [13].

Three conclusions may be made. First, GA performs very close to the optimal bit-loading algorithm as long as the number of iterations in GA is large enough.

The gap between GA with 100 iterations and bit-loading algorithm is approximatively 8 dB at all SNR values. The gap is reduced to 1.5 dB if GA performs 500 iterations. This loss is mainly due to the high percentage rate for the elitism in GA. Whereas elitism of 10% dramatically increases the convergence speed of GA, it penalizes the search toward the global optimal solution. Duplicating the best but still suboptimal solutions among the population in GA may prevent to find a new better solution. Second, our GA implementation outperforms the solution in [13].

The loss is mainly due to the fact that GA implementation in [13] uses the objective function defined by Equation (4.12a) rather than Equation (4.12c). Indeed, three sets of weight were proposed: i) the multimedia mode which favours the throughput at the expense of the power consumption and the reliability, ii) the low-power mode which lower the power consumption ($w_1 = 0.8, w_2 = 0.15, w_3 = 0.05$) and iii) the emergency mode which favours the reliability at the expense of the throughput and power consumption ($w_1 = 0.15, w_2 = 0.05, w_3 = 0.8$). As expected, the set corresponding to the multimedia mode (high throughput, lower reliability) leads to a better solution than the emergency mode set (high reliability, lower throughput) at high SNR. At low SNR, the emergency mode set performs better than the multimedia mode set. The operational SNR range is rather small for both sets, that is if the SNR is lower than 30 dB, the multimedia mode cannot find any solution that satisfies PER*. On the other hand, for SNR larger than 30 dB, the multimedia mode provides a solution with much better PER than PER* at the expense of the throughput. Finally, the conventional approach is penalized by the fact of using the same modulation order over all subcarriers. In this case, performance are dictated by the transmission error over the subcarrier(s) with the deepest fading.

4.4.7 Scenario 2: ARQ-based Cross-Layer Optimization with Adaptive Contention Window Size

Figure 4.15 shows the throughput performance achieved by GA for cross-layer optimization. In addition to the parameters of the physical layer P_i and M_i, where $i = 1, \ldots, N_c$, the minimal contention window size and maximum exponential backoff stage are also considered. Optimization over all these parameters with respect to Equation (4.14) requires one to find the optimal trade-off between conflicting entities: Maximizing the minimal contention window size reduces the number of retransmission per packet and therefore the power consumption. However, it does not necessarily increase the throughput as shown by Bianchi in [121] for large network load. In this example, the target packet error rate PER* is set to 10^{-3} and a network with 10 or 25 users is considered. We compare against the conventional scheme with pre-determined initial contention window size $CW_{\min} = 32$ and maximum backoff stage $m = 5$. For the conventional scheme, the modulation order is determined as in Scenario 1.

We also compare against the scheme referred as “PHY then MAC” which consists to separately optimize the parameters of the physical layer and the link layer. The parameters of the physical layer are optimized with algorithm proposed in [124] and the parameters at the link layer are optimized through an exhaustive search of all possibilities of $CW_{\min}$ and $CW_{\max}$.

We can make two conclusions. First, GA with optimal $CW_{\min}$ and $CW_{\max}$ outperforms the conventional scheme by approximatively 20 dB. The loss essentially occurs at the physical layer. Indeed, the gap between both schemes in Scenario 1 was already around 20 dB. The fact of using fixed contention window size $CW_{\min} = 32$ and maximum backoff stage $m = 5$ seems to have little effect on the performance. However, in different scenarios, it might not be the case. Second, GA outperforms also "PHY then MAC" approach at SNR larger than 15 dB. Although the bit-loading algorithm performs slightly better than GA at the physical layer (Scenario 1), the "PHY then MAC" approach optimizes the minimum and maximum contention window sizes based on target PER* and not the current PER value. Whereas the loss is negligible at lower SNR values (SNR $<$ 15 dB), it becomes significant for SNR values greater than or equal to 15 dB. Indeed, PER is significantly smaller than PER* and considering PER* instead of PER leads to non-optimal contention window sizes.

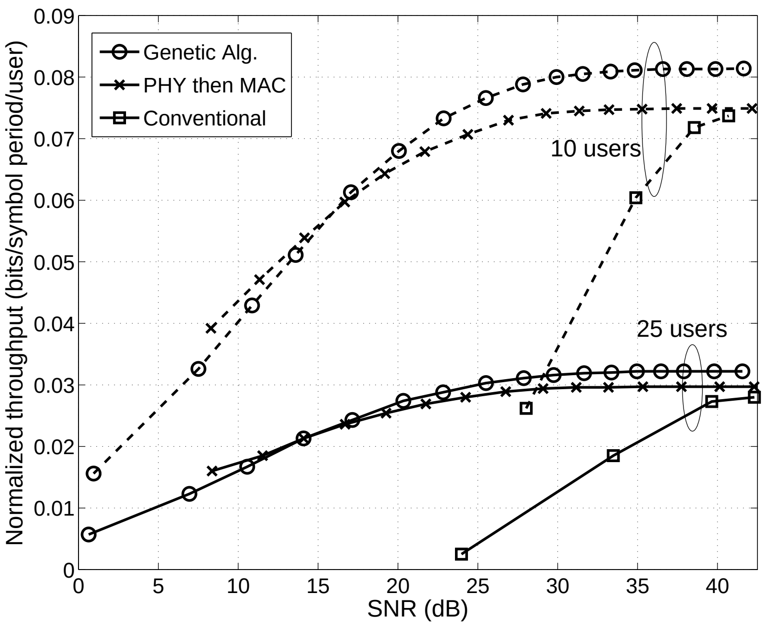

Figure 4.15: Cross-Layer Optimization for OFDM-based CSMA/CA system: Joint Power Optimization / Bit-loading Algorithm / Minimum Contention Window Size / Maximum Backoff Stage. Target PER* $= 10^{-3}$.

4.4.8 Scenario 3: ARQ-based Cross-Layer Optimization with QoS requirements

In the four previous scenarios, we use GA to maximize the throughput at given transmission power consumption and target PER*. In this section, we propose to evaluate GA performance with other QoS requirements. We start with transmission power consumption. Whereas we display GA performance after convergence in the previous scenarios, it might also be interesting to check the tracking properties of our GA implementation.

Case 1: Power consumption. Let us consider the following scenario: for the first 3000 time slots (0.15 seconds), full power is used (2.50 mW), then the transmitter detects that the batteries are half-empty and power controller lowers the power consumption to 1 mW. After 6000 time slots (0.30 seconds), the system is asked to switch to the minimum power consumption 0.15 mW.

In order to satisfy the power consumption requirements, we use the objective function defined with Equation (4.10c) instead of Equation (4.10b). In this case, the values of the weights are as follows: $w_1 = 0.1, w_2 = 0.1, w_3 = 0.8$. The results are plotted in Figure 4.16.

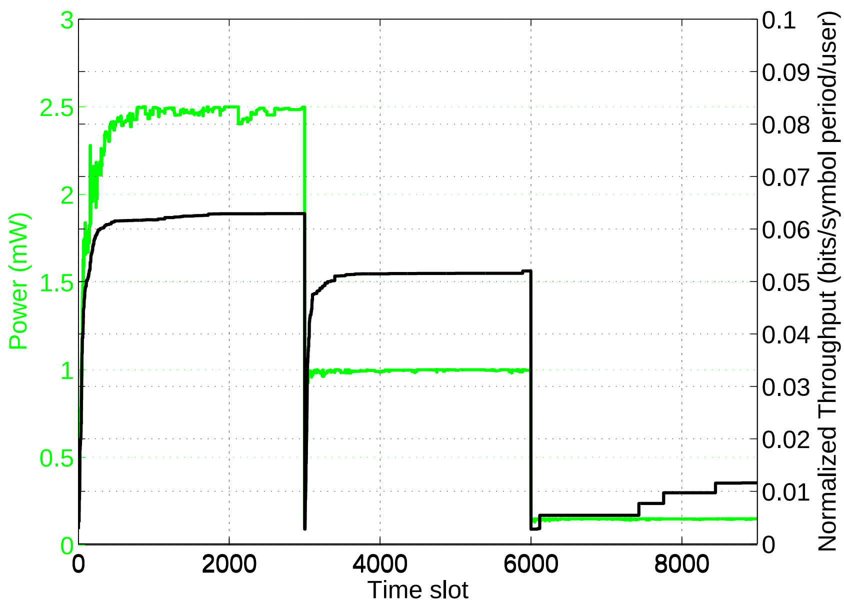

Figure 4.16: Power adaptation in OFDM-based CSMA/CA system with decreasing power constraints: Start with full power utilization (Target POWER = 2.50 mW) for the first 3000 time slots, then switch to 1 mW mode for the next 3K time slots, then lower power (0.15 mW). Target PER* $= 10^{-3}$. Network load = 10.

The light green curve shows the transmission power. As we can see, as soon as the power constraint changes, the optimal solution of our GA implementation satisfies the power constraints within few time slot periods. Moreover, the throughput (black curve) is maximized in tens or hundreds time slots. For very low power constraint, say, 0.15 mW, GA struggles to maximize the throughput and needs more than 1000 iterations to reach optimal throughput.

Case 2: Delay bound for data streaming (video or audio). Another important parameter in wireless networks is the transmission latency which is crucial for real-time audio and video streaming applications. In Figure 4.17, we display the delay bound as a function of the number of active users such that all their transmissions experience a delay less than or equal to this delay bound. Target PER* is 2% and the target throughput T* is set to 640 Kbps, 1.28 Mbps or 2.56 Mbps. The weights for GA are set as follows: $w_1 = w_2 = w_3 = 0.1, w_4 = 0.7$. Our GA implementation almost matches performance of the "PHY then MAC" approach. Basically, it means that GA selected the smallest packet sizes such that the throughput is greater or equal to T^*. Moreover, the delay bound with GA is twice as small as the delay bound for the conventional scheme. For large load, the delay reduction is even more important.

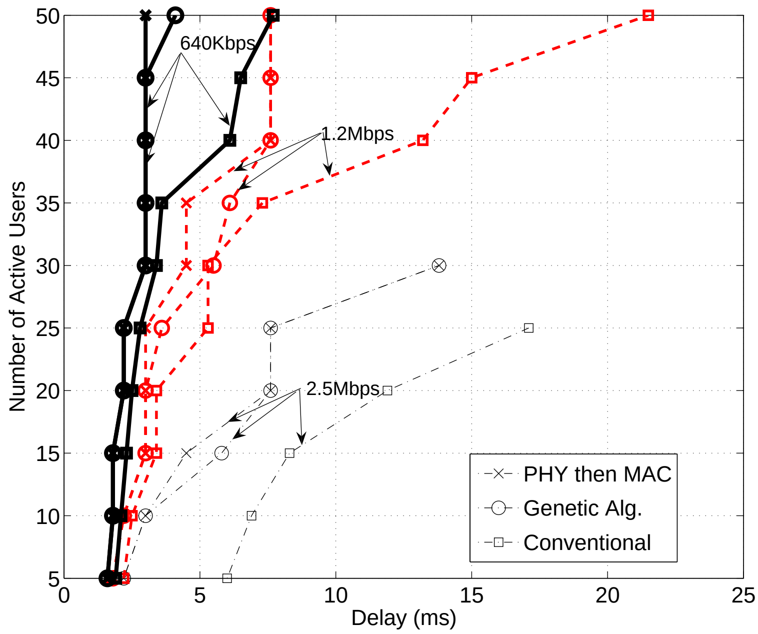

Figure 4.17: Maximum number of active users in the network with respect to the delay bound. Target packet error rate of 2% in OFDM-based CSMA/CA system for high data rate streaming transmission (video).

4.5 GA for dynamic spectrum access in time-varying jammed spectrum scenario

So far, the discussions in this chapter were focused on using genetic algorithms as tools for a multi-objective optimization. We particularly studied three- and four-objective functions, which optimized PHY and MAC parameters of the radio in order to achieve better QoS. The parameters considered included modulation order, power level per subcarrier, contention window, etc. However, in all the simulation scenarios we assumed that the channel allocation problem is handled and the radio is given the best possible channel to operate. Hence, the frequency channel was not so far included as a parameter in the analysis. As will be discussed in Chapter 5, careful channel allocation and spectrum agility is of great importance for interference mitigation and increased system performance.

In this section, we investigate the capability of GA to select transmission channels with lower interference level. We prove the ability of GA to provide spectrum agility. We do this by considering channel allocation as a standalone problem, which we study through experimental tests, where all functionalities are implemented into low-cost SDR platforms. The experimental setup comprises two USRP motherboards that are equipped with RFX 400 daughter boards (see description and references in Chapter 2). One daughterboard is used for transmitting the data and the other one is used as a receiver. All components required for the RF/baseband conversion are implemented on the FPGA of the daughterboard. All baseband operations are executed in the GNU Radio software environment which is running on the host computer. The USRP hardware is connected to the host computer via a USB connection. This interface is utilized for programming the hardware and for carrying the data stream between the GNU Radio flow graphs and the USRP boards. We use the same structure for all packets, which consists of three parts: 1) a preamble for synchronization, 2) the payload length, and 3) the payload itself. The preamble is used by the receiver to detect the beginning of a packet. It consists of a 64-bits pseudo-noise sequence. Our implementation supports variable length packets up to 4096 bytes. The packet length is duplicated in order to mitigate the channel impairments that may occur during the transmission. The modulation order k, used for the current packet is transmitted to the receiver before the packet transmission through a wired Transmission Control Protocol (TCP) control channel. Feedback information from the receiver to the transmitter is also sent through the control channel. In our case the control channel is error-free channel. This is done without losing generality, and the purpose of this is to find out the maximum gain that can be achieved for the data communications.

In the experiment, there were 32, 40 kHz-wide channels as possible candidates for the optimal choice of the GA. The channels are between 428 MHz and 459 MHz, the center frequencies separated by 1 MHz guard intervals.

After L consecutive packets have been transmitted, the objectives in Equation

TABLE 4.4: Internal parameters used in our genetic algorithm implementation.

Parameter	Value
population size	16
number of iterations	11
selection method	tournament
elitism rate	10%
mutation rate μ	$1 - 1.8^{-1/L}$ [123]
crossover rate	90%

(4.4), Equation (4.5) and Equation (4.6) are evaluated for the L corresponding sets of transmission parameters. GA generates L new sets of transmission parameters with better average scores with respect to the overall fitness function in Equation (4.7). The transmission of the next L packets proceeds with these new sets of parameters, which will be evaluated consecutively at the transmitter. This iterative optimization process runs until Equation (4.7) is maximized. Table 4.4 summarizes the set of internal parameters that we use for GA in all our tests.

In our experiments, no channel is initially jammed; all channels have approximately the same background noise and GA iteratively optimizes the transmission parameters. After 48 packet transmissions, which exactly corresponds to 3 full GA iterations, a subset of the available channels is suddenly jammed. For simulating jamming, we set a high resolution, Agilent E4438C, 250 KHz-6 GHz, signal generator to transmit random signal with very high power (8 dBm) in the band of interest such that no transmission is possible. Indeed, the power spectral density in the jammed spectrum is approximately 40 dB higher than the noise power in the "free" channels. In our experiments, we vary the ratio between the number of jammed channels and the number of "free" channels. This ratio is referred as the occupancy level of the system. Our analysis includes occupancy levels of 10%, 30%, 50%, 70%, and 90%. The last case, in which 90% of the channels are jammed is the most challenging optimization problem because only 3 out of 32 channels ensure a reliable transmission. In order to steer the GA population to the solutions that correspond to free channels, we stress the transmission reliability at the expense of the transmit power and the throughput by setting the weights to $w_{BER} = 0.8$ and $w_{power} = w_{tp} = 0.1$. The single fitness f_{BER} is the only objective which is in function of the channel conditions. We thus assure with high probability that the fitness score of the chromosomes that correspond to a high throughput or to a low transmit power level but to a jammed channel is lower than the score of the chromosomes that correspond to free channels. For sake of simplicity, we only use DBPSK and DQPSK modulation for transmission. The transmission power can take different 16 levels normalized between 100 and 15000.

In order to analyze the convergence behavior of GA when the spectrum is partially jammed, the average fitness value f_{min_BER} is evaluated over all chromosomes. The goal of our GA-based implementation is that all chromosomes in the GA population correspond to some free channels for two reasons. First, all packets that are transmitted over jammed channels are lost and require retransmission, which might significantly reduce the throughput. Second, transmitting over jammed channels interferes with other signals which may come, in fact, from primary users. It is essential that all elements in the GA population that correspond to jammed channels are removed in very few iterations.

The convergence behavior of the GA algorithm is illustrated in Figure 4.18. Since f_{min_BER} is the main indicator of the transmission reliability, the average fitness score of f_{min_BER} is plotted as a function of the number of iterations. We average the scores over 6 runs for each occupancy level. A fitness score of 0.98 is observed before jamming part of the spectrum. During this period, transmission is reliable over all 32 channels because all of them are free. Since the ambient noise is similar for the 32 channels in our experiments, GA optimizes the rate and the transmission power and transmits over an arbitrary channel which changes from one packet to another.

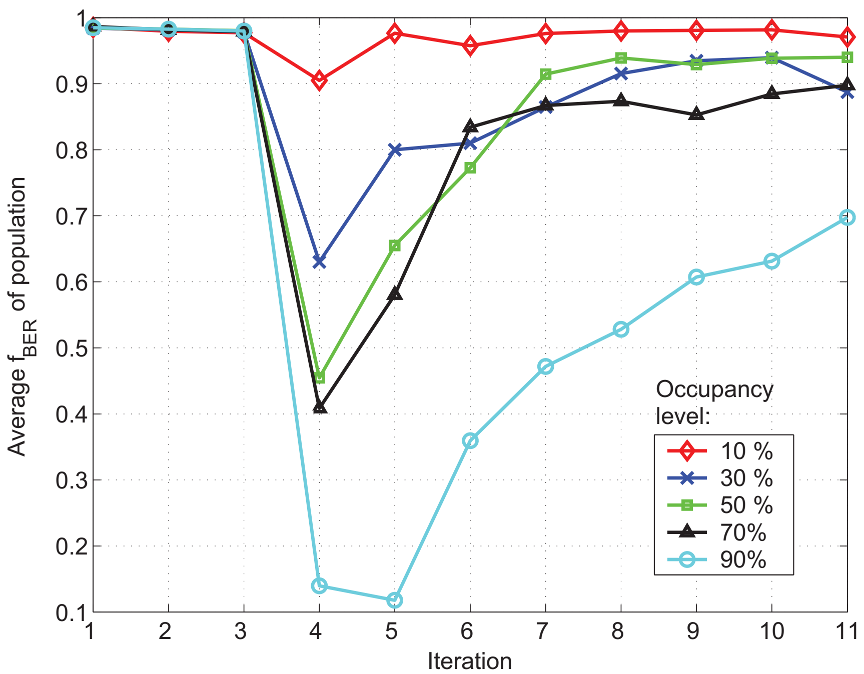

FIGURE 4.18: Channel adaptation for several spectrum occupancy levels. Convergence behavior of f_{min_BER} (4.4) as a function of the number of iterations.

The optimization difficulty rises after iteration 3 when some of the 32 channels remain "free". In order to limit the interference with potential primary users and to guarantee reliable transmission, population of GA has to quickly evolve to free channels. In Figure 4.18, f_{min_BER} drops approximatively to one minus the occupancy level at iteration four. Since the transmission SNRs are similar over all channels, GA population is spread approximatively uniformly over them at iteration 3. Right after jamming a certain ratio of the 32 channels, all packets in the jammed channels are lost and most of the packets in the free channels are successfully transmitted. For example, in the scenario in which 50% of the channels are jammed, the average fitness score drops from 0.9803 down to 0.4546. A similar behavior is observed for the other occupancy cases. In most cases, the average fitness score of f_{min_BER} reaches its initial value (before the jammer) within next few iterations.

When a very large number of channels are jammed (90% occupancy level) GA converges very slowly to the free channels. Before jamming, the channel indices of the whole GA population are concentrated onto a few frequencies instead of being randomly distributed for some runs. Once a good channel is generated, this solution will spread out to the next population with high probability. However, if all channels at iteration 3 are jammed by the signal generator, chromosomes that correspond to free channels can only be introduced through crossover and mutation operations. The probability for this occurrence is rather low with the settings in Table 4.4. Instead of considering constant GA internal parameters in all scenarios, a more sophisticated solution would consist in adapting the internal parameters based on the current occupancy level. However, we used constant rates in this study in order to keep our GA implementation as generic as we could. In Figure 4.19, we illustrate this behavior for two extreme runs as well as the average behavior for several runs for the case when 90% of the channels are jammed. In the first case, the population at iteration 4 contains some chromosomes corresponding to free channels. Most of the solutions correspond to jammed channels, which caused sharp drop from 0.98 to 0.1 as the fitness score at iteration 4. However, the chromosomes corresponding to free channels will quickly propagate during the next iterations due to their high scores.

The second case displays a different behavior. The population at iteration three does not contain any chromosome including free channels. Therefore, the average fitness f_{min_BER} of the population drops to nearly zero at iteration 4. There is no significant improvement until iteration 9. The possibility to generate a chromosome corresponding to a free channel can only occur throughout mutation or crossover operation (and not throughout selection). Up to this occurrence, all packets are lost because their transmission occurs in jammed channels for any chromosome of the GA population. Either mutation or crossover introduced a chromosome corresponding to a free channel at iteration 9. The fitness score improves from 0 in iteration eight to 0.058 after iteration 9. This good solution quickly propagates through the new populations within the next iterations. Ultimately, all solutions that correspond to jammed channels will be removed. Finally, the third curve in Figure 4.19 represents the average fitness score over six runs and includes both aforementioned cases.

Based on our analysis and experiments we can conclude that GA is suitable for efficient dynamic spectrum access. For some extreme cases such as for 90%

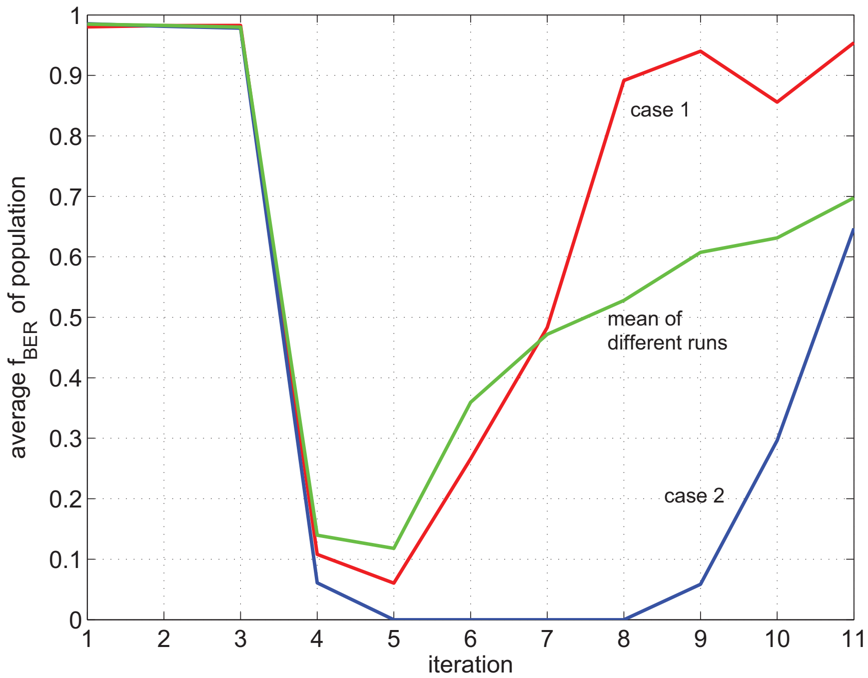

Figure 4.19: Convergence Analysis for 90% jammed channels scenario.

of the channels jammed, GA may converge slowly to free channels. In that case, one solution consists in increasing the mutation rate of GA. Introducing an adaptive mutation rate may overcome this problem. The implementation is however a challenging problem because it requires to include several special cases and thresholds. For implementation complexity reasons, we only considered a constant mutation rate in this paper. It is clear that the channel allocation against jamming based on GA approach may not be always optimal. Energy detectors or feature-detectors combined with simple avoidance algorithms may perform better. The real benefit in CR context is that the jamming avoidance is combined with GA which is performing also other optimizations at the same time.

4.6 Summary

Genetic algorithms (GAs) have a unique niche in the area of multi-objective optimization. Due to the parallel search nature of the algorithms, the approximation of multiple Pareto optimal solutions can be effectively executed. We studied through simulations and experiments the applicability of GA as multi-objective optimization tool. Both PHY and MAC layer parameters were successfully optimized in a cross-layer fashion in order to achieve a better performance of an OFDM system.

Although GAs can be powerful as on-line optimization tools as also shown from our analysis, in the CRM context we see them most of the time as an off-line background computational tools. This means that we could position the GA modules to work in a background to explore *optimization possibilities.* The results could then be fed as an input to our world-models and the actual *dynamic* adaptation is handled, e.g., through decision trees and fuzzy logic controllers. This is due to the fact that latter models provide computationally much faster way to adaptation in practical algorithms.

5

Dynamic Channel Allocation

Frequency is an expensive and limited resource. That is why efficient frequency planning has been an important matter since the very beginning of the wireless systems era. Careful frequency allocation is needed in order to minimize harmful interference, and allow concurrent operation and coexistence of different wireless technologies in a same frequency band.

The spectrum scarcity is a fact for existing ISM-bands and licensed systems. Obviously, frequency and interference management are also critical for future systems such as cognitive radio systems and femtocells. Traditional cellular systems deploy time and frequency division to the available frequency bands and a spatial frequency reuse techniques to serve as many users as possible with a certain quality and reliability. Many commercial and residential Wi-Fi networks still use static channel allocation schemes and very often the access points (AP) are set up to a default channel by the manufacturers themselves. Naturally this creates serious interference problems and service degradation if the number of APs and users operating in the same channel increases. Better channel allocation mechanisms for Wi-Fi networks have potential to increase local capacity and user satisfaction. In this section we describe a specific method developed to solve this problem. We also review the literature in this filed, which is mostly subsequent or contemporary to our work.

There exist static and dynamic channel allocation schemes and they can be realized in distributed or centralized fashion. In this chapter we give a detailed overview of two channel allocation schemes that we have proposed, namely a *dynamic graph coloring channel allocation* and a *distributed load balancing channel allocation* for WLANs and cognitive radios. We show both analytical and real testbed results on the performance of the wireless network using our channel allocation schemes and argue about the advantages and drawbacks of our solutions. Our general goal is to provide solutions that can be used with future cognitive radio systems. However, our focus on WLANs in the beginning of this chapter is motivated by two reasons. First ISM-band operating Wi-Fi networks provide us a useful“toy-model” to quantitatively and sentimentally test our solution. Second, the practical problems with Wi-Fi deployments made the dynamic channel allocation problem an interesting research with direct technology transfer possibilities.

5.1 Channel allocation in wireless networks

Suppressing the interference is one of the major challenge in the wireless communication systems. Deployment of time and space diversity, efficient filtering and equalization, adaptive modulation schemes, etc., are very helpful techniques to fight against interference. Still, co-channel interference, as a result of frequency reuse, and adjacent channel interference can cause severe performance degradation. Hence, efficient channel allocation is critical for the performance of all types of wireless systems. Since the design of the first cellular networks in the 1980es, and the immense success of W-iFi in the beginning of this century, a great effort has been done in developing frequency allocation techniques for interference minimization and increasing the capacity of the wireless systems. With the emergence of cognitive radios and the concept of DSA, the problem of channel allocation gained a new dimension. A new and more complex system setups and increased dynamics come into play, which pose additional constraints in finding the optimal channel acquisition and allocation solutions. The paradigm shift towards opportunistic spectrum access and co-existence of primary and secondary users in the same band poses new challenges to the frequency management in enabling more flexible, efficient and dynamic spectrum reuse and at the same time optimal network performance and user experience.

Traditionally a radio frequency band dedicated to a certain communication technology is divided into set of channels, so that the available band can be used simultaneously while keeping an acceptable SNIR. The very basic techniques for sharing the communication medium by several users are Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA) [125]. In FDMA the available band is divided into subbands so called frequency channels, whereas in TDMA the users time-share a certain frequency band by being assigned one or several time slots. Classical examples of FDMA systems are the first generation cellular systems such as Nippon Telephone and Telegraph (NTT) deployed in 1979 in Japan, Advanced Mobile Phone System (AMPS) introduced in 1983 in the USA, European Total Access Cellular System (ETACS) introduced two years later, etc. The Digital European Cordless Telephone (DECT) and the Global System for Mobile (GSM) are probably the best known TDMA/FDMA systems. In CDMA the channel separation is achieved by using different spreading codes. CDMA is an underlying technology for IS-95, CDMA2000 and Universal Mobile Telecommunications System (UMTS/WCDMA).

In this chapter we are interested in channel allocation techniques, which deal with frequency channels. In particular, we focus on short range radio technology such as Wi-Fi type of networks and future cognitive radio networks. For completeness in the following we will give a comprehensive overview of the existing channel allocation schemes for different wireless systems, outline the most significant solutions and comment shortly on their performance and complexity.

5.1.1 Chancel assignment in cellular systems

Among the fundamental problems in wireless system planning are those regarding channel assignment. The usual questions to address are the following:

- How many frequencies are needed for the system?
- Given a certain number of frequencies, what is the optimal frequency allocation so that a certain number of users get the desired services?
- What is the optimal channel allocation that minimizes the adjacent channel interference?

In a cellular network two cells can use the same channel if only the geographical distance between their centres (base stations) is larger than the minimum reuse distance D_{min} [126]. Hence, the base stations in the neighboring cells are given different set of channels to keep the level of interference low and the number of serviced calls as large as possible [1]. Channel allocation schemes can be categorized in different ways. In cellular networks depending on the manner the co-channels are assigned, three groups can be distinguished: fixed channel allocation (FCA), dynamic channel allocation (DCA) and hybrid channel allocation (HCA) [127].

In FCA a number of channels are assigned to each cell following a certain reuse pattern [128]. The channels can be assigned uniformly, where the same number of channels are on disposal to each cell, or non-uniformly in which case the number of channels per cell is tailored according to the expected load. Furthermore, cells that have all nominal channels in use can, through static channel borrowing or temporal channel borrowing, acquire additional channels from their neighbors to serve new calls [129–131]. Fixed channel schemes are very simple, however they lack flexibility. These schemes do not adapt to changes in traffic volume, a number of users or user distribution. To overcome the shortcomings of FCA, dynamic channel allocation schemes have been designed. At a cost of increased complexity, DCA can flexibly provide additional channels to the cells for new calls [132, 133]. The channels here belong to a common pool and are assigned dynamically to the cells depending on the need. In comparison to FCA, DCAa are less efficient under high traffic loads as they cannot always achieve a maximum re-usability of the channels. The DCA schemes can be implemented in centralized and distributed fashion. In the centralized schemes a controller assigns a channel form the central pool according to a specific cost function as a criterion to select the best channel. Criteria for channel selection such as: first available (FA), mean square (MSQ), nearest neighbor (NN) are just a few examples from the large set of solutions for channel selection [127, 134]. Distributed schemes have drawn a considerable attention and have been intensively studied in the literature because of their good

[1]Here we do not consider CDMA based networks for brevity and because they are out of the scope of this thesis

performance in microcellular systems and their scalability [135–138]. They use local information about the available channels in a vicinity of the base station or signal strength measurements to assign channels. Hybrid channel allocation schemes are a combination of fixed and dynamic ones. A *fixed channel set* is available to each cell as in FCA, but if all of these channels are in use and more capacity is needed for additional users and services in a certain cell, a channel from the *dynamic set* will be assigned. The dynamic channel set is shared by all the cells in the system and different procedure for assignment of the dynamic channels exist in the literature. Some of these schemes are presented in [139–142]. For further reading on different channel allocation schemes in cellular networks the reader is referred to [127] and the references therein.

5.1.2 Channel assignment in 2.4 GHz ISM-band

The 2.4 GHz Industrial, Scientific and Medical (ISM) band is a home of several widely deployed wireless technologies such as Wi-Fi (IEEE 802.11/b/g/n), Bluetooth, low power radio sensors (IEEE 802.15.4) and microwave ovens, just to mention few dominant ones. Mostly due to the fact that the band is unlicensed and the price of the equipments is relatively low we have witnessed an explosion of the number of devices in 2.4 GHz band. At present virtually every hand-held device or notebook has Wi-Fi and Bluetooth integrated. In particular Wi-Fi networks have been deployed to provide wireless Internet access in homes, cafés, airports, hotels, university campuses, corporate organizations etc. The widespread of these short-range technologies inevitably led to increased interference and stimulated a lot of interesting research on finding solutions for interference mitigation.

Increasing number of corporate wireless local area networks (WLANs) are becoming centrally managed, which means there is a central controller coordinating the access points (APs) to mitigate excessive interference and provide higher bit rate. However, large part of the WLANs are still uncoordinated, for example hotspots in restaurants, cafés or residential areas. In those cases there is a danger of high interference among the different access points as the selection of the operating channel or the transmission power are not centrally coordinated. In the literature a number of techniques have been proposed to address performance issues of WLANs such as load balancing (association control), power control, better admission control through enhanced MAC protocols, etc. Obviously one of the most extensively studied technique is channel assignment [143]. A plethora of algorithms have been developed, both centralized and distributed, to fight performance degradation due to interference. Before walking through the most interesting solutions let us have a brief look of the 2.4 GHz frequency band structure.

There are 14 overlapping channels, each 22 MHz wide, in the 2.4 GHz band, which can be used for IEEE 802.11b/g/n. Depending on the country the total number of channels available may vary. In Europe channels 1 to 13 are allowed, whereas North, South and Central America use the band up to channel 11. In Japan, on the other hand, all 14 channels are in use [144]. This channel arrangement ensures only three non-overlapping channels (minimum required separation is 25 MHz) where communication can be carried out without cross-channel interference. Recent studies, however, have shown that even non-overlapping channels can interfere with each other if the interfering transmitter is in a proximity of the receiver in a multi-channel mesh network type of scenario. The authors in [145] demonstrated that non-overlapping channels are not always decoupled and effects such as spurious carrier sensing and increased interference may occur, which can lead to decreased performance of simultaneous transmissions in non-overlapping channels. Similar findings have been also reported in [146, 147].

The interference problem in ISM bands becomes even more complex if we take into account the low power technologies such as Bluetooth and IEEE 802.15.4. In IEEE 802.15.4, 16 channels, 2 MHz wide, span between 2400–2483.5 MHz. Figure 5.1 depicts the channel placement of IEEE 802.15.4 in the 2.4 GHz band in respect to the three non-overlapping channels of IEEE 802.11. Bluetooth, in comparison, uses Frequency Hopping Spread Spectrum (FHSS) and is allowed to hop between 79 different 1 MHz-wide channels. In order to ensure coexistence, the devices operating in the 2.4 GHz ISM band have to comply to number of regulations that limit their transmission power or force them to spread the signals. Additionally MAC and PHY techniques such as carrier sensing or frequency hopping make the coexistence feasible. Yet, these mechanisms cannot completely prevent interference problems.

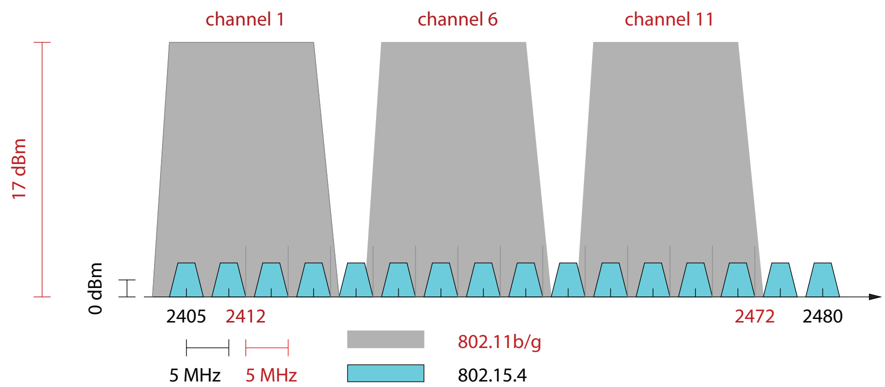

Figure 5.1: IEEE 802.11b/g and IEEE 802.15.4 channels.

As an example, it is a very well known fact that IEEE 802.15.4 links suffer severely in terms of throughput and experience high packet error rate in presence of IEEE 802.11b [148, 149]. In our earlier experimental studies we showed that there should be at least 7 MHz offset between the operational frequencies for a satisfactory performance of the IEEE 802.15.4 [150]. Furthermore in [22], we reported that in an environment with a medium to high IEEE 802.11n traffic load it is very difficult to guarantee the quality of the nearby operating IEEE 802.15.4

based sensor networks. If no care is taken about the operational channels of the two technologies, the IEEE 802.11n itself will have a negative effect on the performance of the IEEE 802.15.4 transmission, resulting in a very poor packet delivery ratio (PDR). We have also observed that even outside of the operating channel the IEEE 802.11n power is high enough to seriously interfere the IEEE 802.15.4 channels. A recent study from Gummadi et al. [151] investigated the influence of narrow-band malicious interferers such as Zigbee devices, wireless cameras and cordless phones on the performance of 802.11 networks. The authors concluded that commodity 802.11b/g receivers suffer from high losses and proposed a rapid channel hopping scheme to improve interference tolerance and mitigate the performance degradation of 802.11b/g networks. Further findings on the effects of mutual interference between the technologies in the 2.4 GHz band and solutions for coexistence can be found in [152–154] and the references therein.

Channel assignment in centrally managed networks

There is a strong need to coordinate the increasing number of WLAN deployments in order to control interference and maximize network capacity. This need motivated an extensive work in finding efficient channel allocation schemes. In the early days of WLANs the channel assignment was either regarded as a part of the corporate network planning problem or was totally ignored particularly in the case of consumer and small-business deployments. Traditionally, in the centrally managed networks, the channel allocation was carried out after the locations of the APs were determined and different techniques were used, including trial-based settings and off-line map coloring to determine the channel allocation [155]. In the beginning of 2000s, some researchers, the author among them, started a systematic work towards providing tools and algorithms to solve the channel allocation problem. In [156], for example, Wertz *et al.* used *integer linear programming* to simultaneously solve the optimal AP placement and channel assignment problems. Somewhat later, in [157], the authors proposed a local searching algorithm called *patching* to jointly optimize the AP placement and channel allocation such that maximum throughput and fairness is achieved among the users in a multi-cell WLAN. In our early work, we demonstrated how graph coloring can be used to design a protocol, which effectively assigns frequencies to APs [24,158]. We showed through extensive simulations and later by using a real testbed measurements that DSATUR coloring algorithm is a very promising solution [26]. It significantly enhances the performance of the network and is yet simple and computationally light. More elaborative analysis of our solution will be presented in Section 5.3.

A client-driven approach for joint channel assignment and load balancing has been proposed in [159]. The work advocates a centralized algorithm called CFAssign-RaC based on conflict set coloring. The algorithm extends the classical idea of using coloring for channel assignment for APs by capturing also the interference and the possible conflicts at the wireless clients side. In a similar setting, Chen *et al.* introduced a channel assignment that minimizes the *weighted interference* measured by the APs and their clients [160]. The weighted interference is calculated by putting different weights on the individual measurements based on the traffic volumes, signal strength and the traffic loads the APs and clients experience. The authors propose three types of channel switching algorithms, which require careful coordination between the APs and intensive information passing towards the central coordinator. Unfortunately, in this case the solution has been tested only through simulations and no real testbed results are available.

Channel assignment in uncoordinated networks

As discussed above, still very large part of the present-day wireless network deployments are unplanned and unmanaged. Typical examples are hotspots in residential areas. Such spontaneous deployments have been also named chaotic networks as each AP has its own parameter settings (very often the default manufacturer configuration) not optimized for coordinated operation with the neighbours [161]. Self-configuration of several key access point parameters among which the channel has been a matter of study for a long time [162]. Many automatic channel assignment algorithms have been proposed in order to improve the end-user experience and quality of service. Techniques such as least congested channel search, weighted coloring, channel hopping and MinMax approaches where the intention is to minimize the most overloaded channels are just selected examples from the large set of solutions for channel (re-)configuration [163–166].

5.2 Channel allocation as a graph coloring problem

Back in 2004, our work on automatic channel allocation for APs using "degree of saturation" (DSATUR) was one of the first attempts to use graph coloring as an efficient channel allocation in context of WLANs [24]. The encouraging simulation results in [25, 158] motivated us to implement and test the algorithm in a office type of environment using off-the-shelf hardware and later USRP software defined radio platforms. Further extensions and refinements towards applicability of our solutions in more dense networks and a complete protocol design were reported in [26].

Of course, graph coloring techniques have a long history in interference aware channel allocation, especially in cellular networks [167–172]. For the interested reader, a comprehensive overview on channel assignment using graph coloring techniques in cellular networks and an exhaustive reference list is given in [173]. By letting each transmitter correspond to a vertex, with edges connecting two nearby interfering transmitters, the problem of channel allocation can be easily modelled as vertex-coloring. The general aim is to color the graph such that the adjacent vertices are not assigned the same color (channel) to avoid interference.

In the following formulate the channel allocation of APs as a graph coloring problem.

5.2.1 Graph coloring preliminaries

Let us consider a simple graph $G = (V, E)$, consisting of a set of vertices V, and set of edges E connecting the vertices so, that loops and multiple edges between vertices are not allowed. Then a *k-vertex coloring* of G is a function $c : V(G) \to F$, where F is a set of k *colors*. A coloring is called *admissable*, if $c(V_i) \neq c(V_j)$ for all *adjacent* V_i and V_j. Furthermore, a graph that permits a k-coloring, is called *k-colorable*.

The minimum number of colors, needed for any k-coloring of the graph G is called *chromatic index*, $\chi(G)$. So if G is k-colorable it means that $\chi(G) \leq k$. An admissable coloring that minimizes the size of the color set, $|c(V)|$ is considered to be an *optimal* coloring.

Given a collection $\{V_i\}$ of access points (or radio transceivers in general), we can now construct an *interference graph* $G = (V, E)$ as follows. The vertex set V is simply identified with the set $\{V_i\}$. The set of edges E is constructed as the union of those pairs $\{V_k, V_l\}$ of vertices, that correspond to access points V_k and V_l that would interfere with each other should they be assigned to use the same channel. Finally, we define the set of "colors", C, to be the collection of channels available to the access points. Now the channel allocation problem is simply finding of an admissable coloring of G with the color set C. Figure 5.2 is an illustration of an interference graph that is 3-colorable.

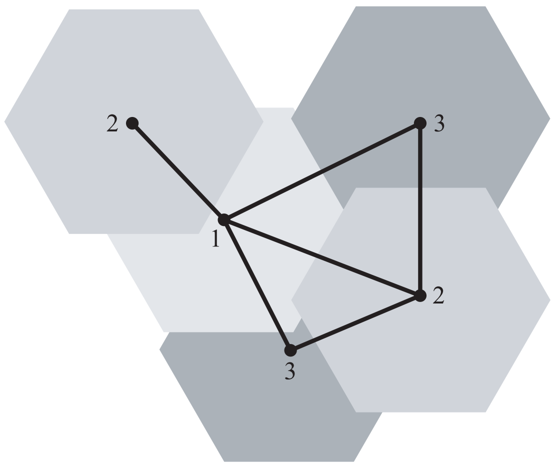

Figure 5.2: Illustration of an interference graph.

The size of the color set is strongly technology and legislation- dependent. As we discussed in the previous section in most European countries, $C = \{1, 2, \ldots, 13\}$ for the IEEE 802.11b and 802.11g technologies, of which the subset $C' = \{1, 6, 11\}$ corresponds to the non-overlapping channels. For IEEE 802.11a the set of non-overlapping channels is larger and equals eight. From the general graph theory we know that an upper bound on the number of colors needed to color different graph classes exists. The Brook's theorem states that if the graph is neither an odd cycle or complete the chromatic index is upper-bounded as: $\chi(G) \leq \Delta(G)$, where $\Delta(G)$ is the maximum degree of the graph. For odd cycles and complete graphs this bound is increased by one. For more extensive background in graph theory and graph coloring, including proofs of the above statements, please see [174,175].

It is well known that finding a minimum coloring of a graph is NP-hard problem [176]. This means that a graph coloring cannot be solved in a polynomial time. Thus, adding more APs in the network would lead to exponential increase of the computational complexity, which at first glance makes graph coloring inapplicable for channel allocation problems. Fortunately, it is possible to develop approximative algorithms to solve the problem. Efficient heuristics have been developed leading to a number of different coloring algorithms that can optimally color an interference graph under different constraints in a polynomial time. In this thesis we studied and developed a heuristic based algorithm suitable for AP interference graph coloring. After the initial study we found out that "degree of saturation" or DSATUR algorithm proposed by Brélaz [177] is a very effective and yet simple technique. We adopted DSATUR in our channel allocation and we showed that it has attractive properties for real implementation. Apart of DSATUR, we will also shortly discuss two other classes of coloring schemes below namely, on-line coloring and T-coloring.

5.3 Automatic channel allocation with DSATUR

5.3.1 The algorithm

DSATUR is a deterministic greedy algorithm that uses a heuristic method called "degree of saturation" according to which the vertices of a given graph are colored. More precisely, the heuristics is used to find the subset of vertices with highest "degree of saturation", that is, the vertices with the largest number of differently colored neighbors. If there is only one vertex in the subset, it is to be colored. If the subset contains more vertices the coloring is performed in the order of decreasing number of uncolored neighbors. In case more than one candidate vertex remain, the selection which vertex to be colored first is done at random. To better illustrate how DSATUR works, let us color the simple graph shown in Figure 5.3. The example graph will be colored as follows:

1. As initially the degree of saturation for all the vertices is zero, the vertex with highest degree is chosen. Consequently vertex V_1 will be colored in a greedy manner, that is, with color 1.

2. Now vertices V_2, V_3, and V_4, all have the degree of saturation of one, so again the degree of the vertex is used as a tie-break rule. As both V_3, and

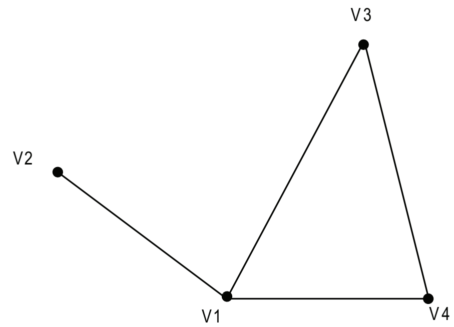

Figure 5.3: Example graph.

V_4, have the same degree, additional tie-break takes place. Suppose that our deterministic selection function picks V_3, to the colored first, using the color 2.

3. In the third step the degree of saturation of V_4, is two, and that of V_2 is one, so V_4, is colored with color 3.

4. Finally, V_2, is colored using the smallest color number consistent with the coloring problem conditions. Color 1 is therefore not possible, as that is already used for V_1, but already 2 is a possible color, which is then applied to V_2. Thus the algorithm terminates, and the final coloring is given by $C(V_1) = 1$, $C(V_2) = 2$, $C(V_3) = 2$ and $C(V_4) = 3$.

In order to give the algorithm more rigorous mathematical formulation we will denote the neighborhood of vertex v by $\gamma(v)$, the total number of vertices by N, the number of already colored vertices by i, and set of uncolored vertices by U. The pseudo-code of DSATUR can be presented in the following form:

We applied DSATUR for automatic channel assignment as the algorithm has several attractive features. The deterministic nature of the algorithm allows the channel allocation to be implemented in a distributed fashion without mandatory use of central coordinator to run the coloring. Additionally DSATUR has low computational complexity and very acceptable memory consumption. The authors in [178] showed that running time of $\mathcal{O}(m \log n)$, where m and n are the size and the order of the graph respectively, is achievable. Another valuable feature of the algorithm is that it produces good colorings. As shown in [178], DSATUR colors optimally almost all k-colorable graphs.

Several modifications and extensions to the original DSATUR algorithm exist in the literature. In [170], the authors presented a randomized DSATUR extended with local search that performers good in standard benchmarks. Nevertheless we chose original DSATUR as for on-line applications like ours local searches may have arbitrary runtimes, leading to more complicated and unpredictable termination conditions.

Input: U, $\gamma(v)\ \forall v \in U$, N
Sort U from largest to smallest degree;
Color first vertex v of U by 1;
$i := 1$;
Delete v from U;
while $i < N$ **do**
  $j := 1$;
  found := "no";
  Select first w from U with maximum degree of saturation;
end while
while found = "no" **do**
  if Some $x \in \gamma(w)$ has color j **then**
    $j := j + 1$;
  else
    found := "yes";
    Color w by j;
    $i := i + 1$;
    Remove w from U;
  end if
end while
All done; Output the coloring

ALGORITHM 1: The DSATUR algorithm.

There are several very natural and simple extensions to the frequency allocation technique so far discussed. First, instead of one color being assigned to every vertex, multiple colors could be assigned, where each combination of colors is still admissible. This will give the APs freedom to locally select the channel. Assignments of color sets are very natural for cellular networks. Second possible extension could be enabling incremental colorings, that is, coloring of newly installed APs in a already colored interference graph. The last extension concerns allowing partially overlapping channels into the color set for larger networks where the set of non-overlapping channels is not enough. There are several classes of coloring algorithms that can address the above introduced extensions, such as on-line coloring and T-coloring. A short walk-through is given in the next subsection.

5.3.2 On-line coloring

It could be very impractical if access points should change channels while serving users. Due to non-existence of a standard technique for user equipments to smoothly change channels while communicating, re-coloring of the graph, should be carefully scheduled if broken connections are to be avoided. Therefore it is best to perform the updates to the colorings as a new access point V_k is added to the interference graph so, that the existing channel allocation are unaffected, if possible.

A class of coloring algorithms, called on-line graph coloring algorithms exists that can be applied for such a channel allocation problem. The algorithm colors a given graph G in a manner such that one vertex at a time is revealed. Once a vertex comes into play, the edges with the previous arriving vertices have to be established and then the color is determined. Unfortunately these algorithms have a very poor worst-case performance (see, for example [179] and references cited therein) in terms of the colors used. The absolute upper limit for the performance ratio r in [179] is shown to be

$$r = O\left(\frac{n}{\log^2 n}\right).$$

Irrespective of this downside, on-line coloring algorithms might be valuable assets in the future, especially in the cases where larger number of non-overlapping channels are available. Detailed analysis of the on-line coloring in the channel assignment context, can be found in [180].

5.3.3 T-coloring

For dense interference graphs where the channel set C' of *non-overlapping* channels is not large enough for admissible coloring, additional channels must be taken into account. One solution is to perform the coloring using the complete channel set, and impose a distance-type of condition for the channels of adjacent vertices. The corresponding coloring problem is called T-coloring problem, which was was formulated by Hale [167] as a model for the channel assignment problem. T-coloring has been extensively studied in the literature (see for e.g., [168,181–183]).

One difference compared to the classical coloring is that interference is modelled by a non-negative integral set T (called a $T - set$), where 0 is forbidden channel separations. Thereafter a valid channel assignment or T-coloring is a mapping c_T from the vertex set of $\{V_i\}$ to the set of non-negative integers such that

$$c_T(V_k) - c_T(V_l) \notin T$$

holds for all neighbor vertices V_k and V_l.

In WLANs, the elements of T determine the smallest allowed difference of the central frequencies used by interfering access points. Separation of five channels (or more) would mean operating in non-overlapping channels for 802.11b/g, while use of smaller values would allow some overlapping of the used channels.

5.3.4 Simulation evaluation

We obtained the first performance analysis of the DSATUR channel allocation algorithm through simulations. The simulations were performed with ns-2 [184]. As ns-2 cannot model intra-channel interference, we put the focus on three-colorable interference graph topologies, and classical graph coloring heuristics.

The coloring algorithm itself was implemented in Java as an external program. From the Tcl-file describing the simulation setup, a script was used to extract only the access points into a separate Tcl-file used to determine the interference graph topology. This was accomplished by attaching sources and sinks into each access point, and recording the pairs between which communication was possible. The graph was then given to the Java-program for coloring, and in the final stage Tcl-files were created to give both the colored and the random channel assignments, based on the original scenario file. Finally, traffic flows consisting of constant bit rate (CBR) traffic over UDP and/or FTP-traffic over FullTCP connections were put into place.

In these simulations the nodes were static. Mobility was not considered to be critical for the performance results as the users WLAN hotspots are typically rather static. Individual simulations were repeated several times with different random number generator seeds, and a large number of different node and traffic configurations were used to ensure statistically significant results. Random channel assignments were used as a baseline, since no well-established alternative frequency assignment scheme existed for WLANs to make comparisons against. In fact, random channel assignment can be even considered as an overly pessimistic comparison, since numerous deployed networks simply use the factory default settings, leading to use of a *single* channel.

We used three parameters to evaluate the effectiveness of the coloring approach:

- The total number of collisions occurring in the network.
- The aggregate throughput of the network.
- In the case of TCP traffic, the perceived connection round-trip-time (RTT) and its variance. Both of these statistics were obtained from the FullTCP implementation of ns-2, and were taken from randomly selected TCP end-nodes.

Before presenting the results it is important to note that these simulations are not meant to give a *precise* quantitative performance evaluation of the coloring approach. To accomplish that an extensive number of simulations have to be performed, with carefully controlled location distributions for the access points and users serviced. Additionally more realistic channel model and reception process at the nodes should be used, which are unfortunately not present in ns-2. However, as shown below, the consistency of the results clearly shows that the coloring algorithm performs significantly better than the random assignment of colors. These motivated our further studies through a real implementation, and measurements, shown later in this chapter.

Number of collisions

The first performance parameter we study here is the number of MAC collisions. The number of collisions is a good indicator for the quality of the links. In Figures 5.4 and 5.5 the number of MAC collisions for different number of APs is presented. The APs are placed randomly in the simulation area and the number of users each AP serves is set to 5. A CBR traffic generator is used to create UDP traffic whereas a File Transfer Protocol (FDP) is used as an application to generate TCP traffic in the network.

The results show that the network can clearly benefit from smarter channel allocation. In comparison to the random case, DSATUR has a great effect on the MAC collisions, particularly for higher number of APs.

Aggregate throughput

Figures 5.6 and 5.7 show the development of the aggregate UDP and TCP throughput as the number of access points is increased. We see that with the application of the coloring algorithm an almost linear growth of throughput is obtained, which is an improvement from the rather jumpy and considerably slower growth in the random case.

TPC RTT

TCP has well-known performance problems when used over wireless connections, that are not readily seen in the performance measures discussed above [185]. These problems manifest themselves typically through spurious retransmissions related to “delay spikes” [186] caused by incorrect estimation of the RTT behavior. Because of this, we also studied the behavior of the TCP RTT estimators in different frequency assignments. The corresponding results are depicted in Figure 5.8, showing a time-evolution of the estimators of a randomly selected node.

Clearly the RTT behavior is substantially more stable when the channel allocation with DSATUR is applied. Due to the structure of the TCP data flow this leads to faster recovery times when transmission errors occur, and the overall responsiveness of the connection is also enhanced. This is especially important for interactive applications, such as web browsing.

5.3.5 Implementation considerations

If our coloring channel allocation algorithm is to be deployed in a real wireless network several components are still needed. In order to construct the interference graph, a protocol for sharing neighborhood information among the APs should be designed. Additionally, a common message format for AP-to-AP communication is also beneficial to have. We propose here a possible message structure and methods the access points can use to communicate among themselves.

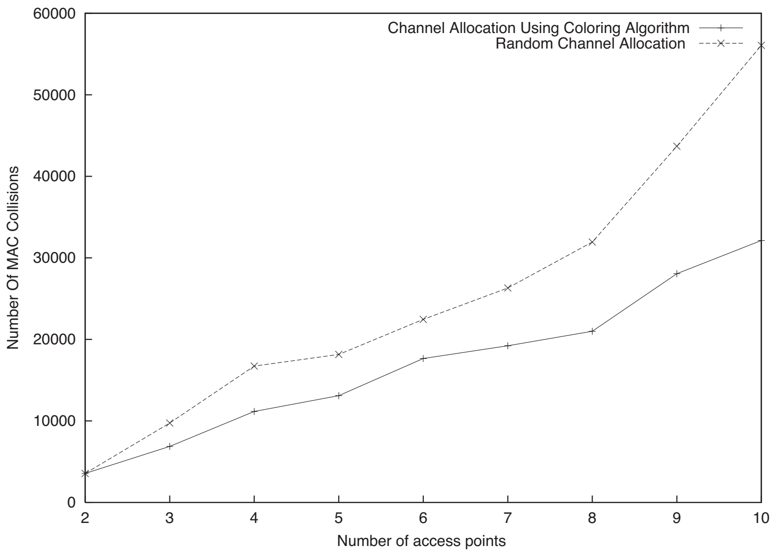

FIGURE 5.4: Number of MAC collisions as a function of the number of access points in a network for UDP traffic.

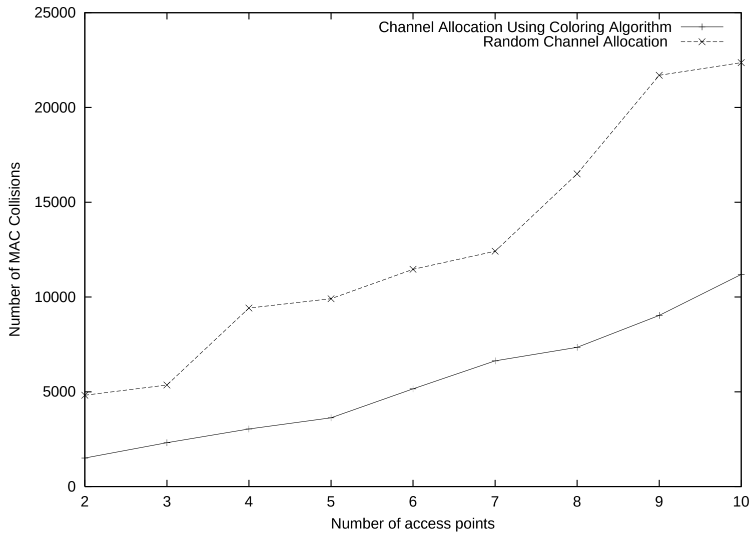

FIGURE 5.5: Number of MAC collisions as a function of the number of access points in a network for TCP traffic.

For constructing the interference graph the access points need neighborhood information. For this, a message consisting of one neighbor entry N_i per neighbor observed is sufficient, preceded by some mandatory header information. The header, can have the structure as in Figure 5.9:

Here the first 16-bit field gives the number of neighbor entries following the header, and the second 48-bit field gives the MAC address of the wireless interface.

In the simplest coloring application the neighbor entries N_i could consist only of the 48-bit MAC-addresses the station sending the message can hear. However,

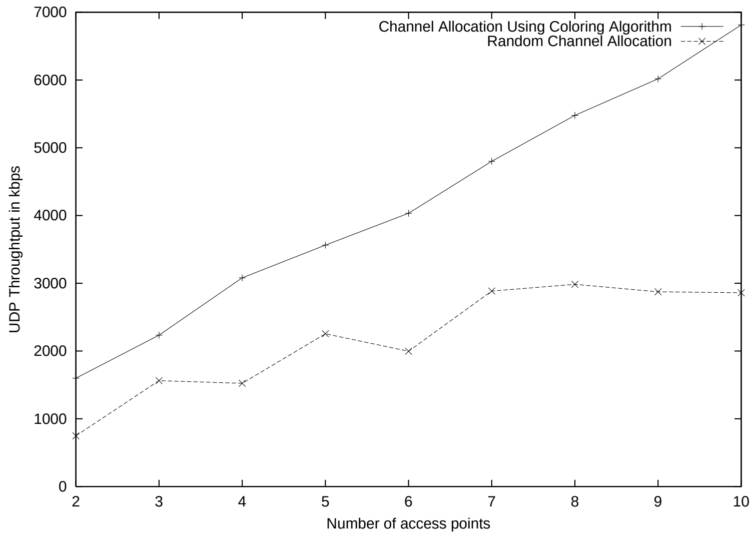

Figure 5.6: Aggregate throughput of the network carrying UDP traffic.

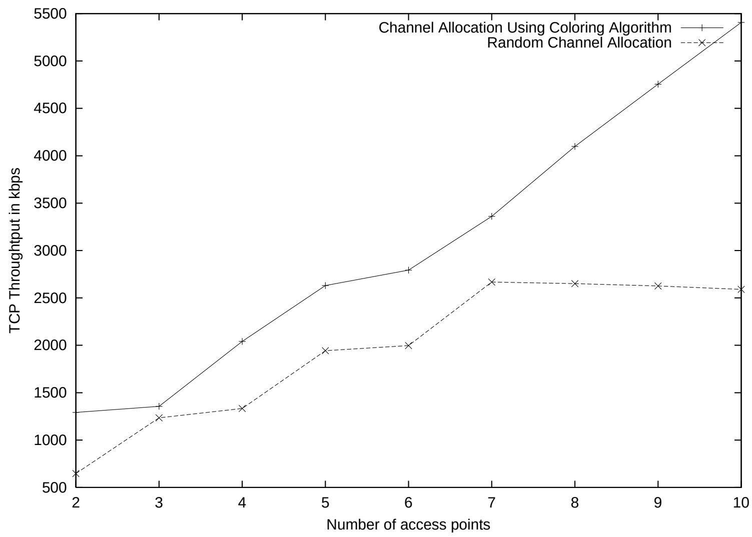

Figure 5.7: Aggregate throughput of the network carrying TCP traffic.

to enable the use of T-coloring and on-line coloring algorithms, information regarding the present channels assigned to the access points, and some measure of the level of the received signal strength is necessary. Thus we propose a format for the neighbor entries as in Figure 5.10:

In this case, the first two 8-bit fields give the channel used by the corresponding neighbor and the measured SNR.

The signalling overhead in terms of messages scales as $O(N^2)$, where N is the number of nodes participating to the frequency coordination process. The size of each message is 64 bits per radio neighbor (plus 64 bits for the header). Thus, with typical access point densities the messages can usually be carried inside

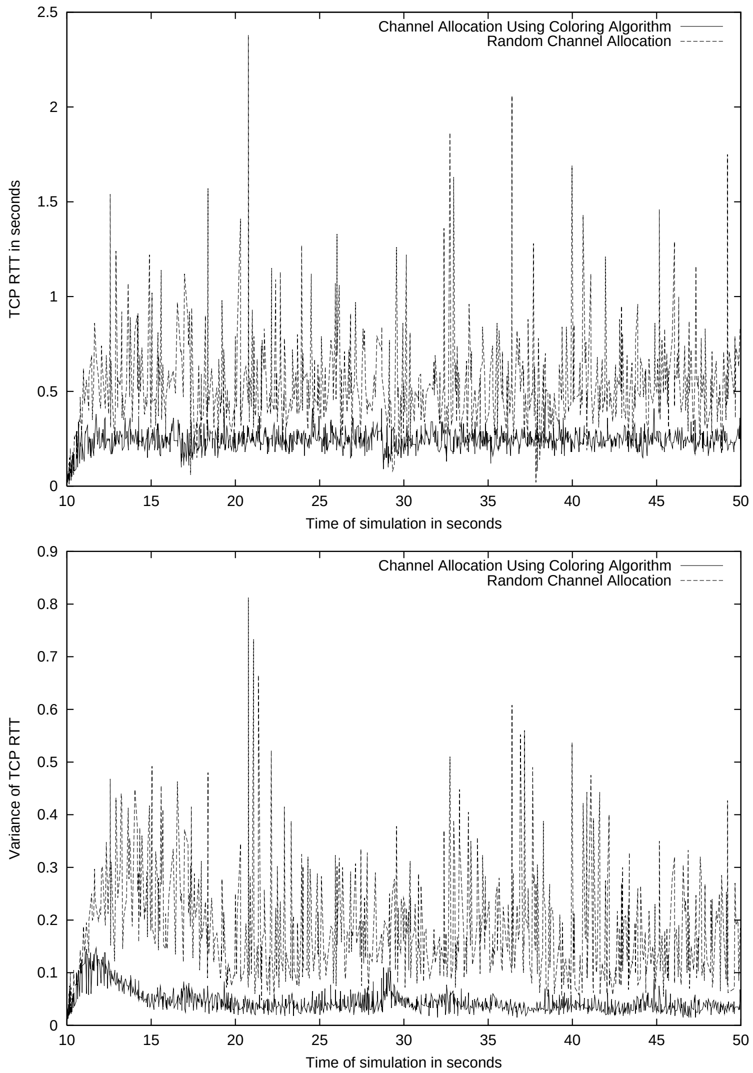

Figure 5.8: TCP RTT and its variance.

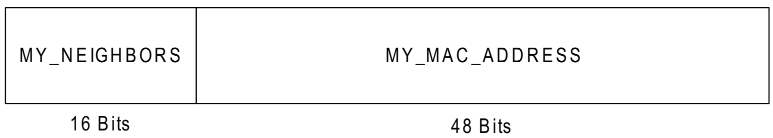

Figure 5.9: Observed neighbors.

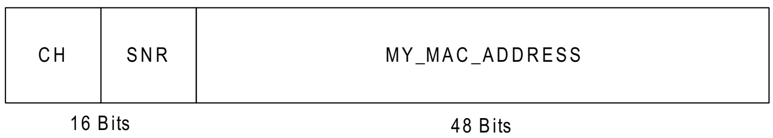

Figure 5.10: Operating channel and SNR of the neighbour heard.

single frame, without having to resort to IP-layer fragmentation.

When specifying a complete protocol for use, a slightly more modular structure, also carrying optional information for the coloring algorithms (such as the T-set used) will be necessary. Also additional management-related information on, for example, network domains could be useful. Especially security aspects should be considered in product-quality deployed systems, both from authentication, and denial-of-service points of view. It is also important to note that the radio resource management group, IEEE 802.11k and the IEEE 802.11v working group have worked on a set of common control commands and protocol messages. In order to enable management of large-scale 802.11 networks these messages can be considered as appropriate places to incorporate the coloring algorithm information as a possible solution for enhancing the performance of the WLANs.

The first step of a fully integrated coloring based channel allocation mechanism is the scanning of the channels for other access points, and formulating the radio neighborhood information message as described above. After the neighborhood messages have been constructed the simplest way of disseminating those messages would be flooding. Standard techniques (sequence numbers and time-to-live fields) can be used to control the extent of the flooding if necessary. However, this will require a common control channel in wireless LANs that all the access points listen to continuously. Thus, in the case of already operational WLAN, the access points should somehow coordinate the sending of the neighborhood information, and change to the same channel for this procedure, or send the information on several channels. Unfortunately this would result in breaking the associations with the stations the access points serve, so this probably cannot be done except during times of very small number of users.

More promising approach is to use the infrastructure network connecting the access points, especially in the case where the same company owns all the access points. In the case of access points being wireless routers as well, the MAC address heard over can be mapped into IP addresses using a simple address resolution protocol (ARP)-call, and any IP-based protocol can then be used for the final delivery of the neighborhood information. Of course, if multicast support in routers (wireless and fixed) becomes a reality, using geography-based multicast groups for access points might also turn out to be a good option.

In a real hot spot environment all the access points will not be connected to the same “Ethernet cable”, as they do not belong tot he same operator. Hence there is no possibility to match the MAC addresses to IP addresses via the ARP protocol and furthermore deliver neighborhood information. In order to be able to establish collaboration between neighboring nodes from different networks it might be convenient having a service discovery protocol providing this capability. As an example, the simple SLP (Service Location Protocol), both in centralized and distributed manner can provide a list of one-hop neighbors of the APs, which then can be used for contracting the interference graph. Alternative possibility can be using location discovery capabilities to find the neighbors. In addition, management techniques that rely on standardized protocols and are vendor independent can be considered for management of the APs from different providers. Simple Network Management Protocol (SNMP) is a good candidate for maintaining the global view of the entire network that consists of several cooperative WLANs[2].

5.3.6 Performance evaluation through measurements

We implemented and tested a full DSATUR graph coloring protocol in two different testbeds, one using standard WLAN cards and the other USRP software defined radio boards. Below a short description on the measurement setups are given followed by a presentation of the measurement results.

Setup 1

Our frequency allocation program tool is of a client-server type and comprises of two parts: one running in the server side (Linux machine connected to the APs via Ethernet) and the other running in the clients and/or APs. During the implementation process we took care to build a flexible software architecture in order to be able to add and remove APs easily. In the testing phase, as access points, we used regular laptops with NETGEAR MA401 PCMCIA IEEE 802.11b cards with Prism 2 chipset, running HostAP [187]. HostAP is a Linux driver that supports so called HostAP mode in which the WLAN card acts as an AP.

The actual DSATUR algorithm calculating the new frequency allocation is running on the server side. It means that the processing of the collected data from all the APs and the construction of the interference graph is done in a centralized fashion. In order to provide the neighboring information to the server we designed a specific protocol for exchanging information called Network Information Exchange (NetIx). The packet structure of the protocol is shown in Figure 5.11. The protocol uses 10 bytes for the header information and up to 40 bytes for the payload where maximum ten IP neighboring addresses can be transmitted. NetIx defines the following header fields:

- **Flag** is a two-byte filed used to identify the packet type. We have defined six different packets showed on the message flow diagram on Figure 5.12 .
- **Colour** is a two-byte field that carries the new colour calculated by the DSATUR algorithm.
- **Degree** carries the information about the number of neigbours that a certain AP has.
- **Own address** field shows the IP address of the AP sending the packet.

[2]IEEE 802.11k group has recently finalized its work and after submitting the IEEE 802.11k Amendment 1, the group is inactive. Some parts of our work can be implemented or supported by the current standard, but our coloring solution is not a part of it. However, using IEEE 802.11k mechanisms one can simplify the implementation of proprietary extensions.

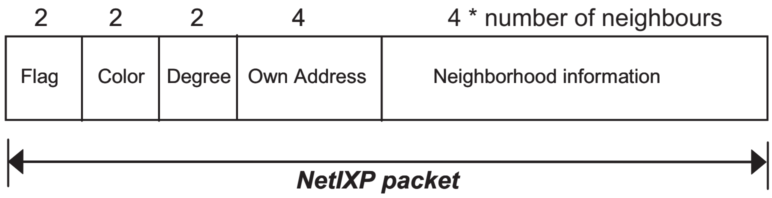

Figure 5.11: NetIx packet.

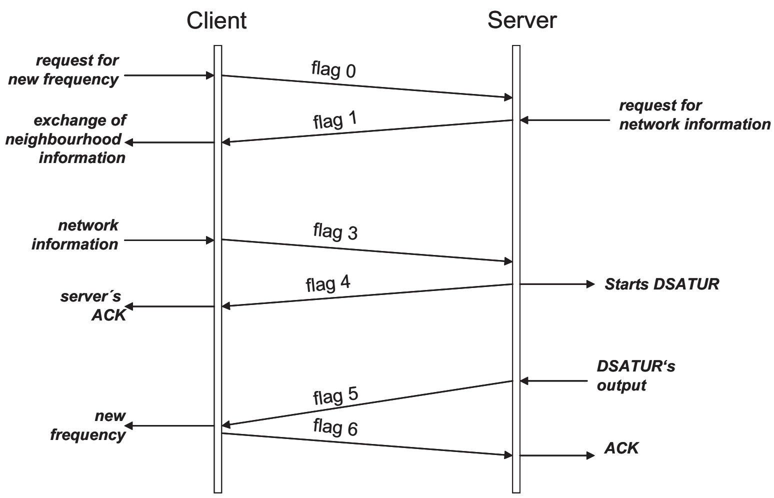

Figure 5.12: Client-server information exchange.

The process of new frequency allocation can be in general triggered by a new client (AP) coming into the network or by any client if the interference level is such that it produces serious impairments in the wireless communication. For our testing and measurements we implemented only the first option. In case the server gets a *request for new frequency* message from a client (`flag=0`), it broadcasts a *request for network information* message back. At this point the access point will start to exchange neighboring information among each other, in an ad hoc mode, by sending three broadcast packets with the `flag=2` via the wireless interface (for example all APs switch to channel 12). It should be shortly mentioned that the entire communication between the clients and the server is done through the Ethernet. The exchange of neighboring information is a one hop communication. When the process is finished, each client sends a neighboring report to the server via a message with the flag set to three. After these packets are acknowledged the neighboring information is fed to the DSATUR which constructs the new interference graph and calculates the frequency allocation. The coloring algorithm unicasts the assigned frequency for each AP. The cycle is finished as long as each client acknowledges the receipt of the new frequency. The transmission of the NetIx protocol messages is done over the UDP/IP packets. In order to increase the reliability of the UDP we added an ACK policy for the unicast packets so that each successfully received packet is acknowledged. Missing ACK will initiate a packet retransmission.

To evaluate the performance of the coloring-based channel allocation we set up a testbed consisting of a collection of laptops. Half of the laptops we running in the master-mode of the driver, acting as access points, and the rest operated as clients, see Figure 5.15[3]. Experiments were repeated using different numbers of access points to estimated the dependency of the performance obtained from the network density. One of the clients was randomly chosen to run `iperf` [188], using UDP traffic at various offered bitrates to probe the capacity available through the access point. Other client-AP pairs were transmitting data at high enough rate to saturate the respective channels.

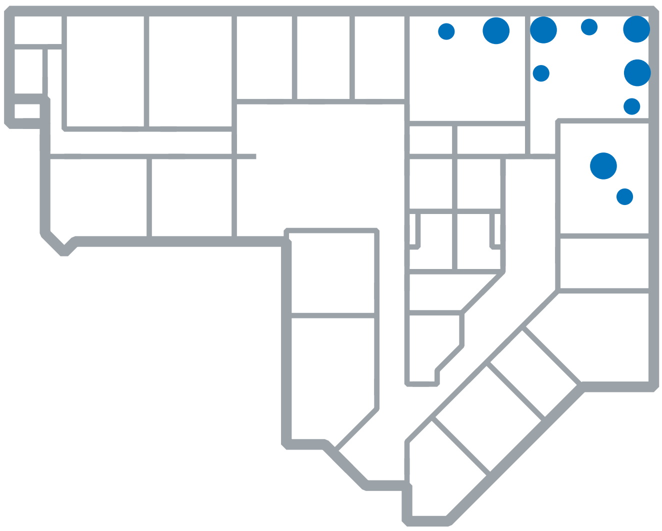

FIGURE 5.13: Placement of APs and clients in test setup 1.

Three channel assignment schemes were used in the measurements: fixed, random, and coloring-based. The fixed allocation corresponds to the situation where all access points are utilizing the same channel. This is actually the case with the majority of present-day access points. In the random case the channels for the access points were randomly drawn from the uniform distribution. This channel selection mechanism can be enabled in some commercially available APs. Finally, the coloring-based approach utilized the prototype software described in the above.

Figure 5.14 summarizes the *worst-case* performance of different schemes observed in the experiments. It is easy to spot the superiority of the coloring method. As expected the bitrate achieved with the fixed scheme is very low as severe degradations happen already when two or three access points share the

[3]The large and small circles correspond to APs and clients respectively.

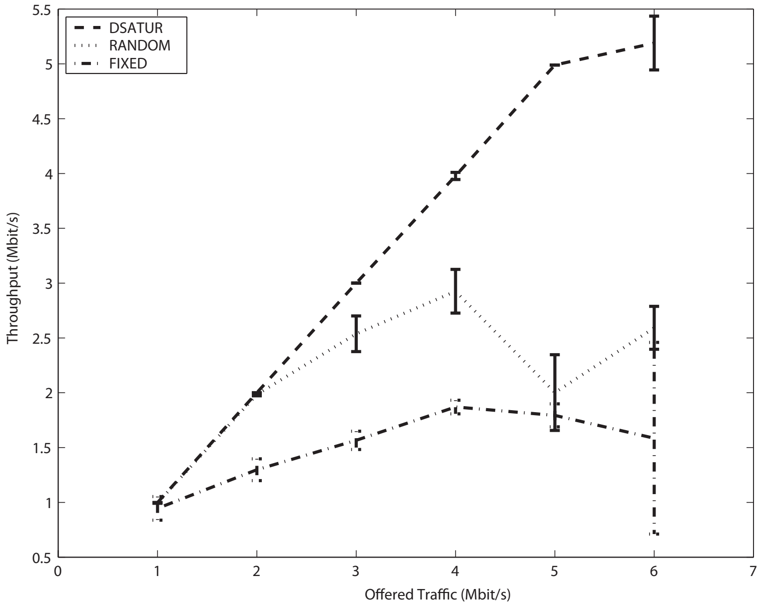

Figure 5.14: Comparison of the worst-case performance of the channel assignment techniques.

same channel. Random channel assignment achieves considerably improved performance. However, from the error bars it is obvious that the performance is highly varying. By far the most stable performance is obtained by the coloring method. The scaling of the average bitrate achieved as a function of offered bitrate is almost linear, and the channel is very stable as shown by the small error bars.

Setup 2

The second testbed consists of 18 USRP nodes, i.e., 6 APs, each serving 2 clients as shown in Figure 5.15[4].

Since we wanted to avoid overlapping channels, we configured five 5 MHz-wide channels in the available 25 MHz band. It is good to notice here that the measured width of the carrier frequency of the USRP is 3.5 MHz. The channels used for coloring are given in the Table 5.1.

The measurements were performed with the largest UDP payload without needing to fragment the packet (i.e. 1472 Bytes). We created two test cases. In the first test case, one of the two clients was randomly chosen to run `iperf` at UDP traffic bitrates from 100 kbps to 1.9 Mbps increasing in steps of 300 kbps. Each measurement was repeated 20 times. The other clients that were

[4]The nodes with bigger circles correspond to APs, whereas the smaller circles correspond to the clients. The dashed lines represent the association of clients to a specific AP.

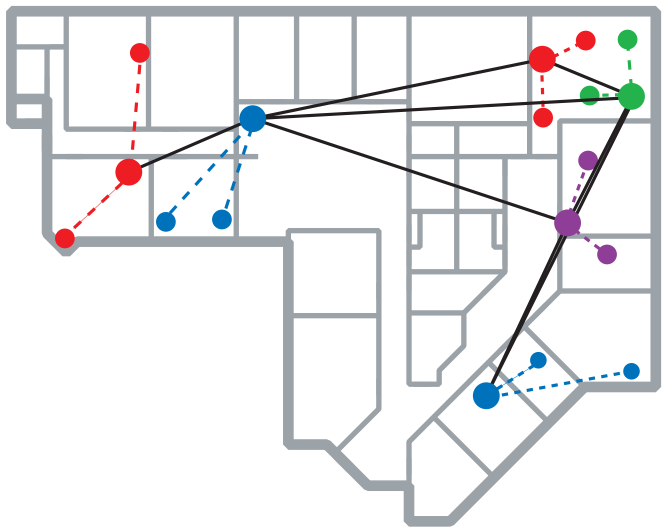

FIGURE 5.15: 6AP interference graph with coloring.

Channel	***Frequency (GHz)***
CH1	2.460
CH2	2.465
CH3	2.470
CH4	2.475
CH5	2.480

TABLE 5.1: Central channel frequencies.

not running `iperf` were transmitting a data flow directed to the AP at a high enough rate to saturate the respective channel. In the second test case each of the two clients run the `iperf` client at the same time. The goal was to emulate the situation where two clients compete for the shared bandwidth. The other measurements parameters were kept the same. The reader should note that here, similarly to the setup 1, we use the Network Exchange Protocol to enable client-service neighborhood message exchange.

In Figure 5.16 we show the results for a case when every node transmitting in the testbed uses a single fixed frequency channel (i.e., CH3 at 2.47 GHz). The experiments are done for different number of APs in an increasing order. Thus, we can compare the performance as the density increases. As one can may expect the bitrate achieved decreases as we add new APs. We observe that the performance is severely degraded after four or five APs use the same channel. The standard deviation error bars show that fixed channel allocation scheme is not stable in dense networks. Consequently, as the number of APs increased also the performance became highly varying.

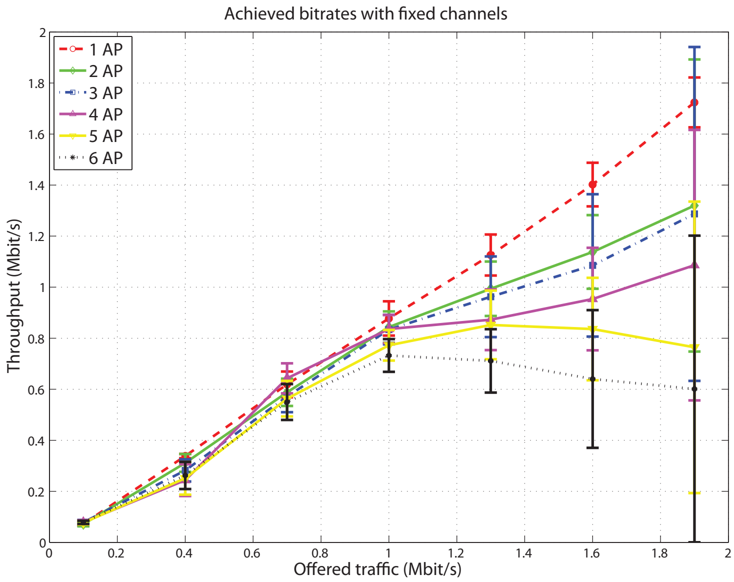

Figure 5.16: Performance of the fixed channel assignment.

It may be more interesting to have a look what happens with the throughput when the channels assigned to the APs are randomly drawn from a uniform distribution.

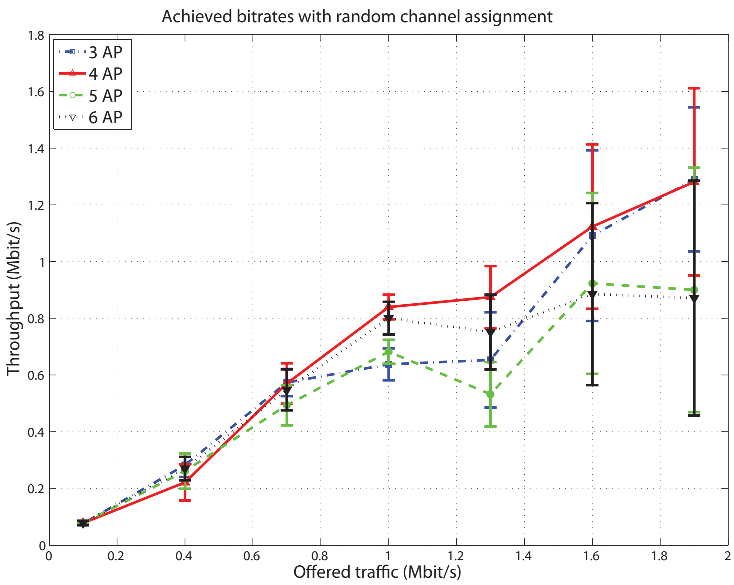

Figure 5.17: Performance of the random channel assignment.

The results in Figure 5.17 show a similar tendency as in the case of fixed assignation. Due to the randomness we can observe situations in which a scenario with fewer APs had better bitrates than another with more APs. In general the performance is clearly degraded when the number of APs increases. However, compared tot he fixed scheme, the random technique provided a more stable behavior.

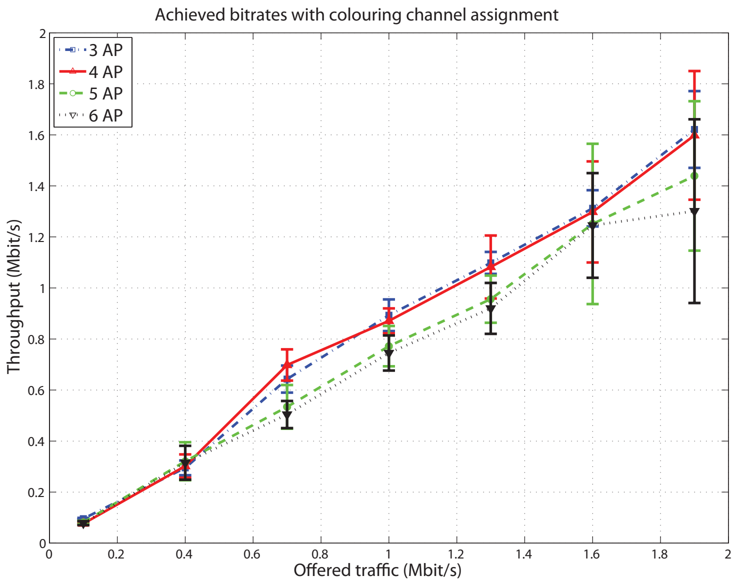

Figure 5.18: Performance of the coloring channel assignment.

Applying the algorithm will result in a vertex coloring as shown in Figure 5.15. From Figure 5.18, one can see that the results obtained with the coloring method exhibit almost a linear behavior, and the performance is by far better than in any of the other two methods analyzed. We obtained very good results even when the topology density in the scenarios increased. Furthermore, the bitrates exhibit smaller error bars, which means stable channel allocation and less fluctuations in the expected throughput.

In Figures 5.19 and 5.20 we summarize the performance of the three channel allocation schemes. Our coloring interference minimizing channel allocation scheme shows superior performance over the fixed and random channel allocation schemes. Moreover using the coloring scheme clearly lowers the datagram error rates in the network. The reader should note that we observed very similar results also for the case, where the clients are contending for the medium. In order to keep this part of the thesis concise the respective results will not be shown here.

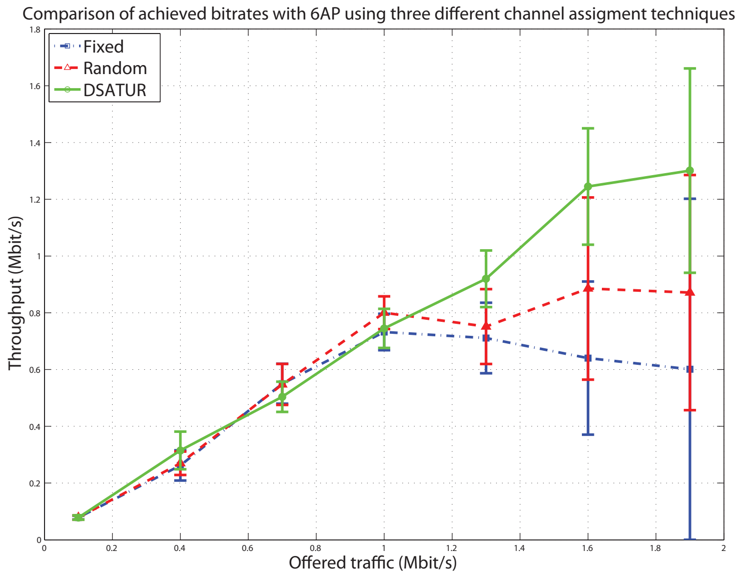

Figure 5.19: Performance comparison in the *worst-case* scenario using the three different channel assignment schemes.

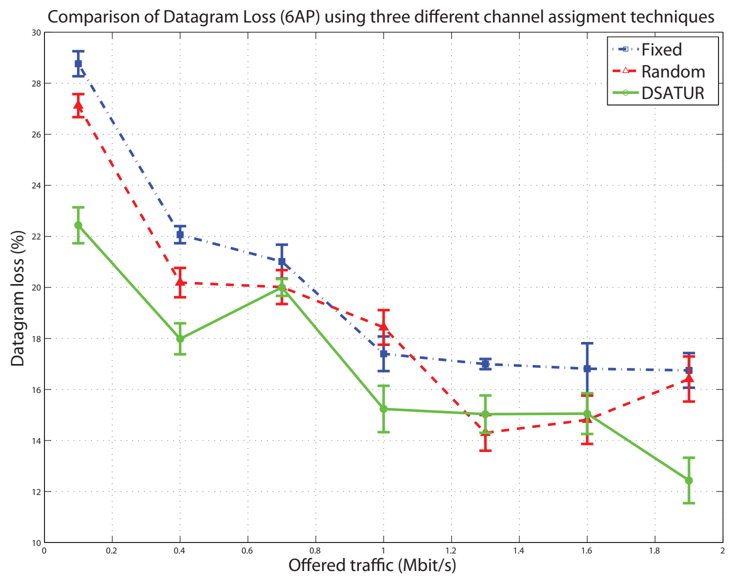

Figure 5.20: Datagram losses in the *worst-case* scenario using the three different channel assignment techniques.

5.4 Channel allocation as a load balancing game

One of the most important issues in the design of channel allocation schemes for wireless networks is fair and balanced distribution of available channels among the users. Especially in dynamic scenarios where the number of users and the available channels frequently changes, it is desirable to avoid situations where some of the channels become overpopulated and the others stay almost empty. Such *flash crowd effects* are very likely to occur in opportunistic spectrum scenarios if no coordination among the secondary users exists [23]. Balanced channel allocation not only improves the channel utilization but also reduces the probability of collisions and can lead to better performance for the users in terms of throughput, delay and fairness. Yet, load balancing channel allocation is a challenging problem to solve and implement, particularly when no central controller is available in the network and the environment is highly dynamic.

Similar and well studied problem in the literature is a fair and balanced distribution of users among the APs in a network. Several studies showed that the problem can be successfully attacked by more efficient user-AP association control, see for example [189] and the references therein. Different metrics and heuristics that maximize these metrics have been used as association criteria. Typically measured RSSI (Received Signal Strength Indicator) values of the links from the APs or the bandwidth the user can obtain from the AP have been selected as a criterion for association.

In this section we propose two novel distributed channel allocation algorithms, which can achieve load balancing among the available channels. The algorithms are simple, dynamic, with very good convergence properties and can apply for both general WLAN type of channel allocation problems and secondary spectrum sharing. Furthermore the algorithms run in distributed fashion and rely mostly on local information, thus not introducing unnecessary signalling overhead in the network.

We model the channel allocation as a *balls and bins* problem borrowed from game theory [190]. Let us assume that we have n balls and we throw them into m bins by placing each ball in a uniformly randomly selected bin as shown in Figure 5.21. One of the original questions asked for this model was to find out the maximum number of balls in each bin as discussed in [190]. This simple model has been used to describe many problems in theoretical computer science one being online load balancing, e.g., assigning requests to servers in such a way that all servers handle approximately same number of requests.

We use the balls and bins model to address the channel allocation problem in wireless networks. In our system model the cognitive radios (users or agents)[5] represent the balls and the available wireless channels represent the bins. Our main aim is to design an algorithm that minimizes the interference and maximizes the throughput performance for the users by allocating the available channels

[5]The terms users and agents will be used interchangeably in this context.

so that a load-balancing is achieved. Below, we elaborate on the convergence properties of the two proposed algorithms through theoretical and simulations. Furthermore we present a design of a distributed load balancing protocol based on one of the proposed algorithms and study its performance in both simulated and real network environment.

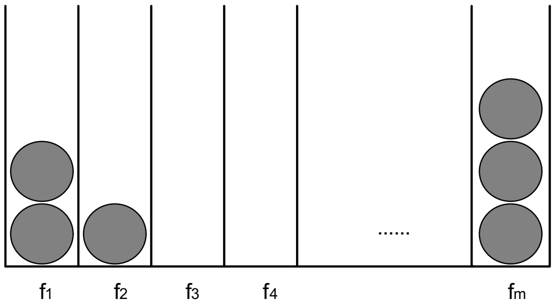

Figure 5.21: Sketch of the balls and bins problem.

5.4.1 System model

We are given n balls and m bins. The n balls correspond to n agents whereas the m bins represent the available channels in our model. For simplicity, but without losing generality, we assume that all balls are equal in size, meaning that they generate the same amount of traffic load in the system. Additionally we assume that all the agents use the same radio technology for communication, e.g., IEEE 802.11a/b/g. One should note that the presented solutions below in this chapter will also apply to unequal loads but these analysis are left for the future work. Let us denote the load in channel i, that is the number of balls that have selected bin $i \in [m]$, with n_i. Now, the whole system can be fully described with the state vector $\mathbf{n} = (n_i)_{i\in[m]}$[6]. The balls have a Poissonian living time and the number of balls n may fluctuate in time if user leave the network or new users join in.

In the initializing phase, we assume that the balls have been placed randomly in the bins. Later on every ball decides to stay in the current bin or change to another one based on the performance experienced. A performance measure can be throughput, delay, packet error rate or any other QoS parameter. In order to quantify the performance of the balls in each channel we define a utility or cost function c_i. The value $c_i(n_i)$ specifies the cost incurred by all balls placed in bin i. For our analysis we normalize the costs such that the maximum measured cost of any c_i is 1 and the minimum value is 0. We define the average sustained cost by the agents to be

$$C(n) = \sum_{i\in[m]} \frac{n_i}{n} \cdot c_i(n_i). \tag{5.1}$$

[6]The state vector is also often referred as an action vector in communications field.

From a game theoretical point of view it is interesting to find the balanced states in this model, which will ensure an existence of stable channel allocation. The driving force towards a stable load balanced channel allocation is the objective of each ball to minimize its cost.[7] A ball has an incentive to move from its current bin to another one only if it will get more utility by doing so. A stable state $\mathbf{n}$ is reached if no agent has such an incentive. In that case, we consider that the system is in Nash equilibrium.

Definition 5.1 (Nash equilibrium)**.** *A state* $\mathbf{n}$ *is at a* Nash equilibrium *if for all players i and j with $n_i > 0$ it holds that $c_i(n_i) \leq c_j(n_j + 1)$.*

The existence of Nash equilibrium for a certain system is interesting from a practical engineering point of view. A lack of stable solutions make the systems harder to analyze and implement. Thus knowing if there exist a Nash equilibrium for our load balancing system model is an important question. Fortunately, we are able to show with relatively simple arguments that this is the case. We start the proof by defining a congestion game. Congestion games are class of games introduced by Rosenthal in [191].

Definition 5.2. *A congestion game is defined by a group of resources E, and a group of players $N =$ 1, . . . , n. The strategy set of player i is defined to be sub-sets of resources: $S \subset 2^E$. For each resource $e \in E$ we have a cost function c_e: $\mathbb{N} \longrightarrow \mathbb{R}$. Each resource e has a cost function $c_e(x_e)$ when x_e players are using it. The cost function for each player is defined by: $c_i(s) = \sum_{e \in s_i} c_e(x_e)$.*

In other words with a congestion game one can define players and resources, where the payoff of each player depends on the resource it chooses and number of other players choosing the same resource. Based on **Definition 1** and the description of the balls and bins game we observe that balls and bins is a congestion game.

Conjecture 5.1. *Balls and bins is a congestion game.*

Definition 5.3. *A game G is defined to be an exact potential game, if there exists an exact potential function $\Phi : S \longrightarrow \mathbb{R}$ such that for every player i, strategy profile $s = (s_i, s_{-i})$ and strategy s_i' we have that: $c_i(s_i, s_{-i}) - c_i(s_i', s_{-i}) = \Phi(s_i, s_{-i}) - \Phi(s_i', s_{-i})$.*

An exact potential game has a function which maps strategy of profiles to real numbers, such that when player i deviates from strategy s_i to s_i', the change in the players cost function is exactly the same as the change in the potential function. It is well known fact that every potential game has a Nash equilibrium.

[7]If the utility is mapped to the throughput then the aim is to maximize the throughput, i.e., minimize the reciprocal value.

Rosenthal [191] proved that every congestion game is a potential game too[8]. Based on this we can state the following conjecture

Conjecture 5.2. *Balls and bins is also a potential game.*

Due to a potential function argument [191], pure Nash equilibria do always exist in this model, even if balls have different sizes [192].

Definition 5.4. *Due to Conjecture 1 and 2, and the fact that potential function argument holds, balls and bins games have always Nash equilibrium.*

Our focus here is to design distributed algorithms by which balls can attain an exact or at least an approximate equilibrium quickly. One should note that Nash equilibria do not necessarily optimize the overall performance of the system, but they have the appealing property of being fair from a local point of view, thus ensuring stability.

Before presenting the the load balancing algorithms below, we will point out several additional assumptions we take into account. In our analytical calculations and simulation analysis we assume that there is no inter-channel interference in the system. Obviously in the real testbed implementation the interference caused from the neighboring overlapping channels is inevitably present since we run the experiments in a dynamic office environment without control of the external interference. Another assumption that we make is that the agents are able to measure the load in the channel, i.e., the number of balls in the bin n_i. In principle it is very challenging to estimate the number of radios using the same frequency in a real, e.g., IEEE 802.11 setup, locally. One could measure the channel utilization, the transmit queue length or the packet delay on a MAC layer to get an indication of the channel load [164]. Here we take for granted that a procedure for measuring the load in the channel is provided and every agent has the ability to estimate the channel load n_i. We are allowed to make this assumption because very precise information on the channel load is not critical for the design or the performance of our load balancing protocol as shown later in this chapter.

5.4.2 Load balancing algorithms

Here we will present in detail two algorithms Compare_And_Balance and Avoid_Contention, which have a good potential to serve as building blocks for load balancing channel allocation in a distributed fashion.

While designing the algorithms we made several assumptions in respect to agent's capabilities. Both of the proposed algorithms rely on the ability of the agents to estimate the load in each channel, by scanning the spectrum. The primitive Measure_Load() returns the estimated load, measured as number

[8]Inverse mapping was shown later by Monderer and Shapley in their paper "Ptential Games", *Games and Economic Behavior*, **14**, 124-143 (1996).

for all balls in parallel **do**
  $c \leftarrow$ MEASURE_COST(*channel*)
  for all channels $i \in [m]$ **do**
    $n_i \leftarrow$ MEASURE_LOAD(i)
  end for
  $n \leftarrow \sum_{i \in [m]} n_i$
  choose j randomly with $\mathbb{P}[j] = n_j/n$
  $c' \leftarrow$ MEASURE_COST(j)
  if $c' < c$ **then**
    with probability $c' - c$: *channel* $\leftarrow j$
  end if
end for

ALGORITHM 2: COMPARE_AND_BALANCE

of agents in a particular channel $\mathbf{n}_i$. We also take into account that $\mathbf{n}_i$ may be observed under uncertainty. There are number of different approaches that real systems can use for estimating load and/or interference in particular channels. For example, SNR estimates or multiple user detectors (MUD) can be used to estimate the number of users in different channels. One fortunate coincidence is that DSA cognitive radios anyway require this type of information, so such technological capabilities would naturally exist in the future CRs.

In order to be able to make the decision of switching the channel, the agent should be able to know the utility of the possible new channel. Although this utility can be measured by jumping to the candidate channel on a trial basis, it might be computationally intensive and energy inefficient transceivers. Alternatively, the agents could decode the packet headers on the target channel to estimate the utility perceived by the agents using that channel. Independent of the method of the utility estimation, we provide a primitive MEASURE_COST(), which returns the value of $c_i(n_i)$.

Algorithm: Compare_and_balance

We assume that MEASURE_COST() can be applied to any available channel to the user. Consider an agent currently using channel i. At first, the agent determines the vector $\mathbf{n}$ using MEASURE_LOAD() and then randomly samples another agent in channel j measure the cost experienced by that agent. Now, the agent compares its own cost with the cost of channel j used by the sampled agent. If $c_j < c_i$, it will migrate towards channel j with probability $c_i - c_j$. A pseudocode of the algorithm COMPARE_AND_BALANCE is given in Algorithm 2.

At first sight, this algorithm might look counter-intuitive. The probability of a channel being sampled increases with the number of agents already utilizing it. This seems to contradict the actual goal of balancing the load. However, there are fundamental theoretical reasons for this approach. Intuitively, if many agents utilize a certain channel it is an indicator that this channel has a high utility. To make that more precise, let us assume, for the time being, that our algorithm knew an assignment $\mathbf{n}^*$ at a Nash equilibrium in advance. In that case, it would be optimal to sample channel i with probability n_i^*/n since this would yield a Nash equilibrium within one round in expectation. Since $\mathbf{n}^*$ is, in general, not known in advance, we must find an estimator for it. Suppose that cost functions are linear, i. e., $c_i(n_i) = n_i/s_i$ where we envision the parameter s_i as the *speed* of channel i. Since at a Nash equilibrium, the values of $c_i(n_i)$ are approximately equal, n_i^* is approximately proportional to s_i which suggests to sample channels proportionally to their speed. In fact, this technique has been successfully applied in [193] in a discrete model, and in [194, 195] in a continuous model. However, the value of s_i may not be known, or it may not even exist for non-linear cost functions. In this case, we may use n_i as an estimator for n_i^*. This estimator becomes better and better quickly as the system approaches a balanced state.

There is a second reason, why load proportional sampling is useful. Assume that we sample channels with some static probability which does not depend on its current load or utility. Then, there exists a channel which has sampling probability at most $1/m$. For this reason, any bound on the time to reach or even approximate a Nash equilibrium must be at least linear in m, since any ball may need expected time at least m to find a better channel. This is proven rigorously in [195].

Algorithm: Avoid_Contention

In the algorithm we used the primitive MEASURE_COST(j) to estimate the utility of the sampled channel j. We now assume that this primitive can only be applied to the channel currently used by the agent. Instead of comparing the utilities of two channels, the behavior of the following algorithm depends only on the utility of the currently used channel. The agent observes its own cost c_i and decides to move to another channel with a probability that increases with c_i. Again, the target channel is sampled proportionally to n_i. The algorithm, called AVOID_CONTENTION, is specified in a pseudocode in Algorithm 3.

Compared to COMPARE_AND_BALANCE, algorithm AVOID_CONTENTION is more energy efficient. Not only does it refrain from invoking MEASURE_COST on channels other than the one currently used, also the measurement of $\mathbf{n}$ by the primitive MEASURE_LOAD is executed only when the decision to switch to a new channel has already been made. In COMPARE_AND_BALANCE this measurement must be executed as the first step in each round.

The energy efficiency of the second algorithm comes at some cost. Whereas algorithm COMPARE_AND_BALANCE certainly stabilizes as soon as a Nash equilibrium is reached, algorithm AVOID_CONTENTION does not. At a Nash equilibrium, there will still be some fluctuations. However, the expected load vector that results from one round starting at a Nash equilibrium is at a Nash equilibrium again. To see this, consider a load vector $\mathbf{n}$. Since any ball in bin j migrates to

for all balls in parallel **do**
  $c \leftarrow$ MEASURE_COST($channel$)
  measure own cost c
  with probability c:
  for all channels $i \in [m]$ **do**
    $n_i \leftarrow$ MEASURE_LOAD(i)
  end for
  $n \leftarrow \sum_{i\in[m]} n_i$
  choose $j \in [m]$ randomly where $\mathbb{P}[j] = n_j/n$
  $channel \leftarrow j$
end for

ALGORITHM 3: AVOID_CONTENTION

bin i with probability $c_j(n_j) \cdot n_i/n$, the expected load n'_i of any channel i after one step is

$$\begin{aligned}\mathbb{E}[n'_i] &= n_i - n_i \cdot c_i(n_i) + \sum_{j\in[m]} n_j \cdot c_j(n_j) \cdot \frac{n_i}{n} \\ &= n_i - n_i \cdot c_i(n_i) + n_i \cdot C(\mathbf{n}) \\ &= n_i \ .\end{aligned} \tag{5.2}$$

The latter equality holds since at a Nash equilibrium, $c_i(n_i) = C(\mathbf{n})$ (or $n_i = 0$). It is easy to check that this property is not preserved if the sampling probabilities used by algorithm AVOID_CONTENTION are modified.

5.4.3 Analysis in the fluid limit

Let us consider the two algorithms described above for the case that the number of users $n \to \infty$, the so-called fluid limit. Rather than considering the number of balls n_i in bin i we now consider the fraction $x_i = n_i/n$ of balls in bin i. By the law of large numbers, as $n \to \infty$, we identify random variables with their expectation. The cost functions $c_i(\cdot)$ generalize naturally in this model by extending their domain to the interval $[0, 1]$. We consider the time derivative $\dot{x}_i$ at which the fraction of balls in bin i changes. For algorithm AVOID_CONTENTION we have seen in Equation (5.2) that

$$\dot{x}_i = x_i \cdot (C(\mathbf{x}) - c_i(x_i)) \ .$$

For the algorithm Compare_And_Balance the probability to migrate from a channel i to a channel j with $c_j < c_i$ is $n_j/n \cdot (c_i(n_i) - c_j(n_j))$. In the fluid limit,

$$\begin{aligned}
\dot{x}_i &= \sum_{j:c_j(x_j)>c_i(x_i)} x_j \cdot x_i \cdot (c_j(x_i) - c_i(x_i)) \\
&- \sum_{j:c_j(x_j)<c_i(x_i)} x_i \cdot x_j \cdot (c_i(x_i) - c_j(x_j)) \\
&= \sum_{j\in[m]} x_j \cdot x_i \cdot (c_j(x_i) - c_i(x_i)) \\
&= x_i \cdot (C(\mathbf{x}) - c_i(x_i)) \ .
\end{aligned}$$

Interestingly, we obtain the same expression for $\dot{x}_i$ for both algorithms in the fluid limit. In [194] it is shown that this dynamics converges towards a Nash equilibrium and reaches a state in which at most an ϵ-fraction of the agents differs by more than an ϵ-fraction from the average cost in time

$$\mathcal{O}\left(\frac{1}{\epsilon^3} \cdot \log\left(\frac{\max_{\mathbf{x}} C(\mathbf{x})}{\min_{\mathbf{x}} C(\mathbf{x})}\right)\right) \ .$$

In [195,196] this analysis is improved by taking into account the fact that information may be out of date by the time it is used. Thus, due to concurrent action of the agents, overshooting and oscillation effects may occur. Still, it can be shown that convergence can be guaranteed if the policy is executed slowly enough.

5.4.4 Simulations

Analysis of these algorithms in the fluid limit are somewhat optimistic due to the fact that the number of agents is finite. This introduces effects that are not present in the fluid limit. Berenbrink *et al.* [197] analyzed a similar algorithm in a discrete round based model with identical linear cost functions. Their analysis yielded an upper bound of $\mathcal{O}(\log\log n + m^4)$ rounds to reach a Nash equilibrium.

Convergence

Simulations show that also for other cost functions the algorithms converge very quickly. We have performed simulations of both algorithms for $n = 500$ and $m = 10$ using both linear and exponential cost functions. This testing was important since we wanted to ensure that the algorithms can support wide variety of different cost functions. We have performed a series of $10,000$ runs for each combination of algorithm and cost function type. Each run starts with a random assignment of agents to channels and consists of 15 iterations. The parameters a_i of the cost functions $c_i(n_i) = a_i \cdot n_i$ and $c_i(n_i) = a_i \cdot \exp(n_i \cdot (m/n))$ were chosen uniformly at random from the interval $[1, 10]$. For each iteration, we considered the random variables C_c and C_a, where C_c is the cost of a channel chosen uniformly at random and C_a is the cost experienced by an agent chosen uniformly at random. We use the standard deviation of these random variables as a measure of the balancedness of the system.

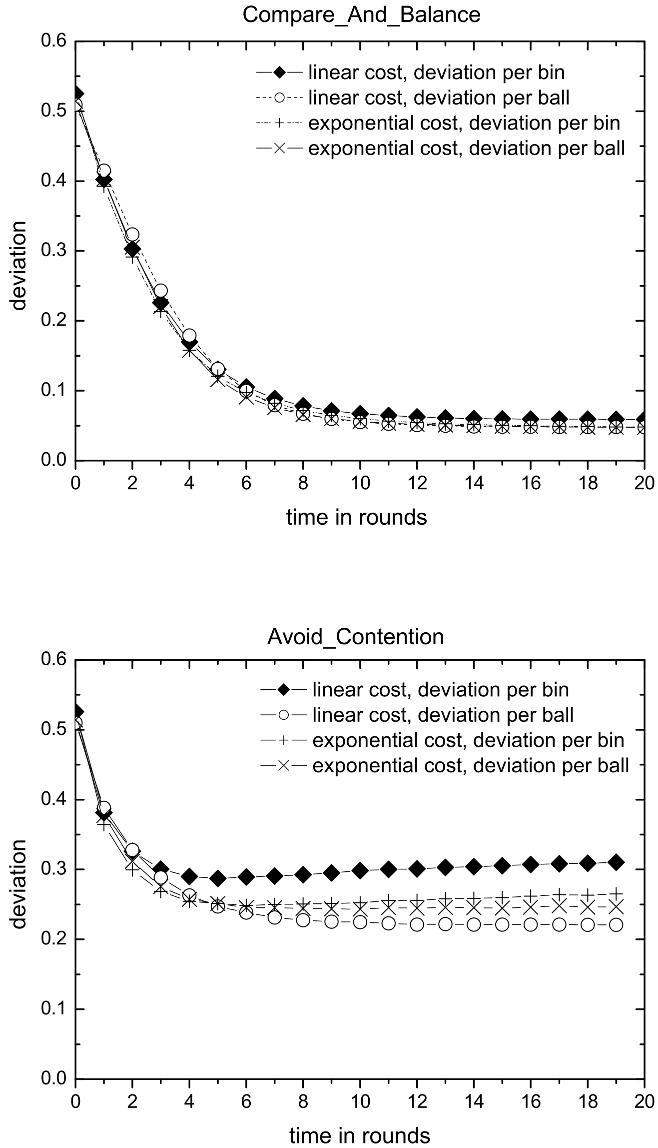

FIGURE 5.22: Relative standard deviation of the cost values for the algorithms COMPARE_AND_BALANCE and AVOID_CONTENTION.

Figure 5.22 shows how the standard deviation (after normalizing the expectation to be 1) decreases over time. We see that the minimum standard deviation is reached after as little as 6 rounds for both algorithms. As we expected, algorithm COMPARE_AND_BALANCE comes close to a Nash equilibrium. The deviation from the average cost decreases to below 6%. As we expected, algorithm Avoid_Contention performs worse. The expected deviation from the average cost decreases only to around 25%. This, however, is still by a factor of more than 2 better than the deviation of the random initial assignment.

Our simulations also suggest that the behavior of the algorithms is largely independent of the class of cost functions.

Dependence on number of users

The quality and convergence speed achieved by algorithm Avoid_Contention depends on the number of users in the system. As the number of users increases, the number of users utilizing a particular channel becomes more and more concentrated around its expectation[9]. This is illustrated in Figure 5.23.

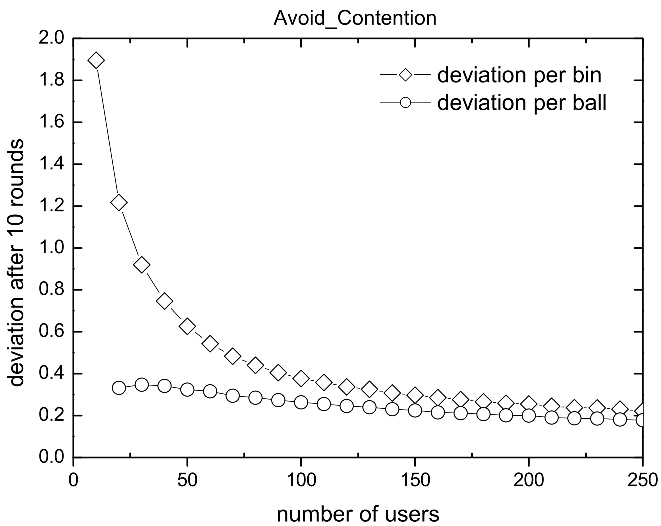

Figure 5.23: Relative standard deviation after 10 rounds depending on the number of users.

Again, we have performed 1000 runs with $m = 10$ and n varying between 10 and 250, i. e., the number of users per channel varying between 1 and 25. We see that, as n increases, the relative deviation from the average cost decreases.

[9]This is simple to understand based on statistical physics arguments. As the number of users increases for a fixed number of channels the system starts to converge towards fluid (gas) limit and deviations are purely statistical noise unless the condition of the channels are changing dramatically.

loop
  for all transmitters in parallel **do**
    Transmit
    After *interval* seconds query receiver for feedback
    Calculate cost c_i in current channel i
    if $c_i > C_{max}$ **then**
      With probability $p = \frac{c_i - C_{max}}{c_i}$ do
      1. Draw $j \in [m] \backslash i$ uniformly at random
      2. $i \leftarrow j$
    end if
  end for
end loop

Algorithm 4: The Threshold Protocol

5.5 The Threshold protocol

Now we will formalize the load balancing channel allocation approaches discussed above in a protocol, which we refer to as a threshold protocol and can be deployed in real network realization. Let us consider a number of communication transmitter-receiver pairs (Tx-Rx), which correspond to the numbers of balls n in our model. The cost assigned to each Tx-Rx pair is calculated as an inverse value of the throughput of the link between the communicating agents. The number of channels to be allocated is m. The maximum cost of communication on each channel is set to be C_{max}. The threshold algorithms works as follows: At the beginning all Tx-Rx pairs are assigned a channel randomly. Each transmitter is periodically calculating the current channel cost c_i and if it exceeds the threshold C_{max} the transmitter will decide to switch to another randomly drawn channel j with a probability p defined as follows:

$$p = \frac{c_i - C_{max}}{c_i}.$$

A pseudocode which describes the protocol flow is given in Algorithm 4. The algorithm is very simple and it ensures load balancing among the available channels in a distributed fashion. Because of its stateless nature, it adapts well to fluctuations of number of agents in time, and reaches stable state quickly. Since the algorithm does not require exchange of information among the Tx-Rx pairs, the generated overhead is minimal and limited to the feedback information needed from the receiver, so that the transmitter can calculate the cost c_i. As the algorithm runs in rounds a synchronization between the transmitters is needed in order to provide calculation of the cost almost at the same time. However, very accurate synchronization is not absolutely necessary. Small synchronization offset can be compensated with a moderately longer convergence time.

5.5.1 Performance evaluation through network simulations

In order to study the performance of the threshold protocol in more details, and verify the promising analytical results we have conducted a large number of network simulations. In the following we will describe the main results from the simulations.

Simulation setup

The simulations are carried out using the Qualnet Wireless Library for WLAN 802.11 [198]. The PHY is set to IEEE 802.11b either at 2 Mbps (DSSS, QPSK) or 11 Mbps (CCK) modes. In the application layer we run a constant bit rate traffic generator for UDP traffic at 1.95 Mbps or 10.9 Mbps, respectively. Necessary coding and modification have been made to the MAC in order to execute the *threshold algorithm*.

The transmitters and receivers are uniformly distributed in 500x500 m area. All nodes transmit with the same output power and they are all in communication range of each other. The simulations are carried out using two-ray propagation model [125].

We use four non-overlapping frequency channels (2.4 GHz, 2.45GHz, 2.47 GHz and 2.5 GHz) and vary the number of Tx-Rx pairs from 4 to 16. In all the simulations the radios have the same cost threshold C_{max} for a particular simulation setup. The threshold is chosen depending on the maximum number of radios a channel has to support given a certain number of channels and users in order to guarantee a balanced load over all available channels. For example if the number of available channels is 4 and the number of communication pairs 5, it is enough if the the maximum number of radios that the channel has to support is 2. This means that if we want to balance the traffic among all the existing channels, the maximum cost needs to match the performance of the most loaded channel and it should not be higher. The maximum cost is defined as the inverse value of the throughput and it is determined heuristically depending on the maximum realizable throughput for the number of users and channels in a particular simulation setting. For 2 Mbps links the maximum cost is calculated to be 0.008. It corresponds to a case when there are 4 channels and 4 Tx-Rx pairs so that each pair will get a channel. If the number of pairs is between 5-8, there will exist at least one channel which will have to serve 2 Tx-Rx pairs, so the maximum cost will be 0.016, etc. For 11 Mbps the logic of calculating the C_{max} is the same. Table 5.2 shows the values for different numbers of transmitter-receiver pairs. One should note that in this paper we consider a simplified case where all the channels are of the same “quality”. More specifically each channel is using the same channel model, and we consider only interference generated by IEEE 802.11 nodes. Any external interference or co-channel interference is not taken into account in the simulations. Obviously in a real setup some of the channels may suffer from external interference and this should be carefully considered in the modelling.

TABLE 5.2: Maximum costs set depending on the number of Tx-Rx pairs in the simulation scenario

No. of Tx-Rx	Max. No of Tx-Rx per Ch.	2 Mbps	11 Mbps
4	1	0.008	0.002
5, 6, 7, 8	2	0.016	0.004
9, 10, 11, 12	3	0.024	0.006
13, 14, 15, 16	4	0.035	0.008

The algorithm is iteratively run in rounds, each of them 10s long. This is done to ensure that simulations reach stable conditions, and no initial transient effects influence the results. In principle the interval can be shortened but we wanted to ensure that we get reliable results and avoid possible numerical artefacts. In every round, the successfully transmitted packets are counted. Based on this information each transmitter calculates the current cost in the channel c_i as the ratio between the round time and the number of successfully transmitted packets. If a switch of the channel is to happen the receiver will be informed of the new channel to be used.

Convergence statistics

Figure 5.24 below shows the average and standard deviation of the convergence time in rounds versus the number of Tx-Rx pairs for ten simulations. Quite naturally the convergence time increases with the number of users. Especially when the number of Tx-Rx pairs is multiple of the number of available channels the algorithm needs more rounds to converge since there are less combinations of valid channel allocations. However, one should note that the increase is mildly higher for the number of users we are considering. The standard deviation of the convergence time is also larger for a large number of users. Increasing the number of simulation repetitions can make the error bars smaller. Similar results are observed when the simulations are run at 11 Mbps.

In Figure 5.25 we show example of a dynamic channel occupation in time for a simulation run with ten Tx-Rx pairs. The algorithm converges relatively fast to a stable and balanced state. It is easy to spot from the figure that before the algorithm converges, the channel allocation is random and no guarantees of the performance or QoS can be assured.

Number of frequency changes

The results of the average number of frequency changes per Tx-Rx pair for a 500s run are plotted in Figure 5.26 and Figure 5.27. As expected, there are more frequency changes for those configurations that take longer to converge. The number of channel changes is relatively small, due to the load balancing algorithm. A

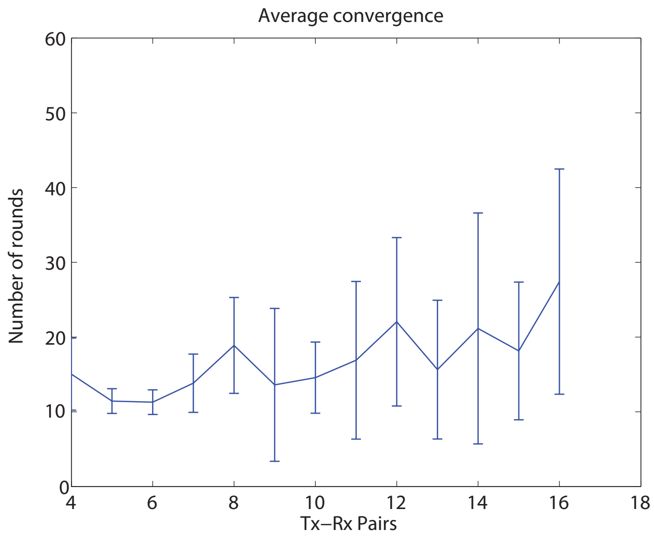

FIGURE 5.24: Convergence time versus number of Tx-Rx pairs with 2 Mbps links.

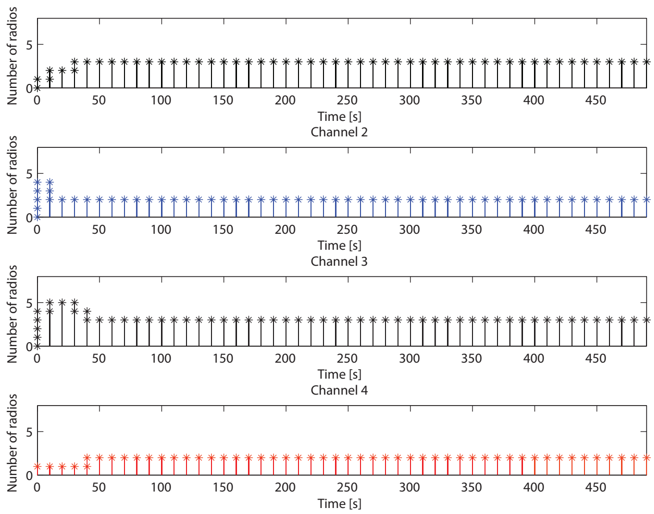

FIGURE 5.25: Channel occupation in time for a run with four channels and ten Tx-Rx pairs.

high probability to switch the channel occurs only when the maximum cost is exceeded by large margin. In the worst case (11 Mbps links and eight Tx-Rx pairs) the average number of channel changes for a 500s run is 2.5, which is a reasonable overhead for a realistic usage of the algorithm. For several configurations the average number of channel changes per radio is even lower than one per run. The reader should note that the perceived periodicity in Figure 5.26 and Figure 5.27 is not an artefact but is induced by the small number of channels the algorithm can use.

It must also be noted that these are the number of channel changes until convergence of the algorithm. After the algorithm has converged, a channel change should happen only if there are external changes, like the arrival of a new agent or interference from another system that affect the channel quality.

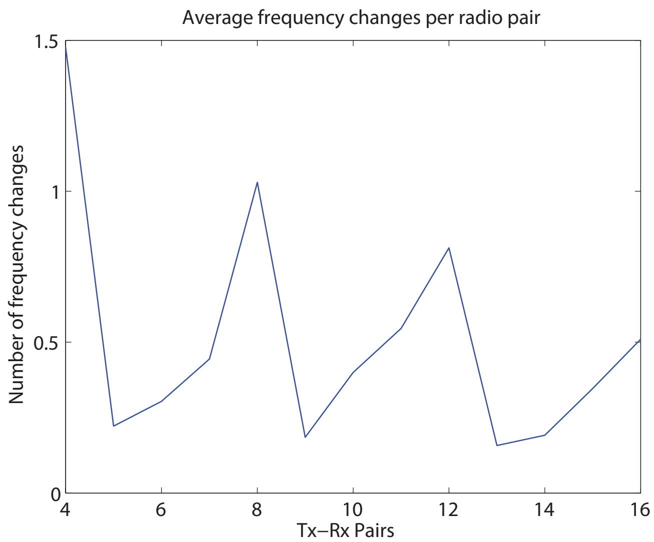

FIGURE 5.26: Number of frequency changes per Tx-Rx pair with 2 Mbps links. The figure shows the number of channel changes till the algorithm converges.

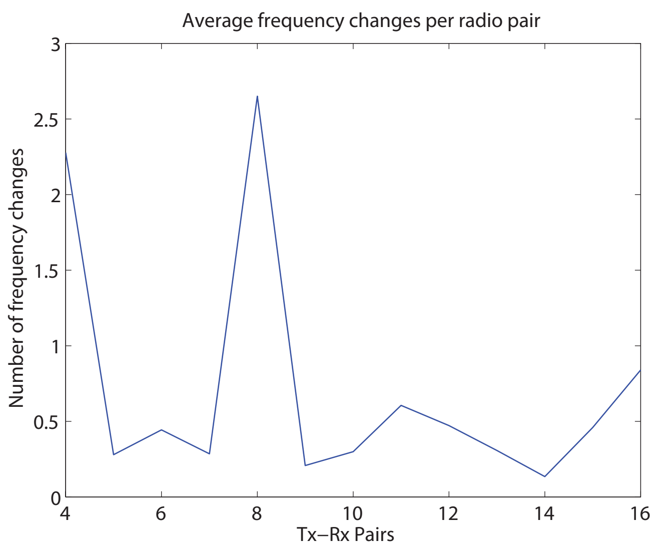

FIGURE 5.27: Average number of channel changes per radio pair for 11 Mbps links. The figure shows the number of channel changes till the algorithm converges.

In this work we have assumed that the cost threshold, C_{max}, is the same for all radios. This can be seen as a universal minimum utility that is required to keep a user happy. One could release this assumption and introduce user specific threshold $C_{i,max}$. Our expectation is that introducing threshold classes is better choice than allowing total freedom in choosing the threshold value, as it could lead to increased convergence time. The use of threshold classes is, however, open for classical problems known from QoS classes that in principle very easily all users can classify themselves as the highest priority. Thus, we left the discussion on adjustable thresholds out from the thesis.

Throughput and fairness

In addition to the convergence properties of the balls and bins algorithm we studied the average throughput over all existing links in the setup after the balls and bins algorithm has run. We compare the throughput results with the those obtained from a WLAN network with the same settings, in the same simulation environment and with the same node distribution in which the Tx-Rx pairs are assigned random frequency and do not execute any further optimization. The results for the minimum and average achieved throughput in both scenarios are shown in Figures 5.28 and 5.29.

We observe both in 2 Mbps and 11 Mbps modes that the average throughput is moderately higher if the balls and bins game is played. However, the performance gain in the worst cases simulated with a minimum achieved throughput is significant and leads to up to 60% increase in the throughput for a small number of Tx-Rx pairs in the 11 Mbps mode.

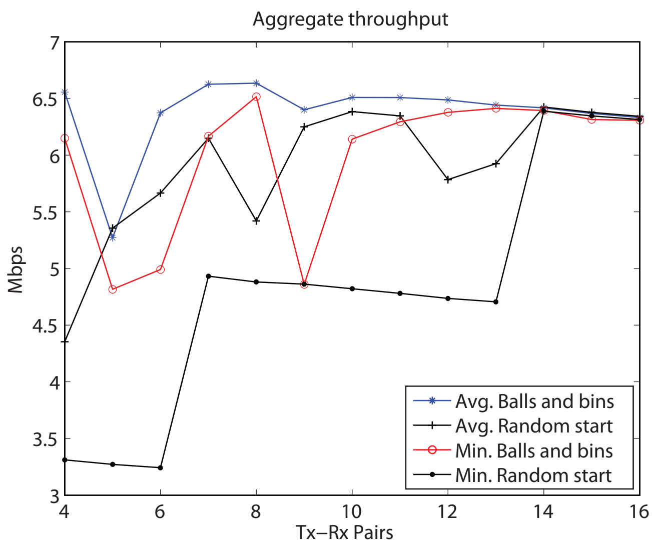

Figure 5.28: Average throughput for 2 Mbps links.

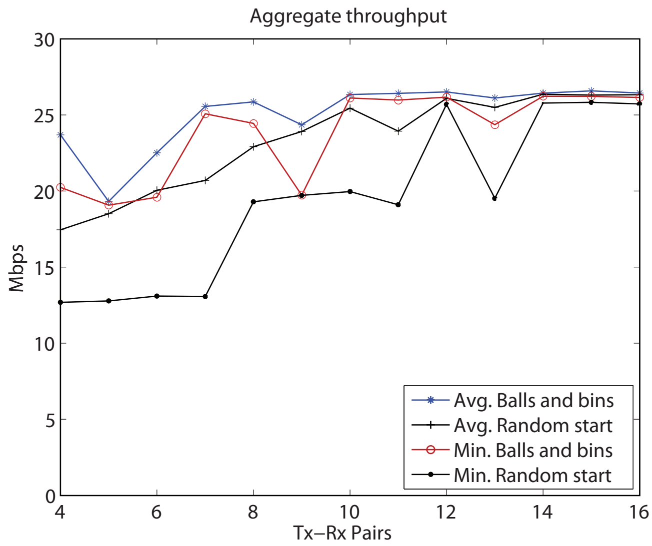

FIGURE 5.29: Average throughput for 11 Mbps links.

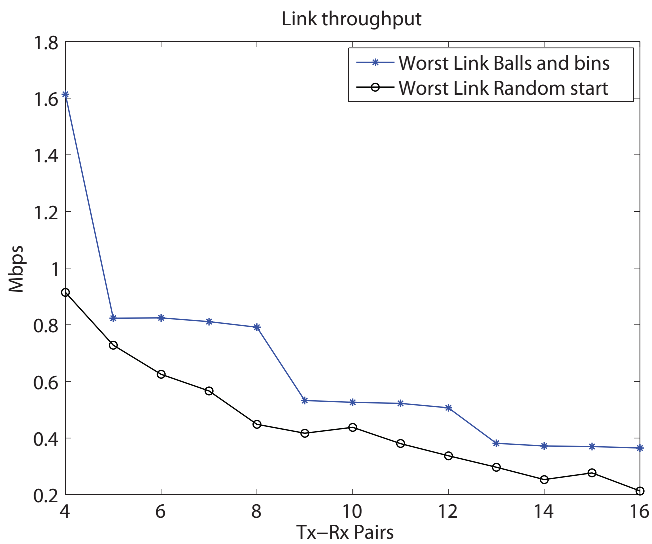

FIGURE 5.30: Average throughput of the worse 2 Mbps link.

Figures 5.30 and 5.31 show the throughput in the worst performing links for both channel allocation schemes at 2 Mbps and 11 Mbps. The balls and bins load balancing channel allocation ensures a throughput increase of 71% to 76% for the worst links.

The fact that balls and bins algorithm is also increasing fairness, especially ensuring that there are no big losers in the system, might in the first though look surprising. However, this is one of the strong benefits of the stochastic load balancing algorithm like balls and bins. The protocol guarantees a certain

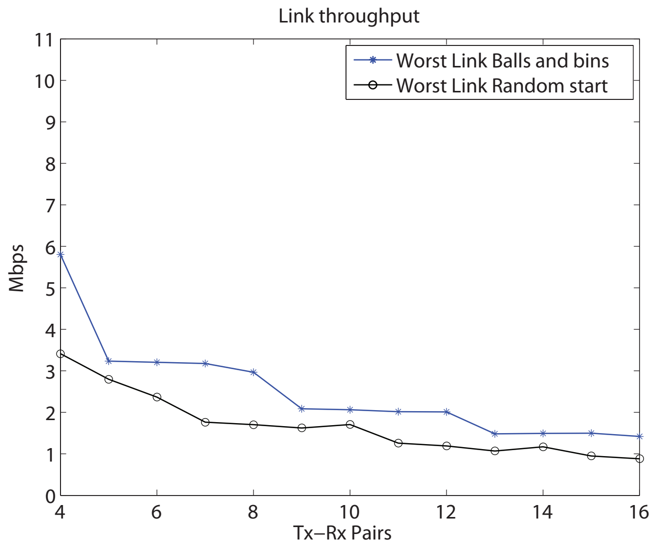

Figure 5.31: Average throughput of the worse 11 Mbps link.

minimum performance to all the links in the network. This is due to the maximum cost associated to each agent. If all the transmitters have the same cost threshold and there exists a suitable link allocation for the cost threshold set then fairness in the convergence state is guaranteed. The algorithm is essentially ensuring that any user experiencing very bad payback from some channel will migrate to the new one. Moreover, the distributed nature of the algorithm removes some of the anomalies the centralized protocols generate.

5.5.2 Experimental evaluation on a SDR GNU platform

In order to find out if the theoretical convergence estimates hold in a setup with realistic channels and to estimate the implementation complexity, we have chosen to implement the algorithm as a part of our more general cognitive radio research. Our cognitive wireless network testbed has been implemented using the GNU software defined radio toolkit [14] and the USRP boards. All the baseband processing including modulation and demodulation is performed on a host CPU using the GNU Radio libraries, while the USRP devices handle up and down conversion, decimation and interpolation on its FPGA. We are using the RFX2400 daughterboard in the 2.4 GHz ISM band. The respective hosts are commodity computers with Intel Core 2 Duo processors at 2.0 GHz and using 2 GB of random-access memory. They are all also connected via an Ethernet hub and synchronized for simplicity using the Network Time Protocol (NTP), another possible approach would be to implement synchronization over the air.

The physical layer uses Gaussian Minimum Shift Keying (GMSK) reaching 500 kbps bitrate. The frame structure contains a 64-bit pseudo-noise sequence for synchronization, a field for the payload length, a sequence number and destination identifier. The MAC layer protocol is carrier sense multiple access with collision avoidance and there is no acknowledgement-and-retransmission mechanism implemented. We use a TAP virtual network interface in order to tunnel the application packets into the GNU Radio PHY and `iperf` [188] to generate UDP traffic.

In our experiment cognitive radios communicate in pairs in an ad hoc fashion. There is a central node, but it is used only for logging purposes. In each round the GNU radios announce over the wired interface the occupied channel during that particular round, the sustained cost for this round, the decision to change channel for the next round and the channel to be used in the next round. This information is then stored in the central node. We are using a wired interface to collect this statistics in order not affect the measurement itself by using wireless channels.

Here we report results for measurements in which the duration of each round is set to ten seconds. At the end of each round, the receiver reports the number of successful packets received during the round. Then the transmitter calculates the cost experienced during the measurement interval and makes the decision whether to leave the channel or not and instructs the receiver to tune to the frequency chosen for the next round. This is done using a wired link in order to assure an error-free feedback channel. The feedback information on link quality could also be retrieved by the receiver as the number of acknowledged packets or using control frames on the air interface. The frequency changes could also be implemented using control frames on the air interface.

The channel allocations before the algorithm reaches convergence are random and there is no assurance of QoS levels until the algorithm has converged. As a result of that, in practical systems it would be beneficial to shorten the interval for evaluation of the cost at the start of the algorithm so that channel allocations are explored and the best solutions are found as quickly as possible.

Our current implementation is aiming at to verify theoretical results on the number of iterations. Hence, we choose rather slow communication and evaluation parameters in order to focus on these aspects before developing full broadband testing capabilities. For getting stable statistics on a sufficient number of transmitted packets for the cost calculation it is necessary to have a long measurement interval as our link is slow. This does not, however, affect the generality of the results. The delays would be around three orders of magnitude shorter on a WLAN commercial hardware platform, yielding to much shorter convergence times. Ideally the measurement interval should be adaptive: shorter at the start of the algorithm when convergence has not been achieved and longer after convergence, to ensure that the situation remains stationary and the QoS requirements are satisfied. We will report on such an implementation in our future work.

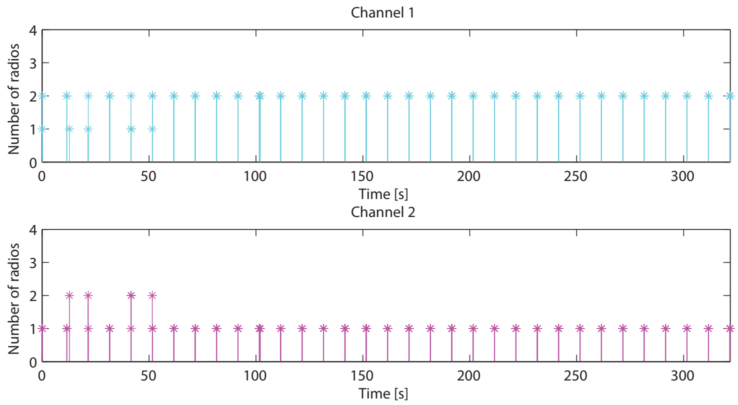

FIGURE 5.32: Channel occupation for a run in test scenario 3.

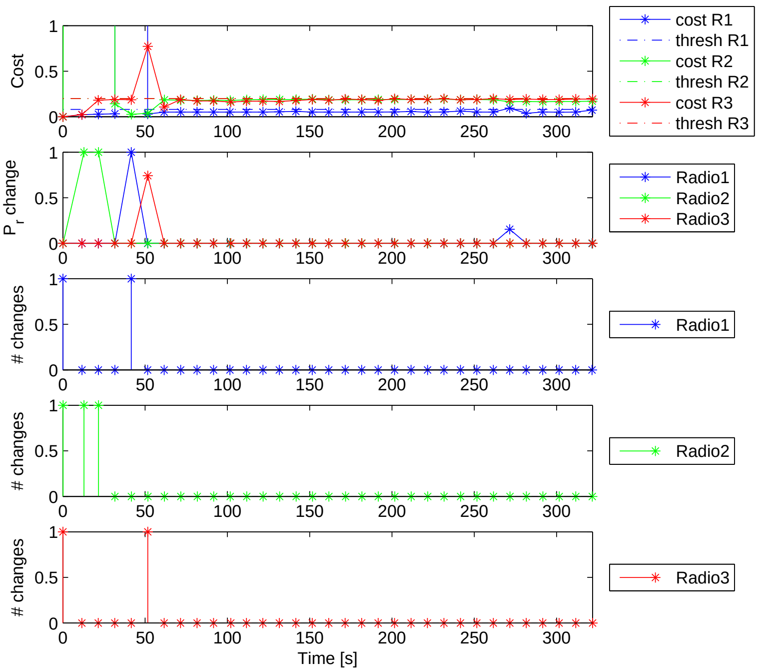

FIGURE 5.33: Sustained cost, probability to change channel and channel changes per radio for a run in test scenario 3.

Experimental results

In order to assess the convergence time and robustness of the protocol, we have performed measurements in four different test scenarios. For each different test configuration, the convergence time has been measured for ten -300s long runs of the algorithm. A large set of measurement were made to ensure statistical validity of the results. The measurements have been performed in a static nonmobile environment, monitoring the spectrum and choosing the non-overlapping channels least prone to interference.

In all the measurements, the cost has been defined as the inverse of the effective throughput, that is the interval duration divided by the number of successfully received packets during this time.

Our test cases are chosen so that they characterize the behaviour of the algorithm and its ability to provide advantageous and robust channel allocations for various traffic patterns. The QoS requirement, expressed as the cost threshold C_{max}, needs to be set accordingly to the generated traffic pattern. The algorithm has been evaluated in the following four different test configurations:

1. Three channels available for three Tx-Rx pairs transmitting at the same constant traffic rate. All the transmitters generate constant traffic at 200 kbps with packet size 1500 bytes, which results in 16.67 packets/s. Their cost threshold is set as $C_{max} = 0.08$ (equivalent to 12.5 packets/s). In this scenario, the protocol always converges to the optimal channel allocation, in which each of the users has a channel for its own. This minimizes the probability of collision and the packet error rate, guaranteeing that the cost threshold and therefore the QoS for the application is satisfied. The algorithm converges in average in 7.5 rounds (75s for our software-defined radio implementation) with a standard deviation of 5.3 rounds (53s). One should note that for the evaluation and verification of the algorithm the rounds are the effective metrics. We have also given the actual measured times for completeness, although those have only particular implementation related meaning.

2. Two channels available for three Tx-Rx pairs transmitting at different constant bitrates. One of the transmitters keeps the above settings and two of them are transmitting at 50 kbps (4.16 packets/s) with cost threshold $C_{max} = 0.24$ (4.16 packets/s). In this case, the optimal channel allocation would be to give the most intensive spectrum user its own channel and let the other two share the remaining channel. In our experiments, the algorithm converges to this allocation 70% of the times, while in the rest of the cases the convergence state is such that the intensive user shares a channel with one of the low-profile traffic users, but they both keep their cost threshold satisfied. The convergence time is 8.27 rounds with 5.62 rounds standard deviation.

3. Two channels available for three Tx-Rx pairs transmitting one of them at a constant traffic rate and the other two with a bursty traffic pattern. One of the transmitters has the same traffic and cost settings as in 1, the other two show a bursty traffic pattern, in which the ON state lasts 2s at a traffic rate of 200 kbps and the OFF state lasts 5s, with a maximum sustained cost of $C_{max} = 0.2$ (5 packets/s). The optimal channel allocation is the same one

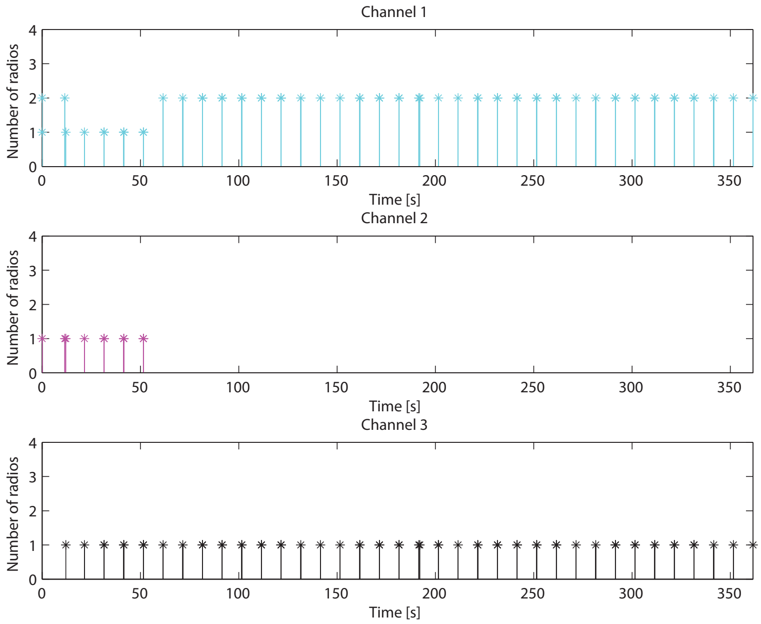

Figure 5.34: Channel occupation for a run in test scenario 4

as in 2 and it is reached in 90% of the runs with an average convergence time of 5.58 rounds and 4.31 rounds standard deviation.

Figure 5.32 shows the channel occupation and Figure 5.33 shows the local information of each of the links during one of our test measurements in Scenario 3. Transmitter 1 (Radio 1 in the graphs) is the intensive user, transmitters 2 and 3 (Radios 2 and 3) have bursty traffic.

All the radios start at the same time. The starting channel is chosen at random from the two available channels. The specific example in Figure 5.33 shows also that the measurement of throughput may sometimes introduce a transient instability which slightly slows down the convergence.

4. In this scenario we have three available channels and let the protocol converge to a stable channel allocation. Then we generate a jammer signal in one of the occupied channels and wait 300 seconds for the algorithm to converge again. The jammer signal has the same bandwidth and transmitted power as the transmitters in the testbed. The convergence time before the jammer is on average 5.27 rounds with a standard deviation of 3.22 rounds. After the jammer has started, a new convergence state needs to be found. In 70% of the tests performed, the algorithm converges to the optimal state in which two less intensive spectrum users share a channel and the most intensive user is on its own in 9.66 rounds with a standard deviation of 9.3 rounds.

   Figure 5.34 and Figure 5.35 represent respectively the channel occupation

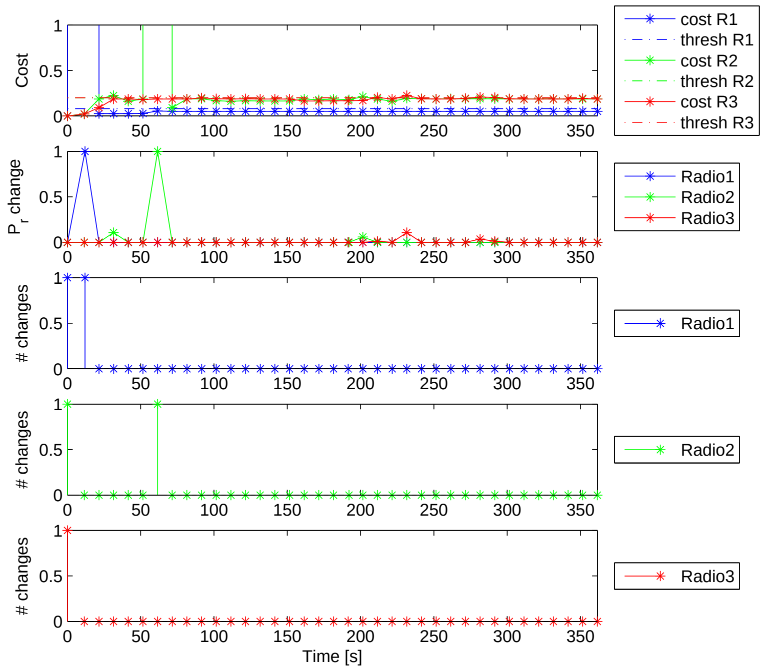

Figure 5.35: Sustained cost, probability to change channel and channel changes per radio for a run in test scenario 4

and local information of each of the links during a representative test measurement of Scenario 4.

The presented algorithm is shown to converge to the equilibrium and it is quite stable in all four scenarios. For more dynamic environments it can be stabilized with different approaches. Our current main implementation relies on the prudent choice of threshold parameter and periodic evaluation of a cost function. The algorithm can be made more responsive by increasing the evaluation frequency or lowering the threshold value. The frequency of cost function estimation and possible change of frequency after that can be interpreted also as an hysteresis timer, which prevents system to switch channels too rapidly and too often. Another possibility would be to allow the system to quickly estimate the costs, but introduce an hysteresis time which is started after equilibrium is reached. This would guarantee that the system stays a certain time in a fixed solution before it has a possibility to search a new configuration.

5.6 Summary

In this chapter we have presented two different methods and specific algorithmic solutions for dynamic channel allocation. The first method, DSATUR based coloring algorithm, was shown to be able to work both in centralized and distributed fashion. Moreover, the algorithm has low complexity and memory consumption and has proven to execute very fast. This work was one of the first published graph coloring based approaches to solve the dynamic channel allocation problem for IEEE 802.11 WLANs. There has been quite much of subsequent work, both derivative and novel done by other groups. In the second approach we use so called balls and bins potential game to model a dynamic channel allocation in wireless networks. The proposed solution has very attractive convergence properties and achieves load balancing among the users in the network by using the network and/or environment condition information in the channel assignment. For both approaches we have been able to conduct a full research cycle: starting from the research idea and initial theoretical analysis, progressing to simulations and finally concluding with real implementations and testbed experiments using WLAN equipments and SDRs.

6

Self-organization through Minority Game

In this chapter we propose a novel self-organizing resource allocation scheme with minimal feedback information based on minority game (MG). Designing wireless protocols with minimal feedback is especially important, because the price for information exchange in wireless networks is high and reduces the overall capacity of the system. This is especially true for networks that may rely strongly on self-organizing communication schemes, such as wireless ad hoc networks and cognitive radio networks. Minority games have a number of suitable mathematical and self-organization properties, and we believe that those can be successfully used to solve a wide spectrum of communication problems. In this thesis we apply a MG resource allocation scheme to schedule user transmissions in the network. We combine our MG scheduling model with CSMA (Collision Avoidance Multiple Access) protocol in order to minimize the number of collisions in dense wireless networks. Simulation results demonstrate that MG enhanced version of the CSMA protocol achieves higher channel utilization compared to a standard slotted CSMA without causing serious additional overhead in terms of delay. The MG-slotted CSMA is shown as a paradigm of how self-organized, game based approaches can be used for distributed resource allocation in ad hoc, mesh and cognitive radio networks.

6.1 Introduction

Self-organization is a crucial feature of the cognitive networks and has a great potential to enable cost effective and reliable maintenance and management of high-density and high-complexity networks. In Chapter 4 and Chapter 5 we discussed algorithmic and protocol solutions, which can be used to achieve a certain degree of self-organization in cognitive networks. We proposed techniques for automatic PHY and MAC parameter (re)configuration and presented both graph colouring-based and game theoretical approaches for modelling channel allocation. Having self-organizing network, however, means also adding more complexity into the system. Furthermore, distributing the coordination and optimization of the system is not without a price. Although network self-organization through cooperation among the users, offers flexibility and efficiency, it is expensive because of the information exchange overhead. In the case of wireless networks this means waste of the available communication bandwidth, which is anyway limited, drain of battery power and decrease of the system capacity. Hence, there is significant benefit from designing distributed minimal-feedback information protocols and algorithms. Recently there has been also renewed interest to provide mathematically more tractable analysis methods and metrics to quantify trade-offs caused by information exchange. In that respect, we have introduced *value of perfect information* metric that is based on expectation value analysis, and can be also combined with Bayesian statistics [10, 30]. At the same time the authors of [11] were introducing a complementary statistics called *price of ignorance*, which is based originally to game theoretical methods.

In this chapter we show that self-organization and cooperation can emerge even through minimal and local information content. We apply minority game (MG) in context of channel access to schedule the users' transmissions in the network and minimize the number of collisions in dense networks. We use IEEE 802.15.4 CSMA protocol [199] as a baseline for our performance analysis.

We adopt a minority game to distribute the users in two consecutive blocks of time slots, which we term *frames*, instead of letting all users to compete for the channel simultaneously[1]. The decision in which frame to transmit is based on the outcome of the minority game, which is played by each user independently. In each round of the game the user will decide to send its data in the current frame or wait until the next one, following its strategies. The users rank their strategies according to the past experience and no further coordination takes place among the users. Our simulation results show that MG distributes the users to access the wireless channel in a fair and balanced way in the time slots and due to this a higher channel utilization is achieved without a big delay overhead.

It is well known that various splitting or clustering algorithms applied to modelling of the user access can enhance the performance of Medium Access Control (MAC) protocols, especially in the highly congested environment [200]. Here, we propose a modification to the traditional slotted CSMA (Collision Avoidance Multiple Access) protocol, which allows for higher channel utilization and fewer collisions among the users transmitting in a certain time slot by automatically allocating users to different time slots. The main purpose of this work is to show how MG can be used as a self-organizing allocation and splitting algorithm in this domain. Although similar kind of performance could be also provided by classical approaches, minority game offers a distinct benefit for self-organizing with minimal feedback information. The minimal feedback not only lowers the overhead, but can also reduce the system complexity and still increase the robustness. As will be discussed below, the randomized strategies will converge to stable solution without synchronizing the whole strategy space between competing users.

[1]We are using two frames without losing generality to extend MG to N-frames systems.

6.2 The Minority Game

Originally, the minority game was mathematically formulated in 1997 by Challet and Zhang [201]. Minority game was inspired by the well known *El Farol bar problem* posed by Brian Arthur in 1994 [202]. Minority game is a model of repeated games where agents select one of two possible decisions in each round of the game according to their strategies. In each round the minority wins the game and the winning strategies are awarded for predicting the winning decision. In economics, MGs and their variants are well known paradigms and are used as market toy models, where each player aims at maximizing the profits from the market. Apart from the economists, minority games have received a fair amount of attention from other communities such as computer scientists and physicists for modelling the dynamics of complex systems. Hence, numerous studies on the dynamics of the minority games have been done and attempts to analytically solve those using methods from statistical physics have been performed [203–206]. Since the initial formulation of the minority games, a great amount of work has been done to make the models more realistic and close to financial markets and physical models. Adding more features in the games has resulted in number of variations of the traditional MG, such as evolutionary minority game, thermal minority game, etc. For the purpose of our analysis we will use the traditional MG formulation, discussed bellow. For further details about minority games including mathematical analysis and applications the reader is referred to [207] and the references therein.

6.2.1 The El Farol Bar problem

The El Farol bar problem was for the first time introduced by Brian Arthur in 1994. In the description of the problem Arthur takes the example of a popular bar in Santa Fe, New Mexico. The background story is as follows: One hundred inhabitants of Santa Fe decide, independently, each week, whether or not to go to the "El Farol" bar, where they know a good entertainment awaits them, if not too many people turn up at the same time. The only information available to every potential bar visitor each week is the number of people that went to the bar in the previous weeks. Each individual goes to the bar if she expects fewer than aN, where $a < 1$ people (60 people) to show up, and stays at home, if this is not the case. This means that the agents can act in two different ways: either go to the bar if they expect less than aN visitors or stay at home if the expect the bar will be overcrowded. This problem is clearly very complex as one has to model a social system beahviour. In order to model the problem, each individual needs a method to forecast the attendance and decide whether or not to go to the bar. However, one important feature of the problem is that no forecasting rule of deterministic attendance can be at the same time correct and available to all agents. In that context, the El Farol bar problem is an example of a inductive reasoning where the agents apply bounded rationality. The rationality is bounded since the agents do not cope with the complexity of the situation. However, the agents adjust their decision to go to the bar or not based both on the experience how the other agents have been behaving in the past and the expectation on what the other agents are going to in the next run[2].

Let us assume that the forecasting rule asserts that aN or more individuals will show up at El Farol bar. Then nobody will go, thus invalidating the forecast. Similarly, if the forecasting rule asserts that less than aN individuals will show up at the bar, then everybody will go, which again invalidates the forecast. To avoid such outcomes each agent is assigned a set of *predictors* or hypotheses that map the past several weeks attendance into the next week's. We list here several examples of possible predictors:

1. Predict next week's attendance to be the same as that of the previous week.

2. Predict an average of the last four weeks.

3. Predict the same as 2 weeks ago.

The main idea behind the predictors is to give the agents the possibility to decide upon experience using the found patterns from the past situations. Each agent monitors its patterns and evaluates them each week by giving scores to those which have predicted going to the bar. Each week the agents uses the predictor with the highest score. Simulation studies have shown that the attendance in the bar fluctuates around aN [208]. This proves that a system with reach dynamics manages to self-organize itself to a stationary state.

The El Farol bar problem has been also interpreted as a market entry game [209]. Market entry games can be seen as two-stage games. In the first stage, players simultaneously decide whether to enter the market or not. In the second stage, the payoffs of the players who entered the market will be determined according to their actions on the market in a continuous way. The difference In the actual El Farol bar problem is that the payoffs of the people attending the bar are determined in a discontinuous manner. Obviously, the El Farol bar can be also seen as a kind of congestion game as agents' payoff depends on the number of attendees in the El Farol bar. Therefore, it can be modelled according to [191], similarly to the balls and bins problem we described in the precious chapter.

The El Farol bar problem is an example of a minority rule driven social system. There exist many other examples of such systems, one being the traffic on the roads where every driver prefers to be on an uncongested road. In communications, typical example is a transmission scheduling where users aim to transmit

[2]The fact that the minority game an be interpreted as a classical inductive reasoning problem links together machine learning and statistical physics as well as applied probability theory and game theory. For the work reported in the beginning of the thesis, the minority game is one suitable method for CRM (see Chapter 3 for details) to use, and shows clearly that not all learning methods used in the cognitive networks need to be heuristic or classical AI based.

in the less crowded channel. In this work we deploy minority game to achieve self-organizing scheduling of users' data transmission in the MAC and increase the channel utilization with minimal overhead. Prior to defining our problem we will give a brief introduction to minority games in the next section.

6.2.2 MG formulation

Let us now define El Farol bar problem more formaly as minority game. An odd number of N players[3] decides in each round of the game between two alternatives "0" or "1". We denote the ith player's *decision* or *action* at time t with $a_i(t) \in \{0,1\}$. The outcome of the game is simply a mapping of all players' decisions to a binary value indicating which group has won. In each round, the players in the minority group win and will be awarded. The total action of all the players at time t of the game is called *attendance* $A(t)$ and can be defined as

$$A(t) = \sum_{i=1}^{N} \{2a_i(t) - 1\}\,. \tag{6.1}$$

The winning group $W(t) \in \{0,1\}$ at time t can be expressed as

$$W(t) = \mathcal{H}\left[-A(t)\right], \tag{6.2}$$

where $\mathcal{H}$ is the Heaviside function. Each player bases the decision about the action in the next round on the last m outcomes of the game, i.e. $\{W(t-m),\ldots,W(t-1)\}$. Thus, all players share the *global information* $\mu(t) \in \{0,\ldots,2^m-1\}$, which is the decimal representation of the binary vector of the last m outcomes of the game. At the end of each round, the players in the minority group gain one point, the others loose one point. The way how the players decide on their actions is determined by the strategies they adopt. A strategy is simply a mapping of $\mu(t)$ to an action $a(t)$. Each player possesses a set of S not necessarily different strategies which are randomly drawn from a pool of strategies at the beginning of the game. We denote by $s_i, s \in \{1,\ldots,S\}, i \in \{1,\ldots,N\}$, the strategy s of player i. Since there are 2^m possible histories $\mu(t)$, the strategy space comprises 2^{2^m} different strategies. In order to decide which of the S strategies to play in the next round, each player keeps track of the *virtual* points each of its own strategies would have gained if it was adopted from the beginning of the game on. Hence, the virtual score $U_{i,s}$ of the strategy s of player i can be updated by

$$U_{i,s}(t+1) = U_{i,s}(t) - \mathrm{sign}[(2a_{i,s} - 1)A(t)]. \tag{6.3}$$

The player then chooses the strategy with the highest virtual score. In case two or more strategies are tied, the player makes a random selection.

[3] We will be using the terms player, agent and node interchangeably.

It is maybe worth to explicitly mention that MG is an adaptive game and the players change their preference of using certain strategies in time based on the past experiences made in the game. The players *do not* have the complete knowledge of the system and neither they know the best global strategies with the highest virtual scores, nor the total number of players in the game. However, the players are considered to be inductive and take decisions according to the best of their knowledge represented by the limited number of strategies in their hands.

6.2.3 Phase transition, volatility and predictability

Due to the minority rule in the game, in each round, the number of winners is always smaller than the number of losers. Averaged over time, the attendance $A(t)$ has a value of $N/2$ if "0" and "1" are adopted as actions. In order to understand the dynamics of the minority game and the self-organizing behaviour of the players it is interesting to observe the variance of the attendance, $\sigma^2 = \langle A^2\rangle - \langle A\rangle^2$, also known as *volatility*. Savit et al. [210] found out that the volatility is only a function of the ratio $\alpha = 2^m/N$. Although it was later suggested that α should be $2^{m+1}/SN$ for $S \geq 2$, the qualitative behavior does not change with the number of strategies each agent hold. Figure 6.1 shows the normalized volatility σ^2/N for different number of players as a function of α for $S = 2$. The volatility value of one marked with the solid line represents the case when the agents are making random decisions in each round of the game (coin-toss). For smaller values of α the value of the volatility is larger than the coin-toss limit and the game is said to be in a so called worse-than-random or crowded regime. In this regime the number of agents is large, their actions start to be correlated and the size of the losing group is much larger than $N/2$. When α increases, the volatility starts to decrease (up to a certain value of $\alpha = \alpha_c$) and the game enters the better-than-random regime. Here the coordination among the agents is improved and they perform better than just playing randomly. For $\alpha = \alpha_c$ the volatility attains the minimum value and at this point a transition between a *symmetric* and *asymmetric* phase of the game takes place. If the value of α increases further, the volatility increases as well up to the point of reaching the coin-toss limit again. Let us now shortly discuss the symmetric and asymmetric phase of the minority game. Analyzing the winning probabilities for a particular action, Savit at al. concluded that for $\alpha < \alpha_c$ it is not possible to find out from the last m bits of the history, that one of the two possible actions was more probable to win. This is so because both actions have equal probability of winning the game. Hence, this phase is also called *unpredictable*. For values of α beyond α_c there is an unequal winning probability for the two actions when analyzing the last m history bits. This is known as asymmetric or *predictable* phase of the MG.

Nevertheless, we would like to stress that coordination in the MG does not actually come through exploitation of the history, i.e., previous outcomes of the game. In the initial formulation of the MG the decisions taken by the players were

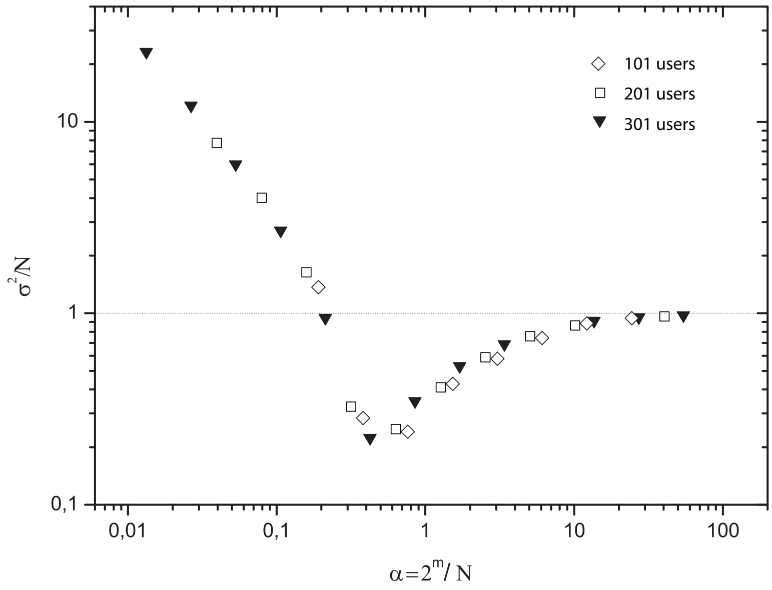

Figure 6.1: Global efficiency of the standard MG. The variance σ^2/N is shown as a function of $2^m/N$ for $S = 2$, for different number of players. The symbols represent $N = 101, 201$ and 301.

indeed based on the last m outcomes of the game, as described in Section 6.2.2. However, later on Cavagna [211] revealed a surprising fact that if the history is replaced by a randomly selected string the behaviour of the MG remains the same regarding time averaged quantities, provided that all the agents utilize the same piece of information in each round. Thus synchronisation of fictitious history is enough to ensure equilibrium for minority games instead of constant feedback of outcomes.

6.3 MG-Model for Channel Access

We will now describe on of the self-organizing protoocls we have developed based on minority games. The aim of using minority game together with Carrier Sensing Multiple Access (CSMA) is to reduce the number of collisions by limiting the number of nodes that may contend concurrently for the channel. We call this scheme MG-slotted CSMA.

We consider that the time is divided into *superframes* and each superframe is subdivided into two equal *frames*. All nodes in the network are the players of an MG, and each superframe corresponds to one round of the game. Essentially, as discussed above, in each round of the MG players are split into two groups. Each node may contend for the channel only within one frame of each superframe, selected according to the node's decision in the current round of the MG; during this frame, we will be addressing the node as *active*. For the remaining frame of the superframe the node is *inactive*. While a node is inactive, it may not gain access to the channel for transmitting, but it may receive packets. Specifically, nodes that belong to the group formed by decisions $\alpha_i(t) = \{0\}$ are considered to be active during the first frame, and nodes in the other group, with decisions $\alpha_i(t) = \{1\}$, are active during the second frame.

During a frame active nodes may allocate the channel by a simple slotted CSMA scheme with Binary Exponential Backoff (BEB). The details follow closely the algorithm as described in the IEEE 802.15.4 standard [199]. A node that has a packet to transmit shall delay for a randomly selected number of time slots in the range 0 to $2^{BE}-1$, where BE is the value of the Backoff Exponent (BE). Then, it senses the channel for a period of two time slots. If the channel is assessed to be idle the node shall start transmitting at the beginning of the next time slot, while if busy, the node selects a new random backoff period after increasing the value of the backoff exponent by one, and starts the backoff counter immediately. For the first transmission attempt, the BE is initialized to a value set as minimum. If the number of selected backoff time slots is greater than the remaining number of time slots in the frame, the MAC shall pause the backoff countdown at the end of the frame and resume it at the start of the next frame within which the node will be active.

The receiver acknowledges the successful reception of data by transmitting an acknowledgement frame. An acknowledgement shall be sent directly upon reception of a packet without using the CSMA/CA mechanism, and no matter whether the node sending the acknowledgement is in active or inactive state.

When an active node gains channel access it starts transmitting regardless of whether its transmission will be finished within the current frame, even if the node will be inactive during the following frame. This last detail ensures that time slots at the end of each frame do not remain unusable.

As discussed above, it is not actually necessary that players take their decisions based on the last m outcomes of the game, but they may rely on random information provided that they share the same piece of information. We exploit this property towards eliminating the amount of feedback packets that need to be exchanged between nodes. In order to utilize the last m outcomes of the game as global information, each node should broadcast its decision so that all nodes could determine the final outcome of the previous round. Instead, we consider that a random string of x bits is generated once every x rounds of the MG by a randomly selected node, and each of the x bits is interpreted as a potential game outcome. Specifically, each node draws a random delay once every x rounds. The node with the smallest delay generates and transmits first the random string of x bits. Then, upon reception of the string the rest of the nodes defer and utilize the received string for playing the following block of x rounds.

In order to get an efficient outcome, in each round of the MG our aim is to divide the nodes into two almost equally sized groups. As discussed in Section 6.2.3, m appears in the control parameter of the MG, α, which influences the efficiency of the result. Therefore, we consider that an appropriate value for m should be selected, or calculated, so that the game is played in a low volatility region, i.e., with $\alpha \approx \alpha_c$ (see Figure 6.1). The value of m should depend on the number of nodes, N, and the number of strategies assigned per node, S, ($\alpha = 2^{m+1}/SN$). Analytical calculations have shown that $\alpha_c(S = 2) = 0.3374$. Therefore, the calculation of m is straightforward. We note that although an appropriate value of m depends on the number of nodes, moderate fluctuations do not influence significantly the efficiency of the game.

At each round of the MG a node updates the virtual points of its strategies by adding or deducting one point from the strategies which give a correct or wrong prediction respectively. Then, the subsequent decision of that node is determined by the strategy with the highest score.

It is interesting to mention that extensions of the original MG appear in the literature, which allow players to choose one out of many alternative actions, and not only one out of two [212]. Within the context of our work a *multiple choice minority game* can be used to extend our model in cases where the number of nodes is very large, or the total traffic load extremely high, in order to split the nodes into more than two groups and limit further the contention. For this purpose, a superframe would be comprised of more than two frames. In fact, the number of frames in a superframe should be imposed by the number of alternatives given in the MG. This extension is trivial, and the dynamics of the multiple choice minority games are well known.

6.4 Performance Evaluation

In order to evaluate the proposed MG-slotted CSMA we performed simulations using the Qualnet network simulator. We compare the performance of the MG-slotted CSMA with the standard slotted CSMA protocol in terms of throughput and packet delay. The standard slotted CSMA scheme we used for the comparison is implemented according to the CSMA/CA mechanism with BEB that we employed in the MG-slotted CSMA protocol.

6.4.1 Simulation Setup

All simulations were run 50 times with different random seeds using a simulation duration of 10 minutes. In this paper we present scenarios with a physical layer data rate equal to 11 Mbps, but we have also tested cases with 2 Mbps. The payload size at the application layer is 1500 bytes. In each scenario all the nodes contribute equally to the offered traffic by generating CBR streams for randomly chosen destinations in the network. Since the MG-slotted CSMA aims to give advantage to relatively congested networks, we consider cases with total offered traffic at least equal to the physical layer data rate, meaning normalized traffic load equal at least to 1.

Table 6.1: The number of clusters in the Thomas distributions.

Number of nodes	Number of clusters
51	2
75	3
101	3
125	5
151	5
175	5
201	6

Furthermore, we consider single-hop and multi-hop networks. In single-hop networks the distribution of the nodes does not have any impact to the results. Therefore, nodes are uniformly distributed within a square region of 150 m×150 m. For the case of multi-hop networks we consider more realistic clustered network topologies following Thomas distributions [213]. Thomas distribution is known to approximate the real world node distributions better than Poisson point distribution (uniformly random node distribution) [214]. Due to this we have chosen to use clustered Thomas distribution model, although we have also tested the algorithm with other distributions. The Thomas distributions are generated in a region with dimensions 850 m×850 m, whereas the communication range of each node is approximately 270m.

In a Thomas distribution, a primary Poisson distribution is first used to generate the initial point distribution for cluster centers. After this a bivariate Gaussian distribution with covariance matrix $\Sigma = diag(\sigma_\chi, \sigma_\psi)$ is used to generate the locations of actual points around the location of each cluster center. The parameters σ_χ and σ_ψ characterize the size and shape of the clusters; here we used $\sigma_\chi = \sigma_\psi = 60$. The number of clusters we generated differs according to the number of users; the values are illustrated in Table 6.1.

The time slot duration and backoff values are set according to the IEEE 802.11 standard. A time slot is $20\mu s$. The minimum backoff exponent is 5 and the maximum is 10, giving backoff values ranging from 32 to 1024 time slots. Furthermore, all communication between nodes is performed over a single-channel with a bandwidth of 20 MHz, in the 2.4 GHz frequency band.

Regarding the configuration paremeters of the MG itself, we present cases with 2 strategies assigned per node. Each strategy is a mapping of the global information to a decision (see Section 6.2.2). At the beginning each player generates randomly its own mappings, i.e., strategies. We conducted also simulations with 4 strategies per node but results are omitted since they almost identical. It is, however, known that qualitative results of the MG are independent of the number of strategies [207,208]. As for the feedback information in the MG-slotted CSMA, the randomly generated global information for the MG is distributed by a counter-based flooding scheme [215] once every 100 rounds of the game, which corresponds to a period of 3.2 s. Therefore, a sequence of 100 bits is disseminated through the network once every 3.2 s, which is definitely a minimal amount of feedback information for providing coordination between nodes.

6.4.2 Simulation Results

In this section we present part of the results we obtained for single-hop and multi-hop networks. Figures 6.2 and 6.3 show respectively the channel utilization and average packet delay in an 11 Mbps network with a normalized network load equal to 1, i.e., total traffic of 11 Mbps.

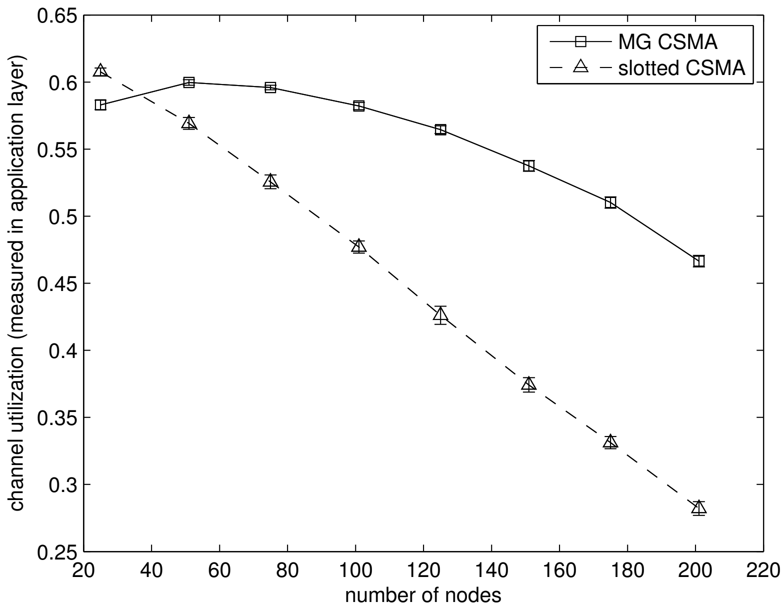

Figure 6.2: Channel Utilization of MG-slotted CSMA and standard slotted CSMA for single-hop network. The data rate at physical layer is 11 Mbps and the normalized network load is equal to 1.

The results illustrate that there is a significant gain in channel utilization obtained by the MG-slotted CSMA in comparison to the standard slotted CSMA. This emerges from the fact that MG-slotted CSMA decreases the number of concurrently contending nodes approximately by half. In a CSMA scheme the number of nodes has a large impact on the number of collisions, and hence on the channel utilization.

Furthermore it is important to observe that the proposed scheme does not introduce serious additional overhead in terms of delay, as one might expect due to the fact that nodes may access the channel only during half of the frames. In fact, with CSMA mechanism a node would probably backoff several times

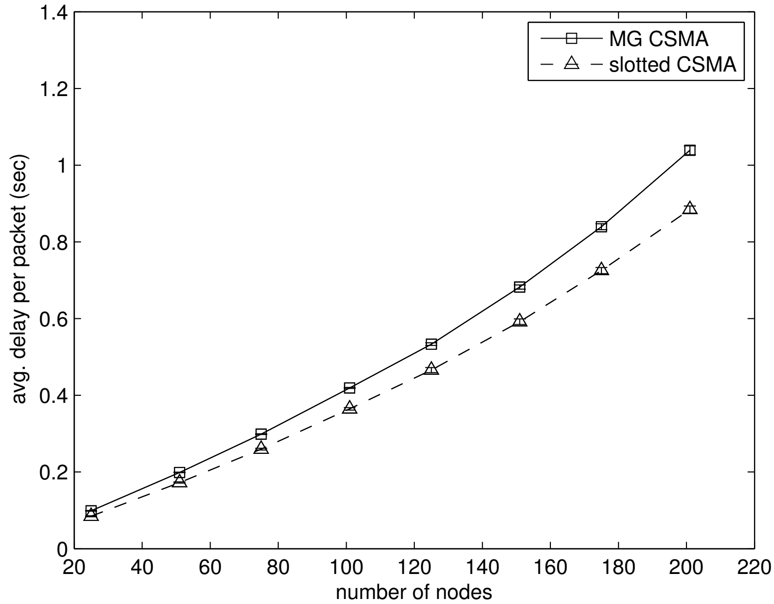

Figure 6.3: Average packet delay of MG-slotted CSMA and standard slotted CSMA for single-hop network. The data rate at physical layer is 11Mbps and the normalized network load is equal to 1.

before gaining channel access under high traffic conditions. In such situations the introduction of the MG halves the number of contending nodes and reduces the amount of instantaneous traffic load, and therefore lowers significantly the waiting periods caused by the backoff mechanism. Additionally, the number of collisions is also decreased; this leads to fewer retransmissions, and hence lower delays.

For the case of multi-hop networks we present results of simulations for a network where the capacity of the links is 11 Mbps and the offered traffic is 22 Mbps (see Figures 6.4 and 6.5). As in single-hop networks, the results verify that MG-slotted CSMA achieves better channel utilization compared to a standard slotted CSMA without generating serious additional delay. At his point we would like to note that the values of standard deviation appearing in Figures 6.4 and 6.5 are comparatively large due to the topology of the network. As already discussed, for the multi-hop scenarios we used clustered network topologies, that were randomly generated. As a result, the degree of connectivity varies significantly in each configuration. We have encountered cases yielding significantly lower throughput than others when the number of nodes that connect disparate clusters is very limited. Therefore, we would like to point out that increasing the number of simulated scenarios does not decrease the standard deviation. In fact, the variances would be much lower if we were using Poisson point distributions, which is typically used in many works as a starting point.

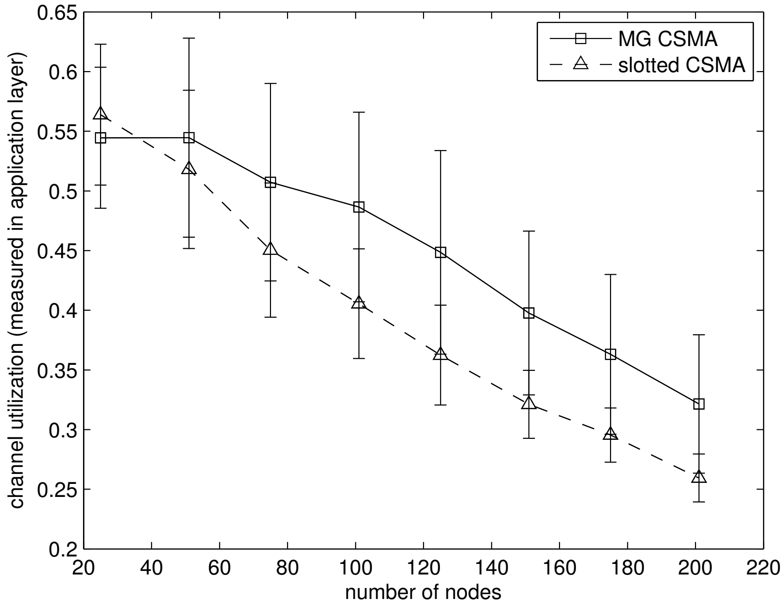

Figure 6.4: Channel Utilization of MG-slotted CSMA and standard slotted CSMA for multi-hop network following Thomas distribution. The data rate at physical layer is 11 Mbps and the normalized network load is equal to 2.

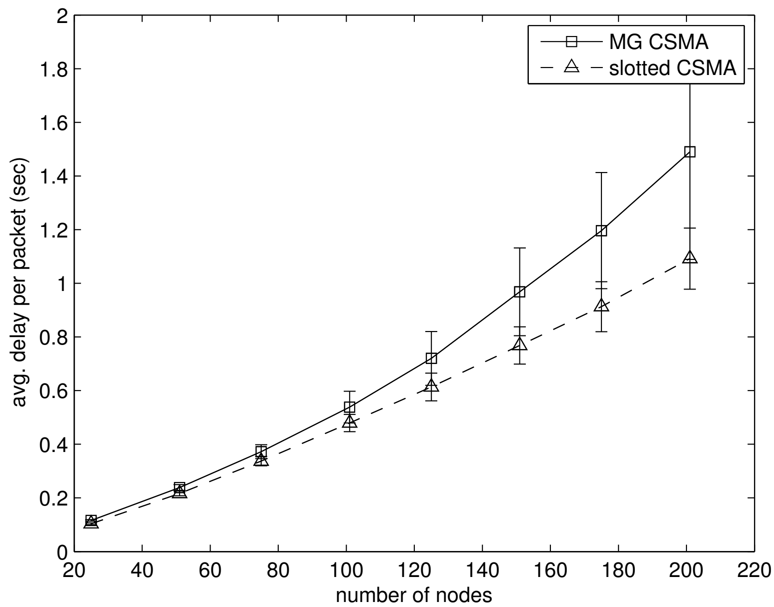

Figure 6.5: Average packet delay of MG-slotted CSMA and standard slotted CSMA for multi-hop network following Thomas distribution. The data rate at physical layer is 11 Mbps and the normalized network load is equal to 2.

6.5 Related work

Several other authors have introduced MG as a tool for modeling the user behavior in the network in order to address different problems in context of communication systems. In [216], Bell et al. have pointed out that a coordination failure among the users in the network can lead to serious problems such as congestion and therefore studied the potential of El Farol bar problem to address congestion and coordination problems which may arise due to overutilization of the Internet. In [217] the authors proposed a new TCP congestion control mechanism by employing MG to selectively reduce the transmission speed of the senders. Furthermore, in [30] the authors argue that resource sharing can be achieved without detailed information exchange or coordination among the users by applying MG in context of cognitive wireless networks. In addition, the paper shows that by using MG as adaptive resource sharing tool for channel access in slotted ALOHA the number of collisions can be decreased and higher throughput can be achieved. Recently, Mertikopoulos and Moustakas [218] have adopted multi-choice minority game to model the users' selection of access points in a heterogeneous network environment.

6.6 Discussion

We conclude this chapter by pointing out that minority game and its dynamical propoerties have also deeper connection with statistical physics and machine learning. The connection with statistical physics is quite obvious and similar behaviours can be observed even in physical systems, e. g., some of the self-organization properties of spin-glass have similarities. From the perspective of cognitive radio the link toward machine learning is even more interesting. In fact, it is possible to show that minority games can be interpreted, in most cases, as reinforcement learning algorithms in a sense of Erev and Roth reinforcement learning models [219]. The link comes through presenting a potential game for reinforcement learning and then showing that learning theory predicts sorting in the El Farol bar problem. Hence, we have provided yet another example of simple algorithm that can perform learning type of self-organization without excessive need for feedback. Through this approach it is quite straigtforward to show that minority games have Nash equilibria. In certain sense our minority game has some similarities of repeated evolutionary games we were using in Chapter 5 for load balancing. We plan to study these similarities analytically in the future work.

6.7 Summary

In this chapter we presented our work on applying minority game in order to coordinate the channel access of multiple users to a shared medium in a self-organizing fashion. Specifically, we introduced a paradigm of an MG enhanced version of the CSMA protocol. Minority games belong, in general, to the class of congestion games. This is the reason we initially considered them to be an interesting agent coordination possibility to handle resource allocation in distributed fashion for resource limited networks. Due to the fact that the players in an MG need to share only a minimal amount of information, we argue that MGs as self-organization schemes in ad-hoc networks are indeed a promising approach.

Simulation results show that the proposed MG-slotted CSMA leads to increased channel utilization. Especially in scenarios with large numbers of contending nodes and high network loads, the gain in channel utilization is considerable. Obviously we expected that the packet delay will be affected by playing the MG, since the nodes are allowed to gain channel access only during limited time intervals. However, the resulting graphs indicate that there is no significant additional overhead in terms of average packet delay in the network.

The main benefit of MG is that it is able to converge fast to an equilibrium solution with minimal external information. In principle, other multiple access schemes like, for instance, carefully adjusted p-persistent CSMA could achieve similar results in *selected* offered traffic ranges, but finding an optimal p value is more complex and information expensive approach than using MG as a control algorithm. Therefore, we believe that our concept of using MG in scheduling resource allocation is promising and could be used in a distributed fashion for wireless ad hoc and cognitive radio networks, not only in the context of multiple access, but also for various resource allocation problems. Also, as the additional computational complexity and required memory for running Minority Game is very low, it is suitable for large and dense networks.

As a future work we plan to extend our idea and employ multiple-choice MGs that allow multiple decisions to be made. Such games shall result in dividing the contending nodes into several groups and decrease further the amount of contention in a network. We believe that this concept may increase significantly the number of nodes that can be handled efficiently by CSMA-based or other existing multiple access protocols, and thus, even expand the capabilities of existing resource allocation mechanisms. Additionally, we plan to publish soon our on-going work which is comparing MG-based user allocation against the classical splitting and clustering algorithms, both analytically and by simulations.

7

Conclusions

The general theme of this thesis has been a study of different self-organization methods that can be applied to both cognitive radios and present day wireless networks. At the beginning we have presented a framework for cognitive resource management called CRM, which can serve as an architectural basic model for enabling efficient implementation of the cognitive cycle in the CRs and more adaptive and flexible use of the network stack for optimization purposes. The CRM itself has been extended beyond reference architecture towards specific reference implementations in several research projects during this thesis. In the later part of the thesis we have concentrated on selected problems such as automatic configuration of PHY and MAC parameters for optimized operation of OFDM-based cognitive radios, dynamic channel allocation and load balancing in cognitive radios and WLANs, and self-organizing resource allocation and scheduling in general with minimal feedback information. Apart of extensive analysis through simulations, a great part of the proposed solutions have been practically implemented and tested both with SDR platforms and commodity Wi-Fi hardware. In this chapter we will summarize the work presented in this thesis and discuss possible directions for the future work.

7.1 Summary of main results

Cognitive radio and cognitive networks are exciting communication paradigms, which have a potential to enable easier management, self-organization and better performance of the future radio networks. Although these paradigms have been actively researched in the last ten years, a large number of challenges have remained unanswered due to the wide scope and the high-complexity of the problems encountered in building a cognitive radio system. The work in this thesis contributes towards practical realization of cognitive radios. We have studied and proposed algorithmic solutions for facilitating adaptation and optimization in key resource sharing problems such as channel allocation, load balancing and channel access control. Additionally, we have introduced a cognitive resource management framework (CRM), which encompasses tools and mechanisms for cross-layer optimization, cognitive operations and advanced management of future networks.

At the beginning of the thesis, as a background, we have introduced the basic concepts of software defined radio and cognitive radio paradigms. We have reviewed Mitola's idea on ultimate cognitive radio and discussed two views on cognitive radio in the community, namely CR-I, which is dynamic spectrum access capable radio and CR-II, which is fully configurable radio, can learn from the radio environment and optimize itself. Additionally, we have briefly discussed on definitions for cognitive networks. The rest of the background analysis in this thesis is spread out through the chapters and is closely related to the particular problems addressed therein.

Following Mitola's idea on fully (re-)configurable radio that can learn, analyze, decide and act upon the environmental stimuli we have proposed a new architectural framework, called cognitive resource management that could enable easier realization of the cognitive cycle. Our CRM framework was one of the first cognitive resource management architectures on a system level that suggested all the necessary components to implement the cognitive cycle. The design of CRM has been performed in a way that several key design principles such as extensibility, flexibility, portability and reasonable complexity are guaranteed. The framework is component based and in comparison to other approaches in the literature is highly modular and easily extensible to include new functionalities. Furthermore, CRM has been designed to facilitate easier interaction and exchange of information between the layers of the protocol stack by means of generic APIs. A central component of the framework is a so called CRM core, which is responsible for coordination of actions, resource scheduling and decision making. The CRM core reaches optimal decisions exploiting information from different sources, namely the protocol stack, the radio environment, regulatory policies and historical data that have been generated from off-line learning. One of the novelties in the framework is the toolbox of optimization, learning and modelling techniques, which is tightly coupled to the core and can be used to perform local and global optimization based on the rich information provided through the APIs. The toolbox has dualistic existence and meaning. In a run-time environment it is a collection of processes and algorithms that are executed, or can be called, to optimize the cognitive radio stack. From the system designer and programmer point of view the existing set of implemented core algorithms can be seen as a programming library that can be used efficiently to implement cognitive radios. A detailed description on the design of three different APIs, ULLA, GENI and CAPRI, has been presented in the thesis. As extensively discussed in Chapter 3, all three interfaces are intended for interaction with the protocol stack and can successfully abstract different technologies and protocols, making technology-independent cross-layer optimization possible. The generic APIs have been also prototyped and experimentally tested as part of a GOLLUM and ARAGORN collaborative projects and are currently in a standardization procedure.

Further in the thesis, we have studied and proposed several algorithmic solutions for optimization of cognitive radio and general wireless networks, which can be included in the toolbox of the CRM framework. In particular, we have addressed the problem of multi-objective optimization and introduced a genetic algorithm based approach for finding the Pareto-optimal settings of the radio and MAC parameters that optimize the cognitive radio in respect to throughput, reliability and power consumption. After formalizing the multi-objective optimization problem mathematically we have discussed pros and cons of the classical and evolutionary multi-optimization techniques. We have argued that evolutionary algorithms are particularly attractive for solving MOO problems because they deal simultaneously with a set of possible optimal solutions and can find Pareto-optimal set in a single run of the algorithm instead of running series of in the case of the conventional approaches. Thus, we have adopted GA to solve MOO problems for several cognitive radio scenarios.

We have shown how to use GAs for optimization of the radio parameters of an OFDM system. Later we extended the analysis towards cross-layer optimization including both PHY and MAC parameters into the optimization problem. We have, furthermore, tested the performance of GAs under different values for the genetic operators and have determined the sensitivity of GA to those operators. Several conclusions were made based on these studies. First, GA successfully compromises between different objectives and finds the optimal solution based on the user preferences. Second, GA was able to very quickly, in few iterations, converge to acceptably good solutions. The analysis showed that only several evolutionary steps are needed to come up with a good solution, and only if the SNR levels are low it pays off to let the GA run for a longer time period. Third, larger population sizes are not always necessary and careful trade-off have to be made in order to justify the marginal increase of the performance against the big delay for transmitting the additional chromosomes through the channel.

In addition to the PHY layer optimization, we have also introduced a cross-layer multi-objective optimization based on GA with minimal feedback. We have proposed an ARQ-based protocol for cognitive radio system that controls the transmission QoS in terms of delay, throughput, packet loss rate and transmission power consumption. Particularly, we have shown that all optimal transmission parameters can be determined through our GA implementation with the acknowledgement signalling (ACK or NAK) of the prior transmitted packets as the only external input. No additional network state information such as the propagation channel transfer function or the number of active users in the network, is needed at the transmitter. We have carried out comparative performance studies of our algorithm in three different test scenarios. In the first scenario, discrete water-filling, we have determined that GA performs very close to the optimal bit-loading algorithm as long as the number of iterations is large enough. The simulation results have also clearly showed that our solution outperforms the conventional rate adaptation in standard IEEE 802.11 and the algorithm presented by Newman et al. in [13]. In the second scenario, ARQ-based cross-layer optimization with adaptive contention window size, we have also included the contention window size in the parameter set. In attempt to optimize the throughput at given transmission power and target PER, we have concluded the following: GA with optimal CWmin and CWmax outperforms the conventional scheme by approximatively 20 dB. Also, GA performs better compared to "PHY then MAC" approach at

SNR larger than 15 dB. In the third scenario, cross-layer optimization with QoS requirements we have studied two cases: First, how well our algorithm can converge to a low power consumption mode of operation as the requirements change online and second, if the GA solution can satisfy the delay bounds in case real audio/video streaming is carried out. The results have shown that our GA implementation satisfies the power requirements within few time slot periods and also satisfies the delay bound in real time applications. This indicates that GA is also suitable as a real-time optimization and adaptation tool.

Apart of multi-objective and cross-layer optimization for cognitive radios, a large portion of the work in this thesis has focused on dynamic and load balancing channel allocation problem. Our goal has been to provide more flexible and adaptive solutions for interference mitigation in cognitive radio systems, but also consider solutions that can be used in the present wireless networks. Both theoretical and experimental studies have been carried out and two concrete solutions have been proposed, namely a dynamic graph coloring channel allocation and distributed load balancing channel allocation. Based on exhaustive analysis we have shown that the graph coloring approach based on deterministic greedy heuristic algorithm, called DSATUR, greatly outperforms any random or fixed channel allocation scheme, which are unfortunately still practice in many present-day Wi-Fi deployments. Our dynamic graph coloring channel allocation was one of the first automatic channel allocation schemes proposed and has also been tested both on commodity Wi-Fi and SDR platforms. From the comparative analysis we have seen that the aggregate throughput in the network increases almost linearly with the number of access points in the network and that the number of collisions in the network is drastically degreased. The algorithm has been also successfully used by commercial manufacturer in their products.

The work dedicated to the load balancing problem has resulted in two novel distributed channel allocation algorithms based on balls and bins congestion game, namely Compare_And_Balance and Avoid_Contention. We have shown, through analysis in the fluid limit, that the algorithms converge very fast for different types of cost functions and that the algorithms are independent of the class of cost functions. Encouraged by the optimistic analytical analysis, as a next step we studied the performance of the load balancing algorithm through network simulations and experiments using SDR platforms. These analysis have validated the theoretical studies and have confirmed the quick convergence properties also over real channels. This is of great importance in real-time operation of the networks. The algorithm has been shown to increase the fairness in the networks and assign the available channels based on their quality and the load in the network. In this context, the protocol guarantees a certain minimum performance to all the links in the network. The algorithm essentially ensures that any user experiencing very bad payback from some channel will migrate to a new one. Moreover, the distributed nature of the algorithm removes some of the anomalies the centralized protocols generate. Following our approach for minimal feedback, one should note that load balancing algorithms can operate with very little or no information in the feedback channel.

Finally in the last chapter of the thesis we summarize out work on achieving self-organization in allocation of the resources in the network among the cognitive radios without causing overhead to the system. In that regard, we have proposed a novel self-organizing resource allocation scheme with minimal feedback information based on minority game (MG). In particular, we apply MG to schedule user transmissions over a wireless channel through enhancing CSMA. In order to minimize the number of collisions in dense networks. The simulation results have demonstrated that MG inspired version of the CSMA protocol achieves higher channel utilization compared to a standard slotted CSMA without causing serious additional overhead in terms of delay. Moreover the results have indicated that minority game based approaches have a great potential in solving distributed resource allocation problems in ad hoc, mesh and cognitive radio networks.

7.2 Future work

The presented work can be pursued and extended in multiple directions. In the following, we focus shortly only on some tangible tasks for short- to mid-term research.

The basic architectural work for Cognitive Resource Manager is quite mature, but there is a lot of implementation specific research that is still required. The interdisciplinary research with experts of operating system design would be highly desirable in order to enhance the performance of CRM-core and to provide better automated tools to support multi-core processors. In our work we did not consider some specific parts of decomposable and reconfigurable protocol stack. One of the key areas for innovation would be to study Medium Access Control (MAC) layer and make it reconfigurable, yet highly integrated with PHY-layer capabilities. This research work should pay high dividends and some early work has been already started towards realizing this research vision. On higher level of abstraction, revisiting part of CRM architecture from the perspective of cognitive architectures such as ICARUS is still warranted in order to understand how to facilitate automatic learning of behavior as efficiently as possible.

On the experimental side the implementation of key algorithms such as channel allocation methods could be tested also in the context of dynamic spectrum access. The recent availability of measured spectrum occupancy traces and synthetic models is opening a possibility to test new scenarios with unprecedented accuracy outside ISM-bands. Some of these spectrum measurements are done thanks to UMIC research cluster in RWTH Aachen University, and the access to the raw data makes this research avenue particularly attractive.

The minimal feedback protocols and the reasoning under uncertain or incomplete information is something that requires more attention. This is a research direction we believe has a long time perspectives. Specifying quantitative measures and algorithms to handle such situations has utmost importance, and this aspect has been unfortunately often ignored in earlier works. One possibility would be to look more deeply into self-organizing and machine learning based algorithms to explore both theoretical and practical topics in this domain.

There are also some specific areas in this thesis that are open for direct continuation research, and some of it has been already started. Load balancing for both secondary and primary users will be increasingly important issue as the wireless networks are becoming overcrowded. The evolutionary game theory based methods are one avenue to continue. One of the short-term goals is to consider combining the load balancing approaches with power control and scheduling problems. As discussed also in the thesis, minority game is quite general evolutionary concept for both reinforcement learning and game theory. Many current problems with cognitive radio and cognitive network could be considered as such problems, or at least they could be (partially) studied approximately using known dynamical solutions of different minority games.

Generally the main themes of the thesis, context sensitivity and self-organization, remain to be key research problem for foreseeable future. The work done for this thesis and discussions with collaborators both from academia and industry clearly show that self-organization is not only a possibility- it will be required for the future networks. However, theoretical foundations and especially practical approaches that are available for communications engineering community are still rather rudimentary and a lot remains to be researched.

A

Simulation parameters for testing GA convergence properties

A.1 Simulation Cycle

The simulation cycle includes the mandatory setup of all parameters (OFDM, GA, Simulink model), creation of the initial population, as well as various loop to iterate over a set of different settings (seeds, Rice factors, SNR values, bandwidths, path lengths, GA iterations). All the summation steps are shown in Table A.1.

Table A.1: Simulation steps.

Step	Action
1	Define OFDM and GA parameters in the header of the simulation's M-file
2	Define the Simulink model to be used
3	Set current seed
4	Set current Rice factor (only Rice channel)
5	Set current bandwidth
6	Create initial population
7	Set current SNR value
8	Set current path length (only Rayleigh and Rice channels)
9	Set current GA iteration
10	Set model parameters
11	Start simulation
12	Store results
13	Start GA

A.2 Simulation Parameters

A.2.1 Seeds

We perform each simulation with three different seeds and average the results. The seed we use to generate different initial conditions are:

- AWGN channel: [67 8432 475]
- Rayleigh/Rice channel: [7321 85691 326744]

A.2.2 SNR Values

In order to get a picture of the GA performance at different SNR values, we run each simulation at eleven distinct SNRs. The values spread over 35 dB, but are not at equidistant positions. We draw a focus between 5 dB and 15 dB. In this range, the step size is 2 dB, otherwise 5 dB. The concrete SNR values are [1 5 7 9 11 13 15 20 25 30 35] dB.

A.2.3 GA Parameters

We try to find out how the different GA parameters influence the optimization of the OFDM system. As a single chromosome for all subcarriers would consist of 600 bits, we decided to run the GA per subcarrier and thus optimize the subcarriers one by one. In a real-world scenario, this is only possible for chromosomes with the same center frequency. In such a scenario one would better use the average score of all chromosomes in the selection process of the GA. The largest impact on the results was naturally expected from the fitness weights. The chosen fitness weights are listed in Table A.2. Besides fitness weights and GA method, we investigate the influence of mutation rate, score threshold, population size, and number of GA iterations.

A.2.4 Default Parameters

As a combination of each and every parameter would take a long time to simulate and does not provide much insight, we defined some default parameters that are fixed in every simulation. Only the parameter, currently under investigation is changed. All simulation parameters are listed in Table A.3. The default values are in bold face. The default fitness weights are highlighted in Table A.2.

TABLE A.2: Fitness weights.

w_{SNR}	w_{tp}	w_{power}
90 %	5 %	5 %
80 %	**15 %**	**5 %**
80 %	5 %	15 %
60 %	20 %	20 %
50 %	30 %	20 %

TABLE A.3: Simulation parameters. Default values in bold face.

Parameter	Value
SNR	[1 5 7 9 11 13 15 20 25 30 35] dB
Bandwidths	[5 20] MHz
Population size N_{PS}	[5 **10** 15 20]
Maximum Iterations	[3 **5** 10 15]
Mutation rate μ	[1 2 **5** 10 15] %
Elitism rate r_{elite}	20 %
Score threshold s_{th}	[0 0.25 **0.5** 0.6]
Target error rate r_{TER}	[10^{-5} $\mathbf{10^{-6}}$]

B

Abbreviations

AI	Artificial Intelligence
AM	Amplitude Modulation
AP	Access Point
API	Application Programmable Interface
ARP	Address Resolution Protocol
ARQ	Automatic Repeat reQuest
ASIC	Application Specific Integrated Circuit
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CAPRI	Common Application Interface
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CSMA	Collision Avoidance Multiple Access
CORBA	Common Object Request Broker Architecture
CR	Cognitive Radio
CRM	Cognitive Resource Manager
CWN	Cognitive Wireless Network
DCA	Dynamic Channel Allocation
DSA	Dynamic Spectrum Access
DSP	Digital Signal Processor
EA	Evolutionary Algorithm

ETSI	European Telecommunication Standards Institute
FCA	Fixed Channel Allocation
FCC	Federal Communication Commission
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FFT	Fast Fourier Transformation
FPGA	Field Programmable Gate Array
FTP	File Transfer Protocol
GA	Genetic Algorithm
GENI	Generic Network Interface
GHz	Gigahertz
GPP	General purpose Processor
GSM	Global System for Mobile Communication (formerly Groupe Spécial Mobile)
HCA	Hybrid Channel Allocation
HDTV	High Speed TV
HSUPA	High Speed Uplink Packet Access
IF	Intermediate Frequency
IEEE	Institute of Electrical and Electronics Engineers
ISM	Industrial, Scientific, and Medical
ISO	International Standardization Organization
JRRM	Joint Radio Resource Management
JTRS	Joint Tactical Radio System
LLC	Logical Link Control Layer
LTE	Long Term Evolution
LTM	Long Term Memory
MAC	Media Access Control Layer

MG Minority Game

MIMO Multiple Input Multiple Output

MOOP Multi-Objective Optimization Problem

MRRM Multiradio Radio Resource Management

OFDM Orthogonal Frequency Division Multiplexing

OSI Open Systems Interconnection

PHY Physical Layer

QAM quadrature amplitude modulation

QoS Quality of Service

QPSK Quaternary Phase Shift Keying

REM Radio Environment Map

RF Radio Frequency

RKRL Radio Knowledge Representation Language

RRM Radio Resource Manager

RTT Round Trip Time

SCC 41 IEEE Standards Coordination Committee 41

SDR Software Defined Radio

SLP Service Location Protocol

SNIR Signal-to-Noise and Interference Ratio

SNMP Simple Network Management Protocol

SNR Signal-to-Noise Ratio

TCP Transmission Control Protocol

TDMA Time Division Multiple Access

TER Target Error Rate

UDP User Datagram Protocol

ULLA Unified Link-Layer API

UMTS Universal Mobile Telecommunications System

USRP Universal Software Radio Peripheral

UWB Ultra Wide Band

W-CDMA Wideband Code Division Multiple Access

WARP Wireless Open-Access Research Platform

WIMAX Worldwide Interoperability for Microwave Access

WLAN Wireless Local Area Network

WM Working Memory

WRAN Wireless Regional Area Network

Bibliography

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Elsevier Computer Networks Journal*, vol. 50, no. 13, pp. 2127–2159, September 2006.

[2] Q. Zhao and B. M. Sadler, "A Survey of Dynamic Spectrum Access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[3] *FCC Adopts Rules For Unlicensed Use of Television White Spaces*, http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-260A1.pdf.

[4] *Unlicensed Operation in the TV Broadcast Bands. Additional Spectrum for Unlicensed Devices Below 900 MHz and in the 3 GHz Band*, http://www.fcc.gov/Daily_Releases/Daily_Business/2010/.../FCC-10-174A1.pdf.

[5] J. Mitola III and G. Maguire, Jr., "Cognitive Radio: Making Software Radios More Personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, August 1999.

[6] J. Mitola III, *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio.* Ph.D. Thesis, KTH (Royal Institute of Technology), 2000.

[7] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," vol. 23, no. 2, pp. 201 – 220, February 2005.

[8] D. D. Clark, C. Partridge, J. C. Ramming, and J. T. Wroclawski, "A knowledge plane for the internet," in *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications.* New York, NY, USA: ACM Press, 2003, pp. 3–10.

[9] C. Rieser, T. Rondeau, C. Bostian, and T. Gallagher, "Cognitive Radio Testbed: Further Details and Testing of a Distributed Genetic Algorithm Based Cognitive Engine for Programmable Radios," in *Proc. of Military Communications Conference (MILCOM '04), Monterey, CA, USA*, vol. 3, November 2004, pp. 1437–1443.

[10] M. Petrova, P. Mähönen, and J. Riihijärvi, "Evolution of Radio Resource Management: A Case for Cognitive Resource Manager with VPI," in *Proc. of the IEEE International Conference on Communications (ICC '07)*, June 2007, pp. 6471–6745.

[11] R. W. Thomas, L. DaSilva, M. Marathe, and K. N. Wood, "Critical Design Decisions for Cognitive Networks," in *Proc. of the IEEE International Conference on Communications (ICC '07)*, June 2007, pp. 3993–3998.

[12] A. B. MacKenzie, P. Athanas, R. M. Buehrer, S. W. Ellingson, M. Hsiao, C. Patterson, and C. R. C. M. da Silva, "Cognitive Radio and Networking Research at Virginia Tech," *Proceedings of the IEEE*, vol. 97, no. 4, pp. 660–688, April 2009.

[13] T. Newman, B. Barker, A. Wyglinski, A. Agah, J. Evans, and G. Minden, "Cognitive Engine Implementation for Wireless Multicarrier Transceivers," *Wiley Journal of Wireless Communications and Mobile Computing*, vol. 7, no. 9, pp. 1129–1142, November 2007.

[14] *The GNU Software Radio*, http://www.gnu.org/software/gnuradio/.

[15] *Ettus Research LLC*, http://www.ettus.com.

[16] M. Petrova and P. Mähönen, *Cognitive Resource Manager: A cross-layer architecture for implementing Cognitive Radio Networks.* Cognitive Wireless Networks (eds. Fittzek F. and Katz M.), Springer, 2007.

[17] P. Mähönen, M. Petrova, and J. Riihijärvi, *Cooperation in Ad-Hoc Networks.* Cooperation in Wireless Networks: Principles and Applications (ed. F. Fitzek, and M. Katz), Springer, 2006.

[18] P. Mähönen, M. Petrova, J. Riihijärvi, and M. Wellens, "Cognitive Wireless Networks: Your Network Just Became a Teenager," in *Poster session of the IEEE International Conference on Computer Communications (INFOCOM '06), Barcelona, Spain*, April 2006.

[19] J. Riihijärvi, M. Petrova, and M. Mähönen, "A Common Application Requirement Interface for Cognitive Wireless Networks," in *Proc. of the 4th Annual IEEE Workshop on Network Technologies for Software Defined Radio (SDR) Networks, in conjunction with SECON 2009*, Rome, Italy, June 2009.

[20] J. Riihijärvi, M. Petrova, V. Atanasovski, and P. Mähönen, "Extending Policy Languages with Utility and Prioritization Knowledge: The CAPRI Approach," in *Proc. of the 5th IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '10), Singapore*, April 2010.

[21] A. de Baynast, P. Mähönen, and M. Petrova, "ARQ-based Cross-layer Optimization for Wireless Multicarrier Transmission on Cognitive Radio Networks," *Computer Networks*, vol. 52, pp. 778–794, March 2008.

[22] M. Petrova, L. Wu, P. Mähönen, and J. Riihijärvi, "Interference Measurements on Performance Degradation between Colocated IEEE 802.11g/n and IEEE 802.15.4 Networks," in *Proc. of the Sixth International Conference on Networking (ICN '07), Sainte-Luce, Martinique*, April 2007, pp. 93–99.

[23] P. Mähönen and M. Petrova, "Flash Crowds in Cognitive Radio Environment: Beware an Unpredictable Mob," in *Proc. of 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC 2007), Athens, Greece*, September 2007.

[24] P. Mähönen, J. Riihijärvi, and M. Petrova, "Automatic Channel Allocation for Small Wireless Local Area Networks Using Graph Colouring Algorithm Approach," in *Proc. of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC '04), Barcelona, Spain*, vol. 1, September 2004, pp. 536–539.

[25] J. Riihijärvi, M. Petrova, and P. Mähönen, "Frequency Allocation for WLANs Using Graph Colouring Techniques," *Ad Hoc and Sensor Wireless Networks*, vol. 3, pp. 121–139, 2006.

[26] J. Riihijärvi, M. Petrova, and P. Mähönen, "Performance Evaluation of Automatic Channel Assignment Mechanism for IEEE 802.11 Based on Graph Coloring," in *Proc. of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC '06), Helsinki, Finland*, September 2006.

[27] S. Fischer, P. Mähönen, M. Schöngens, and B. Vöcking, "Load Balancing for Dynamic Spectrum Assignment with Local Information for Secondary Users," in *Proc. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'08), Chicago, Illinois, USA*, October 2008.

[28] N. Olano, M. Petrova, and P. Mähönen, "SDR Implementation and Verification of Distributed Load Balancing Spectrum Allocation Algorithm," in *Proc. of 20th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC'09), Tokyo, Japan*, September 2009.

[29] M. Petrova, N. Olano, and P. Mähönen, "Balls and Bins Distributed Load Balancing Algorithm for Channel Allocation," in *Proc. of 7th International Conference on Wireless On-demand Network Systems and Services, (WONS'10), Kranjska Gora, Slovenia*, February 2010, pp. 25–30.

[30] P. Mähönen and M. Petrova, "Minority Game for Cognitive Radios: Cooperating Without Cooperation," *Physical Communication*, vol. 1, no. 2, pp. 94–102, June 2008.

[31] M. Petrova, M. Michalopoulou, and P. Mähönen, "Self-Organizing Multiple Access with Minimal Information: Networking in El Farol Bar," in *Proc. of Workshop on Mobile Computing and Emerging Communication Networks (MCECN) in conjunction with GLOBECOM '10, Miami, Florida, USA*, December 2010.

[32] J. Mitola III, *Software Radio Architecture: Object Oriented Approaches to Wireless System Engineering.* John Wiley and Sons, 2000.

[33] J. H. Reed, *Software Radio: A Modern Approach to Radio Engineering.* Prentice Hall, 2002.

[34] Object Management Group, *The Common Object Request Broker: Architecture and Specification, v*2.0, July 1995.

[35] D. S. Jacobs, "The Integrated Communication Navigation Identification Avionics (ICNIA) program summary from a "lessons learned" perspective," in *Proc. of the IEEE National Aerospace and Electronics Conference (NAECON'90), Dayton, OH , USA*, vol. 1, 1990, pp. 221–226.

[36] R. Prill and M. Antonesco, "Programmable Channelized Digital Radio/Modem," in *Proc. of 10th IEEE/AIAA Digital Avionics Systems Conference*, October 1991, pp. 248–251.

[37] R. Lackey and D. Upmal, "Speakeasy: The Military Software Radio," *IEEE Communication Magazine*, vol. 33, no. 5, pp. 56–61, May 1995.

[38] P. Cook and W. Bonser, "Architectural Overview of the SPEAKeasy System," *IEEE Journal on Selected Areas in Communication*, vol. 17, no. 4, pp. 650–661, April 1999.

[39] Modular Software-Programmable Radio Consortium, "MSRC-5000SCA V1.1," Tech. Rep., September 2000.

[40] *Spectrum Signal Processing by Vecima*, http://www.spectrumsignal.com.

[41] *Motorola, Inc.*, http://www.motorola.com.

[42] G. Minden, J. Evans, L. Searl, D. DePardo, V. Petty, R. Rajbanshi, T. Newman, Q. Chen, F. Weidling, J. Guffey, D. Datla, B. Barker, M. Peck, B. Cordill, A. Wyglinski, and A. Agah, "KUAR: A Flexible Software-Defined Radio Development Platform," in *Proc. of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, (DySPAN'07)*, April 2007, pp. 428–439.

[43] *WARP- Wireless Open-Access Research Platform*, http://warp.rice.edu/.

[44] C. Chang, J. Wawrzynek, and R. W. Brodersen, “BEE2: A High-End Reconfigurable Computing System,” *IEEE Design & Test of Computers*, vol. 22, pp. 114–125, June 2005.

[45] E. Alpaydin, *Introduction to Machine Learning.* The MIT Press, 2004.

[46] *DySPAN2005.* Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Baltimore, Maryland, USA, November 2005.

[47] *DySPAN2007.* Proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Dublin, Ireland, April 2007.

[48] *DySPAN2008.* Proceedings of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Chicago, Illinois, USA, October 2008.

[49] Federal Communication Commission, “Spectrum Policy Task Force,” Report ET Docket no. 02-135, November 2002.

[50] F. H. Sanders, “Broadband Spectrum Surveys in Denver, CO, San Diego, CA, and Los Angeles, CA: Methodology, Analysis, and Comparative Results,” in *Porc. of IEEE Symposium on Electromagnetic Compatibility*, vol. 2, August 1998, pp. 988–993.

[51] M. A. McHenry, D. McCloskey, and G. Lane-Roberts, “Spectrum Occupancy Measurements Republican National Convention, New York City, NY,” Report, Shared Spectrum Company, August 30, 2004 - September 3, 2004, Revision 2, August 2005.

[52] M. A. McHenry and P. Tenhula, “Multi-Band, Multi-Location Spectrum Occupancy Measurements,” in *Proc. of International Symposium on Advanced Radio Technologies (ISART), Boulder, CO, USA*, March 2006.

[53] M. A. McHenry, P. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, “Chicago Spectrum Occupancy Measurements and Analysis and a Long-Term Studies Proposal,” in *Proc. of Workshop on Technology and Policy for Accessing Spectrum (TAPAS), Boston, MA, USA*, August 2006.

[54] M. McHenry, E. Livsics, T. Nguyen, and N. Majumdar, “XG Dynamic Spectrum Access Field Test Results,” *IEEE Communication Magazine*, vol. 45, no. 6, pp. 51–57, June 2007.

[55] A. Petrin and P. G. Steffes, "Analysis and Comparison of Spectrum Measurements Performed in Urban and Rural Areas to Determine the Total Amount of Spectrum Usage," in *Proc. of International Symposium on Advanced Radio Technologies (ISART), Boulder, CO, USA*, March 2005, pp. 9–12.

[56] S. W. Ellingson, "Spectral Occupancy at VHF: Implications for Frequency-Agile Cognitive Radios," in *Proc. of IEEE Vehicular Technology Conference (VTC), Dallas, USA*, September 2005, pp. 1379–1381.

[57] M. Wellens, J. Wu, and P. Mähönen, "Evaluation of Spectrum Occupancy in Indoor and Outdoor Scenario in the Context of Cognitive Radio," in *Proc. of IEEE Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom'07), Orlando, FL, USA*, August 2007, pp. 420–427.

[58] M. Islam, C. Koh, S. Oh, X. Qing, Y. Lai, C. Wang, Y.-C. Liang, B. Toh, F. Chin, G. Tan, and W. Toh, "Spectrum Survey in Singapore: Occupancy Measurements and Analyses," in *Proc. of IEEE Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom'08), Singapore*, May 2008, pp. 1–7.

[59] R. Chiang, G. Rowe, and K. Sowerby, "A Quantitative Analysis of Spectral Occupancy Measurements for Cognitive Radio," in *Proc. 65th IEEE Vehicular Technology Conference, (VTC'07-Spring)*, April 2007, pp. 3016–3020.

[60] V. Blaschke, H. Jäkel, T. Renk, C. Klöck, and F. K. Jondral, "Occupation Measurements Supporting Dynamic Spectrum Allocation for Cognitive Radio Design," in *Proc. of IEEE Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom'07), Orlando, FL, USA*, August 2007, pp. 50–57.

[61] R. Bacchus, A. Fertner, C. Hood, and D. Roberson, "Long-Term, Wide-Band Spectral Monitoring in Support of Dynamic Spectrum Access Networks at the IIT Spectrum Observatory," in *Proc. of the 3rd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'08), Chicago, Illinois, USA*, October 2008.

[62] M. Wellens, "Empirical Modelling of Spectrum Use and Evaluation of Adaptive Spectrum Sensing in Dynamic Spectrum Access Networks," Ph.D. dissertation, RWTH Aachen University, 2010.

[63] D. Robinson, A. Shukla, J. Burns, and A. Atefi, "Resource Trading for Spectrum Aggregation and Management," in *Proc. of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'05)*, 2005, pp. 666–671.

[64] J. M. Chapin and W. H. Lehr, "The Path to Market Success for Dynamic Spectrum Access Technology," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 96–103, May 2007.

[65] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, and S. Suri, "A General Framework for Wireless Spectrum Auctions," in *Proc. of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks,(DySPAN' 07)*, 2007, pp. 22 – 33.

[66] D. Niyato and E. Hossain, "Spectrum Trading in Cognitive Radio Networks: A Market-Equilibrium-Based Approach," *IEEE Wireless Communications*, vol. 15, no. 6, pp. 71–80, December 2008.

[67] S. Sengupta and M. Chatterjee, "An Economic Framework for Dynamic Spectrum Access and Service Pricing," *IEEE/ACM Transactions on Networking*, vol. 17, no. 4, pp. 1200–1213, 2009.

[68] H. Xu, J. Jin, and B. Li, "A Channel Portfolio Optimization Framework for Trading in a Spectrum Secondary Market," in *Proc. of IEEE International Conference on Communications (ICC '10), Cape Town, South Africa*, 2010, pp. 1 – 5.

[69] C. Caicedo and M. B. H. Weiss, "The Viability of Spectrum Trading Markets," in *Proc. of the 5th IEEE Symposium on New Frontiers in Dynamic Spectrum, (DySPAN'10), Singapore*, 2010, pp. 1 – 10.

[70] *IEEE 802.22 Working Group on Wireless Regional Area Networks*, http://www.ieee802.org/22.

[71] P. Mähönen, "Cognitive Trends in Making: Future of Networks," in *Proc. of IEEE 15th International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC '04), Barcelona, Spain*, September 2004, pp. 1449–1454.

[72] R. Thomas, L. DaSilva, and A. B. MacKenzie, "Cognitive Networks," in *Proc. of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, November 2005.

[73] R. Thomas, D. Friend, L. DaSilva, and A. B. MacKenzie, "Cognitive Networks: Adaptation and Learning to Achieve End-to-End Performance Objectives," *IEEE Communications Magazine*, vol. 44, no. 12, pp. 51 – 57, December 2006.

[74] P. Magnusson, J. Lundsj, J. Sachs, and P. Wallentin, "Radio Resource Management Distribution in a Beyond 3G Multi-Radio Access Architecture," in *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM '04), Dallas, Texas, USA*, 2004, pp. 3472–3477.

[75] *Vanu Inc.*, http://www.vanu.com.

[76] *The GOLLUM-project website*, http://www.ist-gollum.org [Cited on: 15th of June 2010].

[77] *The ARAGORN-project website*, http://www.ist-aragorn.eu [Cited on: 15th of June 2010].

[78] *Open source release of the Unified Link-Layer API*, http://ulla.sourceforge.net/ [Cited On: 15th of June 2010].

[79] M. Bandholz *et al.*, "Unified Link-Layer API Enabling Wireless-Aware Applications," in *Proc. of IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06), Helsinki, Finland*, 2006.

[80] F. Kelly, "Charging and Rate Control for Elastic Traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[81] F. Kelly, A. Maulloo, and D. Tan, "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[82] D. Elenius, G. Denker, M. Stehr, R. Senanayake, C. Talcott, and D. Wilkins, "CoRaL–Policy Language and Reasoning Techniques for Spectrum Policies," in *Proc. of the 8th IEEE International Workshop on Policies for Distributed Systems and Networks.* IEEE Computer Society Washington, DC, USA, 2007, pp. 261–265.

[83] SRI XG Team, "XG architecture. Request for Comments," SRI International, Tech. Rep., 2006.

[84] XG Working Group, "The XG Vision. Request for Comments. Version 2.0," BBN Technologies, Tech. Rep., 2005.

[85] J. Hoydis, M. Petrova, and P. Mähönen, "The Effects of Topology on the Local Throughput of Ad Hoc Networks," *Ad Hoc & Sensor Wireless Networks*, vol. 7, no. 3–4, pp. 337–347, 2009.

[86] J. Riihijärvi, P. Mähönen, M. Petrova, and V. Kolar, "Enchanting Cognitive Radios With Spatial Statistics: From Radio Environment Map to Topology Engine," in *Proc. of CROWNCOM '09, Hannover, Germany*, June 2009.

[87] R. Thomas, L. DaSilva, and A. B. MacKenzie, "Cognitive Networks," in *Proc. of the IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05), Baltimore, USA*, November 2005, pp. 352 – 360.

[88] C. J. Rieser, "Biologically inspired cognitive radio engine model utilizing distributed genetic algorithms for secure and robust wireless communications and networking," Ph.D. dissertation, Virginia Tech, 2004.

[89] P. Sutton, L. Doyle, and K. Nolan, "A Reconfigurable Platform for Cognitive Networks," in *Proc. of the 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM '06), Mykonos, Greece*, June 2006.

[90] K. Nolan, P. Sutton, and L. Doyle, "An Encapsulation for Reasoning, Learning, Knowledge Representation, and Reconfiguration Cognitive Radio Elements," in *Proc. of the 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM '06), Mykonos, Greece*, June 2006.

[91] D. Raychaudhuri, N. B. Mandayam, J. B. Evans, B. J. Ewy, S. Seshan, and P. Steenkiste, "CogNet- An Architectural Foundation for Experimental Cognitive Radio Networks Within the Future Internet," in *Proc. of the ACM/IEEE MobiArch '06, San Francisco, CA, USA*, December 2006, pp. 11–16.

[92] B. S. Manoj, R. R. Rao, and M. Zorzi, "CogNet: A Cognitive Complete Knowledge Network System," *IEEE Wireless Communications*, pp. 81–88, December 2008.

[93] D. Bourse *et al.*, "End-to-End Reconfigurability- Management and Control of Adaptive Communication Systems," in *Proc. of the SDR'06 Thechnical Conference and Product Exposition, Orlando FL, USA*, 2006.

[94] *End-to-End Efficiency (E^3)-project website*, http://www.ist-e3.eu [Cited on: 15th of June 2010].

[95] Y. Zhao *et al.*, "Applying Radio Environment Maps to Cognitive Wireless Regional Area Networks," in *Proc. of DySPAN '07, Dublin, Ireland*, April 2007.

[96] ——, "Development of Radio Environment map Enabled Case- and Knowledge-based Learning Algorithms for IEEE 802.22 WRAN Cognitive Engines," in *Proc. of CROWNCOM '07, Orlando, FL, USA*, August 2007.

[97] F. Perich and M. McHenry, "Policy-based Spectrum Access Control for Dynamic Spectrum Access Network Radios," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, no. 1, pp. 21–27, 2009.

[98] D. Vernon, G. Metta, and G. Sandini, "A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 151–180, 2007.

[99] A. Newell, "The Knowledge Level," *Artificial Intelligence*, vol. 18, no. 1, pp. 82–127, March 1982.

[100] ——, *Unified Theories of Cognition.* Harvard University Press, Cambridge, Massachusetts, 1990.

[101] J. E. Laird, A. Newell, and P. Rosenbloom, "Soar: An Architecture for General Intelligence," *Artificial Intelligence*, vol. 33, no. 1, pp. 1–64, 1987.

[102] *A Cognitive Architecture for Physical Agents*, http://cll.stanford.edu/research/ongoing/icarus/ [Cited on: 15th of June 2010].

[103] Langley, P. and Choi, D. and Shapiro, D., "A Cognitive Architecture for Physical Agents," Computational Learning Laboratory, CSLI, Stanford University, CA, Tech. Rep., 2004.

[104] R. Brooks, *Cambrian Intelligence: The Early History of the New AI.* MIT Press, Boston, MA, 1999.

[105] J. F. Lehman, J. Laird, and P. Rosenbloom, "A Gentle Introduction to SOAR, an Architecture for Human Cognition: 2006 Update," http://ai.eecs.umich.edu/soar/sitemaker/docs/misc/GentleIntroduction-2006.pdf.

[106] H. Q. Chong, A. Tan, and G. W. Ng, "Integrated Cognitive Architectures: A Survey," *Artificial Intelligence Review*, vol. 28, no. 2, pp. 103–130, 2007.

[107] K. Deb, *Multi-objective Optimization using Evolutionary Algorithms.* John Wiley & Sons, LTD, 2004.

[108] M. Farina, K. Deb, and P. Amato, "Dynamic Multiobjective Optimization Problems: Test Cases, Approximations, and Applications," *IEEE Transactions on Evolutionary Computation*, vol. 8, pp. 425–442, Oct. 2004.

[109] E. Zitzler and L. Thiele, "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, Nov. 1999.

[110] C. A. Coello Coello, *20 Years of Evolutionary Multi-Objective Optimization: What Has Been Done and What Remains to be Done.* Computational Intelligence: Principles and Practice, Chapter 4 (eds. G. Y. Yen and D. B. Fogel), IEEE Computational Intelligence Society, 2006, pp. 73–88.

[111] T. Mitchell, *Machine Learning.* McGraw Hill, 1997, ch. 9, pp. 249–273.

[112] K. F. Man, K. S. Tang, and S. Kwong, "Genetic Algorithms: Concepts and Applivations," *IEEE Transactions on Industrial Electronics*, vol. 43, no. 5, pp. 519–534, October 1996.

[113] D. E. Goldberg and K. Deb, "A Comparison of Selection Schemes Used in Genetic Algorithms," in *Foundations of Genetic Algorithms 1 (FOGA-1)*, pp. 69–93.

[114] R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.

[115] M. N. H. Siddique and M. O. Tokhi, "Training neural networks: Backpropogation vs. genetic algorithms," in *Proceedings of IEEE International Joint Conference on Neural Networks, Washington DC*, 2001, pp. 2673–2678.

[116] C. Fonseca and P. Fleming, "Multi-objective Optimization and Multiple Constraint Handling with Evolutionary Algorithms," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 28, no. 1, pp. 26–37, January 1998.

[117] T. W. Rondeau, C. J. Rieser, B. Le, and C. W. Bostian, "Cognitive Radios with Genetic Algorithms: Intelligent Control of Software Defined Radios," *SDR Forum Technical Conference*, vol. C, pp. 3–8, 2004.

[118] IEEE 802.11, "IEEE 802.11-1999," Institute of Electrical and Electronics Engineers, Tech. Rep., 1999, Available: http://standards.ieee.org/getieee802/802.11.html.

[119] Ying Jun Zhang and K. Letaief, "Cross-Layer Adaptive Resource Management for Wireless Packet Networks With OFDM Signaling," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3244–3254, Nov. 2006.

[120] R. Sonalkar and R. Shively, "An Efficient Bit-loading Algorithm for DMT Applications," *IEEE Communications Letters*, vol. 4, pp. 80–82, March 2000.

[121] G. Bianchi, "IEEE 802.11-saturation Throughput Analysis," *IEEE Communications Letters*, vol. 2, pp. 318–320, Dec. 1998.

[122] D. Whitley, "A Genetic Algorithm Tutorial," *Statistics and Computing*, vol. 4, pp. 65–85, 1994.

[123] C. Fonseca and P. Fleming, "Multiobjective Optimization and Multiple Constraint Handling with Evolutionary Algorithms-Part II: Application Example," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 28, no. 1, pp. 38–47, January 1998.

[124] R. F. H. Fischer and J. B. Huber, "A New Loading Algorithm for Discrete Multitone Transmission," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 1, 1996, pp. 724–728.

[125] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2002.

[126] V. H. MacDonald, "Advanced Mobile Phone Service: The Cellular Concept," *Bell System Technical Journal*, vol. 58, pp. 968–980, 1979.

[127] I. Katzela and M. Naghshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunications Systems: A Comprehensive Survey," *IEEE Personal Communications*, vol. 3, pp. 10–31, 1996.

[128] A. Gamst, "Homogeneous Distribution of Frequencies in a Regular Hexagonal Cell System," *IEEE Transactions on Vehicular Technology*, vol. VT-31, no. 3, pp. 132–144, 1982.

[129] M. Zhang and T. S. Yum, "The Non-uniform Compact Pattern Allocation Algorithm for Cellular Mobile Systems," *IEEE Transactions on Vehicular Technology*, vol. VT-40, pp. 387–391, 1991.

[130] T. S. Yum and W. S. Wong, "Hot Spot Traffic Relief in Cellular Systems," *IEEE JSAC*, vol. 11, pp. 934–940, 1993.

[131] H. Jiang and S. Rappaport, "Prioritized Channel Borrowing Without Locking: A Channel Sharing Strategy for Cellular Communications," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 163–172, 1996.

[132] D. Cox and D. Reudnik, "Dynamic Channel Assignment in High-capacity Mobile Communication Systems," *Bell System Technical Journal*, vol. 50, pp. 1833–1857, 1971.

[133] K. Okada and F. Kubota, "On Dynamic Cchannel Assignment in Cellular Mobile Radio Systems," in *Proc. of IEEE International Symposium on Circuits and Systems*, vol. 2, 1991, pp. 938–941.

[134] D. Cox and D. Reudnik, "Dynamic Channel Assignment in Two Dimension Large-scale Mobile Radio Systems," *Bell System Technical Journal*, vol. 51, pp. 1611–1628, 1972.

[135] Y. Furuya and Y. Akaiwa, "A Distributed Channel Allocation Scheme for Mobile Communication Systems," *IEICE Transactions*, vol. E74, no. 6, pp. 1531–1537, 1991.

[136] C.-L. I and P. H. Chao, "Distributed Dynamic Channel Allocation Algorithms with Adjacent Channel Constraints," in *Proc. of PIMRC'04, Den Haag, The Netherlands*, vol. B2.3, September 1994, pp. 169–175.

[137] G. Cao and M. Singhal, "Distributed Fault-tolerant Channel Allocation for Cellular Networks," *IEEE JSAC*, vol. 18, no. 7, pp. 1326–1337, 2000.

[138] J. Jiang, T. H. Lai, and N. Soundarajan, "On Distributed Dynamic Channel Allocation in Mobile Cellular Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 10, pp. 1024–1037, 2002.

[139] D. Cox and D. Reudnik, "Increasing Channel Occupancy in Large Scale Mobile Radio Systems: Dynamic Channel Reassignment," *IEEE Transactions on Vehicular Technology*, vol. VT-22, pp. 218–222, 1973.

[140] T. J. Kahwa and N. Georganas, "A Hybrid Channel Assignment Scheme in Large Scale Cellular-structured Mobile Communication Systems," *IEEE Transactions on Communications*, vol. COM-26, pp. 432–438, 1978.

[141] J. Sin and N. Georganas, "A Simulation Study of a Hybrid Channel Assignment Scheme for Cellular Land-mobile Radio Systems with Erlang-C Service," *IEEE Transactions on Communications*, vol. COM-9, pp. 143–147, 1981.

[142] W.Yue, "Analytical Methods to Calculate the Performance of a Cellular Mobile Radio Communication System with Hybrid Channel Assignment," *IEEE Transactions on Vehicular Technology*, vol. VT-40, pp. 453–459, 1991.

[143] S. Chieochan, E. Hossain, and J. Diamond, "Channel Assignment Schemes for Infrastructure-based 802.11 WLANS: A Survey," *IEEE Communications & Tutorials*, vol. 12, no. 1, pp. 124–136, 2010.

[144] "IEEE Standard for Information Technology-Telecommunications and Information Exchange between Systems-Local and Metropolitan Area Networks-specific Requirenments - Part 11: Wireless LAN Medium Access Control (MAC) and Pysical Layer (PHY) Specifications," *IEEE STD 802.11-2007 (Revision of IEEE STD 802.11-1999)*, pp. C1–1184, June 2007.

[145] P. Fuxjäger, D. Valerio, and F. Ricciato, "The Myth of Non-Overlapping Channels: Interference Measurements in IEEE 802.11," in *Proc. of 4th Annual Conference on Wireless On demand Network Systems and Services (WONS'04), Madonna di Campiglio, Italy*, January 2004.

[146] S. Liese, D. Wu, and P. Mohapatra, "Experimental Characterization of an 802.11b Wireless Mesh Networks," in *Proc. of the 2nd Annual International Wireless Communications and Mobile Computing Conference (IWCMC'06), Vancouver, British Columbia, Canada*, July 2006.

[147] K. Chebrolu, B. Raman, and S. Sen, "Long-distance 802.11b Links: Performance Measurements and Experience," in *Proc. of the 20th Annual International Conference on Mobile Computing and Networking (MobiCom'06), Los Angeles, CA, USA*, September 2006, pp. 74–85.

[148] I. Howitt and J. A. Gutierrez, "IEEE 802.15.4 Low Rate - Wireless Personal Area Network Coexistence Issues," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'03), New Orleans, LA, USA*, vol. 3, March 2003, pp. 1481–1486.

[149] S. Shin, S. Choi, H. S. Park, and W. H. Kwon, "Packet Error Rate Analysis of IEEE 802.15.4 Under IEEE 802.11b Interference," in *Proc. of the 3rd International Conference on Wired/Wireless Internet Communications, (WWIC'05), Xanthi, Greece*, May 2005, pp. 279–288.

[150] M. Petrova, J. Riihijärvi, P. Mähönen, and S. Labella, "Performance Study of IEEE 802.15.4 Using Measurements and Simulations," in *Proc. of IEEE Wireless Communications & Networking Conference WCNC '06, Las Vegas, USA*, April 2006.

[151] R. Gummadi, D. Wetherall, B. Greenstein, and S. Seshan, "Understanding and Mitigating the Impact of RF Interference on 802.11 Networks," in *Proc. of the annual conference of the Special Interest Group on Data Communication (SIGCOMM '07), Kyoto, Japan*, August 2007, pp. 385–396.

[152] T. Karhima, Silvennoinen, A. Hall, and S.-G. M. Haggman, "IEEE 802.11b/g WLAN Tolerance to Jamming," in *Proc. of Military Communications Conference (MILCOM'04), Monterey, CA, USA*, vol. 3, November 2004, pp. 1364–1370.

[153] L. Sydanheimo, M. Keskilammi, and M. Kivikoski, "Performance Issues on the Wireless 2.4 GHz ISM band in a Multisystem Environment," *IEEE Transactions on Consumer Electronics*, vol. 48, no. 3, pp. 643–658, August 2002.

[154] C. F. Chiasserini and R. R. Rao, "Coexistence Mechanisms for Interference Mitigation Between IEEE 802.11 WLANs and Bluetooth," in *Proc. of IEEE INFOCOM '02, New York, USA*, vol. 2, June 2002, pp. 590–598.

[155] R. C. Rodrigues, G. R. Mateus, and A. A. F. Loureiro, "On the Design and Capacity Planning of a Wireless Local Area Network," in *Proc. of IEEE/IFIP Network Operations and Management Symposium*, April 2000, pp. 335–348.

[156] P. Wertz, M. Sauter, F. M. Landstorfer, G. Wlfle, and R. Hoppe, "Automatic Optimization Algorithms for the Planning of Wireless Local Area Networks," in *Proc. of the IEEE Vehicular Technology Conference (VTC'04)*, September 2004, pp. 3010–3014.

[157] X. Ling and K. L. Yeung, "Joint Access Point Placement and Channel Assignment for 802.11 Wireless LANs," *IEEE Transactions on Wireless Communications*, vol. 5, no. 10, pp. 2705–2711, August 2006.

[158] J. Riihijärvi, M. Petrova, and P. Mähönen, "Frequency Allocation for WLANs Using Graph Colouring Techniques," in *Proc. of the 2nd Wireless On-demand Network Systems and Services Conference WONS'05, St. Moritz, Switzerland*, January 2005.

[159] A. Mishra, V. Brik, , S. Banerjee, A. Srinivasan, and W. Arbaugh, "A Client-driven Approach for Channel Management in Wireless LANs," in *Proc. of the 25th IEEE International Coference on Computer Communications (INFOCOM'06)*, April 2006.

[160] J. K. Chen, G. D. Veciana, and T. S. Rappaport, "Improved Measurement-based Frequency Allocation Algorithms for Wireless Networks," in *Proc. of IEEE GLOBECOM'07, Washington, DC, USA*, November 2007.

[161] A. Akella, G. Judd, S. Seshan, and P. Steenkiste, "Self-management in Chaotic Wireless Deployments," in *Proc. of the 11th annual international conference on Mobile computing and networking (MobiCom '05), Cologne, Germany*, 2005, pp. 185–199.

[162] F. Gamba, J.-F. Wagen, and D. Rossier, "A Simple Agent-based Framework for Adaptive WLAN Frequency Management," in *Proc. of MobiCom'03, San Diego, CA, USA*, September 2003.

[163] M. Achanta, "Method and Apparatus for Least Congested Channel Scan for Wireless Access Points," *US Patent No. 20060072602*, April 2006.

[164] M. Ihmig and P. Steenkiste, "Distributed Dynamic Channel Selection in Chaotic Wireless Networks," in *Proc. of European Wireless, Paris, France*, April 2007.

[165] A. Mishra, V. Shrivastava, D. Agrawal, S. Banerjee, and S. Ganguly, "Distributed Channel Management in Uncoordinated Wireless Environments," in *Proc. of the 20th annual international conference on Mobile computing and networking, (MobiCom'06)*, September 2006, pp. 170–181.

[166] K. Leung and B. Kim, "Frequency Assignment for Multi-Cell IEEE 802.11 Wireless Networks," in *Proc. of VTC'03, Orlando, Florida, USA*, October 2003.

[167] W. K. Hale, "Frequency Assignment: Theory and Applications," *Proceedings of the IEEE*, vol. 68, pp. 1497–1514, 1980.

[168] M. Cozzens and F. S. Roberts, "T-colorings of Graphs and the Channel Assignment Problem," *Congressus Numerantium*, vol. 35, pp. 191–208, 1982.

[169] C. McDiarmid and B. Reed, "Channel Assignment and Weighted Coloring," *Networks*, vol. 36, no. 2, pp. 114–117, 2000.

[170] R. Battiti, A. Bertossi, and D. Cavallaro, "A Randomized Saturation Degree Heuristic for Channel Assignment in Cellular Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 2, pp. 364–374, March 2001.

[171] A. Eisenblätter, M. Grötschel, and A. M. Koster, "Frequency Planning and Ramifications of Coloring," *Discussiones Mathematicae, Graph Theory*, no. 22, pp. 51–88, 2002.

[172] G. Even, Z. Lotker, D. Ron, and S. Smorodinsky, "Conflict-Free Colorings of Simple Geometric Regions with Applications to Frequency Assignment in Cellular Networks," in *FOCS '02: Proceedings of the 43rd Symposium on Foundations of Computer Science*. Washington, DC, USA: IEEE Computer Society, 2002, pp. 691–700.

[173] L. Narayanan, "Channel Assignment and Graph Multicoloring," *Handbook of Wireless Networks and Mobile Computing*, pp. 71–94, 2002.

[174] R. Diestel, *Graph Theory*. Springer-Verlag, 2000.

[175] B. Bollobás, *Modern Graph Theory*. Springer-Verlag, 2000.

[176] *Algorithms Course Materials*, http://compgeom.cs.uiuc.edu/~jeffe/teaching/algorithms.

[177] D. Brélaz, "New Methods to Color the Vertices of a Graph," *Communications of the ACM*, vol. 22, pp. 251–256, 1979.

[178] J. Turner, "Almost All k-colorable Graphs are Easy to Color," *Journal of Algorithms*, vol. 9, pp. 63–82, 1988.

[179] M. M. Halldórsson and M. Szegedy, "Lower Bounds for On-line Graph Coloring," in *Proc. of the 3rd ACM-SIAM Symposium on Discrete Algorithms*, 1992.

[180] Y.-T. Tsai, Y.-L. Lin, and F. R. Hsu, "The On-line First-fit Algorithm for Radio Frequency Assignment Problems," *Information processing letters*, vol. 84, pp. 195–199, 2002.

[181] F. Roberts, "T-colorings of Graphs: Recent Results and Open Problems," *Discrete mathematics*, vol. 93, no. 2–3, pp. 229–245, November 1991.

[182] D. D.-F. Liu, "T- colorings of Graphs," vol. 101, pp. 203–212, 1992.

[183] G. J. Chang, D. D.-F. Liu, and X. Zhu, "Distance Graphs and T- Colorings," vol. Series B, 75, 1999.

[184] *ns-2 simulator*, http://nsnam.isi.edu/nsnam/index.php/Main_Page.

[185] G. Xylomenos, G. P. Polyzos, P. Mähönen, and M. Saaranen, "TCP Performance Issues Over Wireless Links," *IEEE Communications Magazine*, vol. 39, no. 4, pp. 52–58, 2001.

[186] A. Gurtov, "Effect of Delays on TCP Performance," in *Proc. of IFIP Personal Wireless Communications (PWC'01)*, August 2001.

[187] *HostAP Driver*, http://hostap.epitest.fi/.

[188] *Iperf-The TCP/UDP Bandwidth Measurement Tool, National Laboratory for Applied Network Research (NLANR)/Distributed Application Support Team (DAST)*, http://sourceforge.net/projects/iperf/.

[189] Y. Bejerano, S.-J. Han, and L. E. Li, "Fairness and Load Balancing in Wireless LANs Using Association Control," in *Proc. of the 10th annual international conference on Mobile Computing and Networking (MobiCom '04), Philadelphia, PA, USA*, September 2004.

[190] M. Raab and A. Steger, "Balls into Bins– A Simple and Tight Analysis," *Lecture Notes in Computer Science*, vol. 1518, pp. 159–170, 1998.

[191] R. W. Rosenthal, "A Class of Games Possessing Pure-strategy Nash Equilibria," *International Journal of Game Theory*, vol. 2, pp. 65–67, 1973.

[192] D. Fotakis, S. Kontogiannis, E. Koutsoupias, M. Mavronicolas, and P. Spirakis, "The Structure and Complexity of Nash Equilibria for a Selfish Routing Game," in *Proc. 29th Int. EATCS Coll. on Automata, Languages and Programming (ICALP)*, Malaga, Spain, 2002, pp. 123–134.

[193] E. Even-Dar and Y. Mansour, "Fast Convergence of Selfish Rerouting," in *Proc. 16th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA)*, Vancouver, British Columbia, Canada, pp. 772–781.

[194] S. Fischer and B. Vöcking, "On the Evolution of Selfish Routing," in *Proc. 12th Ann. European Symp. on Algorithms (ESA)*, ser. Lecture Notes in Comput. Sci., S. Albers and T. Radzik, Eds., no. 3221. Bergen, Norway: Springer-Verlag, September 2004, pp. 323–334.

[195] S. Fischer, H. Räcke, and B. Vöcking, "Fast Convergence to Wardrop Equilibria by Adaptive Sampling Methods," in *Proc. 38th Ann. ACM. Symp. on Theory of Comput. (STOC).* Seattle, WA, USA: ACM, May 2006, pp. 653–662.

[196] S. Fischer and B. Vöcking, "Adaptive Routing with Stale Information," in *Proc. 24th Ann. ACM SIGACT-SIGOPS Symp. on Principles of Distributed Computing (PODC)*, M. K. Aguilera and J. Aspnes, Eds. Las Vegas, NV, USA: ACM, July 2005, pp. 276–283.

[197] P. Berenbrink, T. Friedetzky, L. A. Goldberg, P. Goldberg, Z. Hu, and R. Martin, "Distributed Selfish Load Balancing," in *Proc. 17th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA)*, 2006.

[198] *Qualnet Network Simulator*, http://www.scalable-networks.com/products/qualnet/.

[199] *Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (WPANs)*, no. 802.15.4. IEEE, 2006.

[200] D. Bertsekas and R. Callager, *Data Networks.* Prentice Hall, 1992.

[201] D. Challet and Y. C. Zhang, "Emergence of Cooperation and Organization in an Evolutionary Game," *Physica A*, no. 246, pp. 407–418, 1997.

[202] W. B. Arthur, "Inductive reasoning and bounded rationality," *American Economic Review*, vol. 84, no. 2, pp. 406–411, 1994.

[203] P. Jefferies, M. L. Hart, and N. F. Johnson, "Deterministic Dynamics in the Minority Game," *Physical Review E*, vol. 65, 016105, pp. 1–8, December 2001.

[204] M. L. Hart, P. Jefferies, and N. F. Johnson, "Dynamics of the Time Horizon Minority Game," *Physica A*, vol. 311, pp. 275–290, 2002.

[205] M. Marsili, D. Challet, and R. Zecchina, "Exact Solution of a Modified El Farol's Bar Problem: Efficiency and the Role of Market Impact," *Physica A*, vol. 280, no. 3-4, pp. 522 – 553, June 2000.

[206] D. Challet, M. Marsili, and R. Zecchina, "Statistical Mechanics of Heterogeneous Agents: Minority Games," *Physical Review Letter*, vol. 84, no. 3-4, pp. 1824 – 1827, 2000.

[207] C. H. Yeung and Y.-C. Zhang, "Minority Games," *arXiv:0811.1479v2 [physics.soc-ph]*, November 2008.

[208] E. Moro, "The Minority Game: An Introductory Guide," in *in Advances in Condensed Matter and Statistical Physics*, E. Korutcheva and R. Cuerno, Eds. Nova Science Publishers, Inc., 2004.

[209] R. Franke, "Reinforcement Learning in the El Farol Model," *Journal of Economic Behavior and Organization*, vol. 51, pp. 367–388, 2003.

[210] R. Savit, R. Manuca, and R. Riolo, "Adaptive Competition, Market Efficiency, and Phase Transitions," *Phys. Rev. Lett.*, vol. 82, no. 10, pp. 2203–2206, 1999.

[211] "Irrelevance of Memory in the Minority Game), author = A. Cavagna, journal = Phys. Rev. E, volume = 59, number = 4, pages = R3783–R3786, year = 1999, month = Apr."

[212] F. K. Chow and H. F. Chau, "Multiple Choice Minority Game," *Physica A*, vol. 319, pp. 601 – 615, 2003.

[213] M. Thomas, "A Generalization of Poisson's Binomial Limit for Use in Ecology," *Biometrika*, vol. 36, no. 1-2, pp. 18–25, 1949.

[214] M. Petrova, P. Mähönen, and J. Riihijärvi, "Connectivity Analysis of Clustered Ad Hoc and Mesh Networks," in *Proc. of IEEE Global Communications Conference, GLOBECOM '07*, November 2007, pp. 1139–1143.

[215] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, and J.-P. Sheu, "The Broadcast Storm Problem in a Mobile Ad Hoc Network," in *Proc. of the 5th annual ACM/IEEE international conference on Mobile computing and networking, MobiCom '99*, August 1999, pp. 151–162.

[216] A. M. Bell, W. A. Sethares, and J. A. Bucklew, "Coordination Failure as a Source of Congestion in Information Networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 875–885, March 2003.

[217] H. Kutsuna, S. Tagashira, and S. Fujita, "A Fair and Efficient Congestion Control Scheme Based on Minority Game," in *Proc. of the 14th IEEE International Conference on Networks, ICON '06*, vol. 1, September 2006, pp. 1–6.

[218] P. Mertikopoulos and A. Moustakas, "Correlated Anarchy in Overlapping Wireless Networks," *IEEE Journal on Selected Areas in Communications,*, vol. 26, no. 7, pp. 1160–1169, September 2008.

[219] I. Erev and A. E. Roth, "Reinforcement Learning in Experimental Games with unique Mixed Strategy Equilibria," *American Economic Review*, no. 4, pp. 848–881, 1998.

Curriculum Vitae

Personal Information

Name	Marina Petrova
Date of birth	11th June 1978
Place of birth	Skopje, Macedonia
Nationality	Macedonian

Education

April 2003 – November 2010	RWTH Aachen University Institute for Networked Systems Doctoral studies	Aachen, Germany
October 1997 – June 2002	Ss. Cyril and Methodius University Diploma (Dipl.-Ing.) Electrical Engineering and Information Technology	Skopje, Macedonia
June 1997	Rade Jovcevski-Korcagin Natural Sciences and Mathematics High School (Abitur)	Skopje, Macedonia

Employment

October 2006 – November 2010	RWTH Aachen University Institute for Networked Systems Chief Research Scientist	Aachen, Germany
April 2003 – September 2006	RWTH Aachen University Department of Wireless Networks Research Assistant	Aachen, Germany
October 2002 – March 2003	RWTH Aachen University Department of Communication Networks Research Trainee	Aachen, Germany
July 2001 – October 2001	Elmos Semiconductor AG Internship	Dortmund, Germany

July 2000 – September 2000	Menta Cable y Televisio de Catalunya Internship	Barcelona, Spain

List of Publications

Book chapters relevant to the thesis

M. Petrova and P. Mähönen, "Cognitive Resource Manager: A Cross-layer Architecture for Implementing Cognitive Radio Networks" in *Cognitive Wireless Networks (eds: Fitzek F. and Katz M.)*, Springer, 2007.

P. Mähönen, **M. Petrova** and J. Riihijärvi, "Cooperation in Ad Hoc Networks," in *Cooperation in Wireless Networks: Principles and Applications (eds: Fitzek F. and Katz M.)*, Springer, 2006.

Journal papers relevant to the thesis

J. Hoydis, **M. Petrova** and P. Mähönen, 'The Effects of Topology on the Local Throughput of Ad Hoc Networks," *Ad Hoc & Sensor Wireless Networks*, vol. 7, no. 3-4, pp. 337-347, 2009.

P. Mähönen and **M. Petrova**, "Minority Game for Cognitive Radios: Cooperating without Cooperation," *Elsevier Physical Communications*, vol. 1, no. 2, pp. 94-102, June 2008.

A. de Baynast, P. Mähönen and **M. Petrova**, "ARQ-based Cross-layer Optimization for Wireless Multicarrier Transmission on Cognitive Radio Networks," *Elsevier Computer Networks*, vol. 52, no. 4, pp. 778-794, March 2008.

J. Riihijärvi, **M. Petrova** and P. Mähönen, "Frequency Allocation for WLANs Using Graph Colouring Techniques," *Ad Hoc & Sensor Wireless Networks*, vol. 3, no. 2, pp. 121-139, 2007.

Journal papers in addition to the work in this thesis

E. Meshkova, J. Riihijärvi, **M. Petrova** and P. Mähönen, "A Survey on Resource Discovery Mechanisms, Peer-to-Peer and Service Discovery Frameworks," *Elsevier Computer Networks*, vol. 52, no.11, pp. 2097-2128, August 2008.

P. Mähönen, J. Riihijärvi, **M. Petrova**, and Z. Shelby, "Hop-by-Hop towards Future Mobile Broadband IP," *IEEE Communication Magazine*, vol. 42, no. 3, pp. 138-146, March 2004.

L. Muñoz, R. Agüero, J. Choque, J. Irastorza, L. Sánchez, **M. Petrova**, and P. Mähönen, "Empowering Next-Generation Wireless Personal Communication Networks," *IEEE Communication Magazine*, vol. 42, no. 5, pp. 64-70, May 2004.

Conference papers relevant to the thesis

M. Petrova, M. Michalopoulou and P. Mähönen, "Self-Organizing Multiple Access with Minimal Information: Networking in El Farol Bar," in *Proc. of Workshop on Mobile Computing and Emerging Communication Networks (MCECN) in conjunction with GLOBECOM '10, Miami, Florida, USA (accepted).*

J. Riihijärvi, **M. Petrova**, V. Atanasovski and L. Gavrilovska, "Extending Policy Languages with Utility and Prioritization Knowledge: The CAPRI Approach," in *Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'10)*, Singapore, April 2010.

M. Petrova, N. Olano and P. Mähönen, "Balls and Bins Distributed Load Balancing Algorithm for Channel Allocation," in *Proc. of the 7th Annual Conference on Wireless On Demand Network Systems and Services (WONS'10)*, Kranjska Gora, Slovenia, February 2010.

J. Riihijärvi, **M. Petrova**, P. Mähönen, "A Common Application Requirement Interface for Cognitive Wireless Networks," in *Proc. of the IEEE Workshop on Networking Technologies for SDR Networks in conjunction with IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, Rome, Italy, June 2009.

M. Ahmed, V. Kolar, **M. Petrova**, P. Mähönen and S. Hailes, "Component-based Architecture for Cognitive Radio Resource Manager," in *Proc. of the 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM'09)*, Hannover, Germany, June 2009.

J. Riihijärvi, P. Mähönen, **M. Petrova** and V. Kolar, "Enhancing Cognitive Radios with Spatial Statistics: From Radio Environment Maps to Topology Engine," in *Proc. of the 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM'09)*, Hannover, Germany, June 2009.

C. Sokolowski, **M. Petrova**, A. de Baynast and P. Mähönen, "Cognitive Radio Testbed: Exploiting Limited Feedback in Tomorrow's Wireless Communications Networks," in *Proc. of CogNet 2008 workshop in conjunction with IEEE International Conference on Communications (ICC)*, Beijing, China, May 2008.

M. Petrova, P. Mähönen and J. Riihijärvi, "Connectivity Analysis of Clustered Ad Hoc and Mesh Networks," in *Proc. of IEEE Global Communications Conference (GLOBECOM'07)*, Washington D.C., USA, November 2007.

M. Petrova, P. Mähönen and J. Riihijärvi, "Evolution of Radio Recource Management: A Case Cognitive Radio Manager with VPI," in *Proc. of IEEE CogNet 2007 Workshop in conjunction with IEEE International Conference on Communications (ICC)*, Glasgow, UK, June 2007.

P. Mähönen and **M. Petrova**, "Flash Crowds in Cognitive Radio Environment: Beware an Unpredictable Mob," in *Proc. of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)*, Athens, Greece, September 2007.

M. Petrova, P. Mähönen, M. Wellens and J. Riihijärvi, "Cognitive Resource Management : A Cross-Layer Paradigm for Cognitive Networks ," in *Proc. of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)*, Athens, Greece, September 2007.

P. Mähönen, **M. Petrova** and J. Riihijärvi, "Applications of Topology Information for Cognitive Radios and Networks," in *Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'07)*, Dublin, Ireland, April 2007.

S. Fischer, **M. Petrova**, P. Mähönen and B. Vöcking, "Distributed Load Balancing Algorithm for Adaptive Channel Allocation for Cognitive Radios," in *Proc. of the International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom'07)*, Orlando, USA, August 2007.

M. Petrova, L. Wu, P. Mähönen and J. Riihijärvi, "Interference Measurements on Performance Degradation between Co-located IEEE 802.11g/n and IEEE 802.15.4 Networks," in *Proc. of the International Conference on Networking (ICN'07)*, Sainte-Luce, Martinique, April 2007.

P. Mähönen, **M. Petrova**, J. Riihijärvi and M. Wellens, "Cognitive Wireless Networks: Your Network Just Became a Teenager," Poster abstract in *Proc. of IEEE Conference on Computer Communications (INFOCOM'06)*, Barcelona, Spain, April 2006.

M. Petrova, J. Riihijärvi, P. Mähönen and S. Labella, "Performance Study of IEEE 802.15.4 Using Measurements and Simulations, " in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'06)*, Las Vegas, Nevada, USA, April 2006.

J. Riihijärvi, **M. Petrova**, P. Mähönen and J. de Almeida Barbosa, "Performance Evaluation of Automatic Channel Assignment Mechanism for IEEE 802.11 Based on Graph Colouring," in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06)*, Helsinki, Finland, September 2006.

J. Riihijärvi, **M. Petrova** and P. Mähönen, "Frequency Allocation for WLANs Using Graph Colouring Technique," in *Proc. of Conference on Wireless On Demand Network systems and Services (WONS'05)*, St. Moritz, Switzerland, January 2005.

J. Riihijärvi, P. Mähönen and **M. Petrova**, "Automatic Channel Allocation for Small Wireless Local Area Networks, " in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'04)*, Barcelona, Spain, September, 2004, pp. 536 - 539.

Conference papers in addition to the work in this thesis

M. Petrova, P. Mähönen and Alfredo Osuna, "Multi-Class Classification of Analog and Digital Signals in Cognitive Radio using Support Vector Machines," in *Proc. of Cognitive Communications Workshop (COGCOM) in conjunction with the International Symposium on Wireless Communication Systems (ISWCS'10)*, York, UK, September 2010.

J. Hoydis, **M. Petrova** and P. Mähönen, "Effects of Topology on Local Throughput-Capacity of Ad Hoc Networks," in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08)*, Cannes, France, September 2008.

J. Riihijärvi, **M. Petrova** and P. Mähönen, "Influence of Node Location Distributions on the Structure of Ad Hoc and Mesh Networks," in *Proc. of IEEE Global Communications Conference (GLOBECOM'08)*, New Orleans, USA, November 2008.

J. Ansari, J. Sanchez, **M. Petrova**, J. Riihijärvi, O. Raivio, K. Rerkrai, C. Jardak, F. Oldewurtel, M. Wellens, L. Wu and P. Mähönen, "Flexible Hardware/Software Platform for Tracking Applications," in *Proc. of IEEE Vehicular Technology Conference (VTC'07)*, Dublin, April 2007.

J. M. Sanchez Santana, **M. Petrova** and P. Mähönen, "UPnP Service Discovery for Heterogeneous Networks" in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06)*, Helsinki, Finland, September 2006.

J. Ansari, J. Sanchez, **M. Petrova**, J. Riihijärvi, O. Raivio, K. Rerkrai, C. Jardak, F. Oldewurtel, M. Wellens, L. Wu and P. Mähönen, "Flexible Hardware/Software Platform for Tracking Applications," Demonstration paper in *Proc. of ACM Conference on Embedded Networked Sensor Systems (SenSys)*, Boulder, Colorado, USA, November 2006.

M. Wellens, **M. Petrova**, J. Riihijärvi and P. Mähönen, "Building a Better Wireless Mousetrap: Need for More Realism in Simulations," in *Proc. of Conference on Wireless On Demand Network systems and Services (WONS'05)*, St. Moritz, Switzerland, January 2005, pp.150-157.

M. Petrova, M. Wellens, L. Wu and P. Mähönen, "Hop of No Return: Practical Limitations of Wireless Multi-Hop Networking," in *Proc. of IEEE Workshop on Multi-hop Ad hoc Networks: From Theory to Reality (REALMAN), in conjunction with IEEE International Conference on pervasive Services (ICPS'05)*, Santorini, Greece, July 2005.