ON DYNAMIC SPECTRUM ACCESS IN COGNITIVE RADIO NETWORKING

Rutabayiro Ngoga Said

Blekinge Institute of Technology
Doctoral Dissertation Series No. 2013:12

School of Computing

On Dynamic Spectrum Access in Cognitive Radio Networking

Rutabayiro Ngoga Said

Blekinge Institute of Technology doctoral dissertation series
No 2013:12

On Dynamic Spectrum Access in Cognitive Radio Networking

Rutabayiro Ngoga Said

Doctoral Dissertation in
Telecommunication Systems

School of Computing
Blekinge Institute of Technology
SWEDEN

2013 Rutabayiro Ngoga Said
School of Computing
Publisher: Blekinge Institute of Technology,
SE-371 79 Karlskrona, Sweden
Printed by Printfabriken, Karlskrona, Sweden 2013
ISBN: 978-91-7295-267-6
ISSN 1653-2090
urn:nbn:se:bth-00568

To my family

Abstract

The exploding increase of wireless communications combined with the existing inefficient usage of the licensed spectrum gives a strong impetus to the development and standardization of cognitive radio networking and communications. In this dissertation, a framework for Dynamic Spectrum Access (DSA) is first presented, which is the enabling technology for increasing the spectral efficiency of wireless communications. Based on that, Cognitive Radio (CR) can be developed as an enabling technology for supporting the DSA, which means that the wireless users are provided with enhanced capability for sensing the operating radio environment and for exploiting the network side information obtained from this sensing.

The DSA concept means that the users of a wireless system are divided into a multi-tiered hierarchy with the primary users (PUs) entitled to protection and with cognitive radio capable secondary users (SUs). The improved spectrum efficiency is obtained by means of a medium access control protocol with knowledge about the statistical properties or available local information of the channels already occupied by PUs as well as knowledge about the interference tolerance within which the interference to PUs is kept to a given level. Related to this, emphasis is laid on the protocol capability to determine the efficiency of the secondary sharing of spectrum. Based on the type of available local information, the capacity of opportunistic communication is investigated for three models. These

are: with dynamic, distributed channels information; with dynamic, parallel channels information; and under a dynamic sub-channels allocation scheme.

The results indicate that this capacity is robust with reference to the uncertainty associated with localized sensing of distributed dynamic channels and with timely sensing of parallel dynamic channels. The extension to dynamic parallel sub-channels enables resource allocation to be carried out in sub-channels. The analytical results on the performance of sub-channel allocation indicate a robust traffic capacity in terms of blocking probability, drop-out probability and delay performance as function of PUs traffic loads.

Acknowledgements

I would like to thank first and foremost my advisor, Prof. Adrian Popescu. His guidance and generous and continuous support have been a great source of encouragement through the course towards the Ph.D. degree. I am particularly grateful for his patience and humility in his dealing with me.

I would like to also express deep gratitude to the members of my supervisory committee, including Prof. Hans-Jürgen Zepernick and Dr. David Erman. I am particularly grateful for their inputs during the supervisory meetings, which have brought new perspectives and understanding in the process of forming my dissertation.

My time at Blekinge Tekniska Högskola (BTH) was made enjoyable in large part due to the colleagues and friends that have become a part of my life. In particular, I would like to thank Yong Yao and Alexandru Popescu for the many supportive discussions, which have took me to the appreciation of rigorous theoretical work. Also, I would like to thank Tahir Nawaz Minhas, Junaid Junaid, Selim Ickin, Louis Sibomana, Charles Kabiri and Christine Niyizamwiyitira. I significantly learned from each of you.

I would like to extend my thanks to my former teacher, Dr. Felix Akorli, for the decisive impact he had on my academic formation in Rwanda. I gratefully acknowledge the funding source that made my Ph.D. work possible, the National University of Rwanda (NUR) in partnership with the Swedish International Development Agency (SIDA).

Finally, I would like to express my deepest gratitude towards my family, who have always been my best support and encouragement even if I could not be with them most of the time during the Ph.D. study. Without you, I would not achieve what I have today.

Rutabayiro Ngoga Said

Karlskrona, December 2013

CONTENTS

PAGE

List of Figures

List of Tables

Table | Page

Acronyms

DSA	Dynamic Spectrum Access
CR	Cognitive Radio
PU	Primary User
PUs	Primary Users
SU	Secondary User
SUs	Secondary Users
RSA	Radio Spectrum Access
FDMA	Frequency Division Multiple Access
SDMA	Space Division Multiple Access
TDMA	Time Division Multiple Access
CDMA	Code Division Multiple Access
LMR	Radio Land Mobile
QoS	Quality of Service
LAN	Local Area Network
MIMO	Multiple Input Multiple Output
GSM	Global System for Mobile Communication
LTE	Long Term Evolution
ISM	Industrial, Scientific and Medical
CSMA	Carrier Sense Multiple Access
MAC	Medium Access Control
CRN	Cognitive Radio Network
PFA	Probability of False Alarm
PMD	Probability of Miss Detection
OSA	Opportunistic Spectrum Access
RRI	Round Robin Information
OCI	Ordered Channels Information
ICI	Independent Channels Information
SHI	Skewed Hopping Channels Information
CTMC	Continuous Time Markov Chain
DTMC	Discrete Time Markov Chain
RRH	Round Robin Hopping
OFDMA	Orthogonal Frequency Division Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
FSM	Finite State Machine
BS	Base Station

Chapter 1

Introduction

Abstract

This chapter defines the scope of this research. The main critical issues that need to be solved in cognitive radio networks, so as to build a wireless infrastructure supporting spectrum capacities tailored to today's societies and economies, are presented. Existing solutions and current approaches are discussed. The contributions of the thesis are presented.

1.1 Introduction

For more than a century, the radio spectrum, which formally resides in the frequency range below 3000 GHz in the electromagnetic spectrum, has been unique and valuable natural resource enabling a wide range of wireless applications and services. The earliest and most enduring communication services that use radio spectrum are radio and television broadcasting.

Doubtless, for many decades, they have had an enormous impact on society, e.g., by providing instantaneous access to news and entertainment to virtually every home.

Over the last decades, however, radio spectrum use has increased and expanded immensely beyond the classic wireless systems. The proliferation of many new wireless applications and services spawned by enormous technological breakthroughs, particularly in digital devices, has added impetus to this trend. Today, the wireless communication industry includes cell phones, Global Positioning System (GPS), smart phones and mobile broadband, WiFi, Bluetooth, satellite radio and TV, automotive radar, and Radio-Frequency Identification (RFID), to name just a few.

As economies and societies have become more mobile and information centric, wireless systems have found pervasive use and have became indispensable contributors to increasing the productivity and well being of people around the world. With further proliferation of wireless systems and the emergence of diverse wireless applications and services, the finite radio spectrum resources will be in even greater demand [1]. However, the ability of technological developments to continue to support the ever increasing demand of wireless products and services is challenged by *spectrum overcrowding*, that is, the plateauing capability of the available radio spectrum [2].

Moreover, this picture is further complicated by the fact that *wireless technology* is much more regulated at both national and international levels than other technologies in the field of information and communication technology. The detailed regulation of wireless spectrum access technology means that technical innovations do not achieve practical use unless they are consistent with regulatory requirements or designed in concert with the right policy environment [3]. Even when a new technology is compatible with the existing spectrum policy, that particular technology may be prohibited because regulations require taking cautious actions about allowing a new user that may cause harmful interference to other users who are in compliance with the regulations.

On top of that, measurement of radio spectrum use consistently shows that actual spectrum occupancy is actually low when averaged over space and time [4]. There are many reasons for this, including the spacial heterogeneity of demand and the need to set aside spectrum for military and safety uses, which are occupied based on peak demand rather than average demand. Regardless of the root causes of low spectrum occupancy, it is essential to remove the spectrum gridlocks by using appropriate means such as Cognitive Radio (CR), so as to build a wireless infrastructure supporting spectrum capacities tailored to today's societies and economies.

The key to unlock the various benefits enabled by CRs is the implementation of effective and efficient means viewing the spectrum as an open medium for multiple access by numerous simultaneous users and providing multiple services to a population with diverse requirements and resources. The main challenge is therefore to better understand the nature of traffic, technological capabilities and available tradeoffs. New communication models that exploit, in the most effective way, the spectrum sharing capabilities are needed, so that the performance benefits of this sharing can be fully realized and in which any marginal cost of such sharing provisions is paid by the user who benefits from them.

1.2 Radio Spectrum Access Methods

For more than a century, engineers and scientists have been developing Radio Spectrum Access (RSA) techniques to help better utilize and manage the radio spectrum. The origins of RSA date back in the early days of radio with the advent of Tuned Circuits, as the enabling technology for Frequency Division Multiple Access (FDMA). FDMA, Space Division Multiple Access (SDMA), Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA) were the only technical achievements that allowed use of radio spectrum by multiple simultaneous transmissions that are orthogonal with each others, and they have remained so for most of the history of radio technology. Implicitly, they became the focus of radio regulation in the sense that they provided the foundation for the band channelization approach of regulating spectrum. Governments around the world began to strictly regulate transmissions, to designate which bands are used for which purposes, and to license the use of certain frequencies. Based on that, today, we have generally two methods for managing and accessing spectrum: one method is *licensed* access, the other one is *unlicensed* access.

1.2.1 Licensed Access

A licensed system must get permission from the regulator to operate within a given frequency band. The licensing process is an opportunity for the regulator: a) to divide the spectrum into frequency bands, b) to determine which bands are used for which purposes, and c) to assign a licensee to each band, who is generally granted exclusive use of the band. Systems in this category include Radio Land Mobile (LMR) such as radios used by emergency services, cellular communications, and satellite communications.

The main advantage of licensing access is that a licensee has complete control of the allotted frequency band, and can thus unilaterally manage interference between its users and therefore enhance the Quality of Service (QoS) to its users. Among other things, this also means that a common access protocol must be defined, and must be supported by all users in this frequency band. Moreover, the need for a common access protocol forces some degree of homogeneity among devices, which in turn forms a major barrier for new technological innovations. This is also the problem when a new user attempts to share spectrum with a legacy equipment, which was not designed or deployed with new sharing arrangements support.

1.2.2 Unlicensed Access

A few spectrum have been set aside in specific frequency bands for unlicensed access, to be more precise, bands where devices are "licensed by rules". In other words, any device that complies with the regulator rules, such as maximum power per Hertz or use of contention-based algorithms, can be deployed without explicit permission. Equivalently, a regulator can explicitly grant an unlimited number of nonexclusive licenses with some rules associated with these licenses. Unfortunately, the unlicensed bands can be killed by their own successes, because devices can be deployed anywhere, which means there is no limit to the number of devices that might be operating in a given location, so also is there no limit to the potential congestion.

1.3 Radio Spectrum Access Techniques

Over the past century, the major drivers of RSA technological breakthroughs have been the quest for improving the radio spectrum efficiency. This has resulted in transmissions over larger distances with better quality, less power, and smaller, cheaper devices, and thereby adding impetus for increased and immensely expended use of radio spectrum. In specific, means for improving spectrum efficiency divide in three general classes: more geographic reuse of spectrum, increasing transmission throughputs, the packing of offered traffic in a more effective way into available transmission capacity.

1.3.1 Radio Spectrum Reuse

Frequency reuse is a simple spectrum-efficient technique that allows to increase the amount of traffic that a frequency band can carry within a geographic area, measured in bps/Km^2. The basic idea behind frequency reuse is to exploit the fact that the power of a transmitted signal falls off with distance. Thus, by adding more infrastructure, more users at spatially-separate locations can operate at the same frequency band with minimal interference to one another. In particular, for the cellular infrastructure, the entire cellular band is divided into N sub-bands and these bands are assigned to cells. Cells that use the same band are separated in distance by cells that are using different bands. In this way, interference among cells is minimized by virtue of propagation losses while spectrum efficiency increases.

1.3.2 Raw Throughput Increase

In cell splitting, however, when the cell size is small enough interference may be too large so as to prevent communications. In other words, one can only increase spectrum reuse efficiency until the density of cells reaches a certain level. A better alternative leading to higher transmission throughput gains for a given frequency band is the implementation of more advanced spectrum-efficient technologies such as spread spectrum [5], Multiple Input Multiple Output (MIMO) [6] and Orthogonal Frequency Division Multiple Access (OFDMA) [7].

Spread spectrum is a spectrum-efficient technique that grew from military applications as a technique to prevent jamming or to protect against hostile interception and interference. The basic idea is to isolate a user's signal by using a unique code which when decoded restores the original desired signal, while totally removing the effect of the other user's coded signal. Efficiency gain is obtained by reducing the effectiveness of the interfering signal by a factor known as spreading factor which is the ratio of the code rate to the original source's bit rate. For example, if the data rate is 10 Kbps with a spreading code rate of 1 Mbps the spreading factor corresponds to 100. Compared with the classic spectrum reuse, in which contiguous cells must allocate different frequency bands, a spread spectrum based technique such as CDMA can reuse the entire cellular spectrum in each cell, and thereby improving the spectrum efficiency by multiples.

For instance, in the case of MIMO, a 2x2 MIMO places two antennas at the base station and two "antennas at the user device". By antennas at the user device we mean that the antennas may belong to one user or multiple users. Because the transmitting antennas are slightly displaced from each other and that the receiving antennas are also slightly displaced from each other, each sent and received signal is subject to different multipath characteristics. By examining the four signals together, the effect is that throughput gain increases by multiples.

OFDMA can be viewed as multiple channels (hundreds, or even thousands) that are spaced with minimal distance from each other and modulated with data. By taking the advantage that the channel conditions at different users are subject to different multipath characteristics, one can dynamically allocate the channels according to the user's channel conditions. This approach provides a great flexibility to accommodate more users based on their respective QoS, and the effect is that system throughput increases by multiples.

1.3.3 Effective Data Packing

One approach to obtain spectrum efficiency gain is to use spectrum-effective techniques applicable to data contents, such as source-coding and scheduling. The use of source-coding allows to optimize data content sought by a user into a lower bandwidth format. To do so, the amount of data that needs to be transmitted is reduced by using a model of the data that can be reconstructed. The effect is a much lower data rate to support data content.

Efficiency can also be gained from providing different data applications with QoS that is tailored more precisely to their individual needs. For example, voice packets are generally produced in a steady time-synchronous mode, with traffic characterized by long holding times and modest setup times. Hence, they may accept large access delay, but put stringent requirement on the network in terms of transmit time (i.e., delay), delay variation (i.e., jitter) and bandwidth. Furthermore, they have also flexible loss sensitivity, but are generally uncontrollable (i.e., they cannot stop or slow down, vary their traffic rate during a call). In this case, they must be delivered with common QoS for all users, no matter where they are located (near of far from a base station, in a shadowed or in a clear location). Therefore, more resources are required for disadvantaged users.

With data packets, on the other hand, demands are generally produced from very low duty cycle sources, i.e., with a low activity factor, the source is active only a fraction of the time. In addition, such sources usually generate bursty information of random length. Hence, they are much more controllable than voice packets, i.e., they can stop or slow down transmissions if requested from the network. Being able to offer variable levels of service by advanced network measurement, and then by using QoS based packet scheduling, without overly penalizing the weakest users, the effect is that throughput gain increases by multiples.

1.4 Radio Spectrum Access Applications

By far, the most prevalent applications of RSA have been the cellular systems and wireless Local Area Network (LAN)s.

1.4.1 Cellular Systems

The cellular telephone system is a prevalent example of one of the most successful applications that evidences the technological impact of RSA. The first generation (1G) of cellular systems employed FDMA as spectrum sharing technology to handle analog communications. The second generation (2G) of cellular system moved from analog to digital transmissions to embrace the many advantages provided by digital systems. The components are cheaper, faster, smaller, and require less power. Gradually, these system evolved to support techniques such as TDMA and CDMA as alternative technologies for spectrum sharing. The impetus that began with Global System for Mobile Communication (GSM) system was towards TDMA, which allowed up to eight users to share one TDMA channel in the alloted frequency.

A major breakthrough in cellular systems came with the advent of interference suppression techniques such as spread spectrum or CDMA that grew from military origins. Such techniques not only permit more users to be added to the system, but also solve the efficiency reduction of FDMA and TDMA necessitated by the requirement to assign different frequency allocations to neighboring cells so as to avoid mutual interference.

While second generation cellular systems initially provided mainly voice services, these systems have been gradually enhanced to support data services such as email, Internet access, and short messaging as well. The ability to incorporate effective data packing techniques into 2G system provided the foundation to the deployment of the third generation (3G) of cellular system. The coming of new strategies for improving spectrum efficient of the alloted frequency bands such as OFDMA and MIMO have steered the need to deploy the fourth generation systems, such as Long Term Evolution (LTE) and WiMAX, and beyond.

1.4.2 Wireless Local Area Networks

The advent of wireless LANs is another example that evidences the technological impact of RSA to increasing spectrum use. Wireless LANs date back to around 1985 when the public Industrial, Scientific and Medical (ISM) bands were authorized in U.S.A for wireless LAN products [8]. The ISM band became very attractive to wireless LAN vendors since they did not need to obtain license to operate in this band.

Before that, it was believed that these bands were nearly useless for communications due to interference from the primary users of these bands such as microwave ovens and radio frequency heating devices. Spread spectrum techniques, such as frequency hopping and direct sequence, long established for military applications as techniques to thwart intentional jamming, made the means to use these otherwise polluted bands feasible. As a result many applications and technologies, including devices for 802.11/b/g/n, ZigBee, Bluetooth, and cordless phones have emerged from research ideas to concrete systems.

1.4.3 Cognitive Radio Technology

The ever growing demand for wireless communications along with the recognition that classic approaches to spectrum management leads to inefficient use of spectrum has spurred interest in CR technologies. Based on this, far more inexpensive, flexible, and reconfigurable hardware and software can be developed into more flexible RSA to support far more improved spectrum sharing.

Although there have been many technological breakthroughs in RSA and policies, since the early days of radio, FDMA use of spectrum has remained the dominant method to keep operators from causing interference to one another to date. While a *gridlock* approach, such as FDMA use of spectrum, was appropriate to reduce interference from dissimilar systems in the early days of radio, and reasonable within the technological capabilities of that time, it does not now lend itself to efficient spectrum use, unless significant adjustments in how spectrum is used and shared are considered.

Therefore, the advent of CR technology, a framework for adding intelligence in communication systems, holds a tremendous promise for both greater access to and greater utility from the radio spectrum resource through new approaches to spectrum access that embrace spectrum sharing in its various forms [9]. Specifically, the key to unlocking the many benefits of CR is the implementation of efficient means viewing the radio spectrum as an open resource for multiple access by numerous simultaneous users and providing multiple services to a population with diverse requirements and resources. This, however, imposes the search for new architectural communication models.

The traditional approach to wireless networking follows the current Internet that provides a general transport service capable of supporting many different applications [10]. This "data layer" is not designed or optimized for any single application, but it is designed for generality and adaptability. Parallel to this data layer sits a control layer for which the main function is to manage the network infrastructure and to ensure reliable flow of data.

Although a powerful approach for functionality allocation, in that it facilitates innovation and the deployment of new applications, often both the data and control planes are equally oblivious to detailed understanding of the environments in which the protocols or services are used. Cognitive radio networking is a new construct, a distributed layer or a pervasive system within the network that builds and maintains high-level view of what the network is supposed to do, i.e., the goals of its designers, of the applications running on it, and of its users, and thereby provide services to other elements of the network.

More specifically, the CR approach differs from the traditional, algorithmic approach in that a network is augmented with a new set of mechanisms in the form of cognitive capability and reconfigurability. The cognitive capability enables a network to enhance its awareness of the environment, which means the ability of the network to gain necessary information from the operating radio environment. This includes knowledge about the transmitted waveform, radio frequency (RF) spectrum, communication network type/protocol, geographical information, locally available resources and services, user needs, security policy, and so on. Whereas reconfigurability enables the network to dynamically change its transmission parameters according to the sensed environment variations so as to achieve optimal performance.

1.5 Solution Approach

Much of the interest in RSA comes from the need to match the increasing wireless demands of today's societies and the recognition that the current gridlock approach to spectrum use results in inefficient use of the radio spectrum.

Therefore, in this dissertation, the alternative approach to RSA, called Dynamic Spectrum Access (DSA), is first presented as a means for alleviating spectrum gridlocks by using spectrum sharing. Based on this, CR can be developed to support DSA. The DSA concept means that the users of a wireless system are divided into a multi-tiered hierarchy with the Primary Users (PUs) entitled to protection and cognitive radio capable Secondary Users (SUs). DSA views the spectrum as an open resource to be shared among PUs and SUs in such a way that the performance benefits of such a sharing arrangement can be fully realized while, at the same time, avoiding excessive interference caused to PUs. The performance sharing gains are obtained by means of a medium access control protocol with knowledge of the statistical properties or available local information of the channels already occupied by PUs as well as knowledge of the interference tolerance within which the interference to PUs is kept to a given level.

For the purposes of our study, three classes of communication models, that describe several situations arising in a cognitive communication process, are chosen considering the nature of information available about the channels over which transmissions of Secondary User (SU) take place. These are: with dynamic, distributed channels information; with dynamic, parallel channels information; and under a dynamic sub-channels allocation scheme. Related to this, emphasis is laid on the Medium Access Control (MAC) protocol capabilities to determine the efficiency of the secondary sharing of spectrum.

The various channel models are studied in-depth in order to obtain and to evaluate the performance measures. Detailed theoretical studies of modeling and performance analysis are conducted. The performance aspects of the MAC protocols, which enable reliable communications over these channels, are evaluated with the purpose of investigating the limits and advantage of the DSA approach.

The performance results clearly indicate the feasibility of the DSA concept as a candidate for future wireless infrastructures that have the potential to support economic growth and evolving societal needs. This approach offers the choice of performance that is restricted to the capabilities of spectrum sharing arrangements, and less on the spectrum gridlocks. Specifically, it offers robust performance with reference to the uncertainty associated with localized sensing of distributed, dynamic channels and with timely sensing of parallel, dynamic channels. The extension to dynamic, parallel sub-channels enables resource allocation to be carried out in sub-channels. The analytical results on the performance of sub-channel allocation indicate a robust traffic capacity in terms of blocking probability, drop-out probability and delay performance as function of PUs traffic loads.

1.6 Related Work

In this section, we discuss related work in the literature. Since for SUs communications, the transmitter and the receiver must be designed with the occupancy knowledge of the channels of Primary User (PU)s over which transmission takes place, we first present the physical and statistical nature of cognitive radio channels. We further discuss about recent efforts to model such channels and the corresponding capacities measured in terms of throughput to SUs and the constraint over collision caused to PUs. Special emphasis is laid on the MAC, which enable reliable communication over these channels. We also introduce the work related to our two major applications, namely opportunistic access over multiple wireless channels and admission control in subchannels.

1.6.1 CR Channel Characteristics

In the following, we highlight the defining features of a cognitive radio channel.

- A cognitive channel is a local concept defined with respect to a particular pair of SUs and resources available. It depends on the location of not only the SU transmitter but also the secondary SU receiver. It also depends on the hardware and the energy cost associated with the spectrum monitoring capability of SUs.
- A cognitive channel is determined by the communication activities of PUs rather that of SUs. At any time and location whether the channel is an opportunity or not depends on the PUs activity on that particular time and location.
- Only failed communications caused by collisions to a PUs disqualify a channel from being an opportunity. A cognitive channel is defined based on the interference constraint through packet loss that a PU can tolerate.

Due to the physical separation between the SU transmitter and SU receiver, cognitive channels are *dynamic* and *distributed* [11]. Considering the diverse activity of different PUs in various spectrum frequency bands, cognitive channels can also be *dynamic* and *parallel* [12].

1.6.2 CR Channel Models

We now present a variety of mathematical models for SU communication over a channel of PUs, as found in the literature. Focus is put on those models and assumptions which are relevant for a MAC approach, and we elaborate on those assumptions leading to the required simplifications, giving rise to a rigorous mathematical treatment.

As SU has to be aware of the channel status at all times, periodic sensing is essential in a cognitive radio network. This means that spectrum sensing should execute frequently to quickly detect the presence of PUs when they return to occupy a channel. Usually, this case translates into a time channel model of duration corresponding to the minimum interval between two consecutive sensing activities. This can be achieved by using a slotted protocol [13]. In such a protocol, the slots are of fixed duration corresponding to the packet used in the SU system, in particular each slot consists of a sensing period and a transmission period [14].

The problem is complicated by considering that sensing and communication share the same transceiver (antenna and RF front end), for which a tradeoff between sensing performance and spectrum access needs to be considered [15,16]. This problem can be simplified by hiding the distributed nature of the channels and imposing a slotted structure on both the SUs and PUs. As such, at the beginning of each slot there is a designated time period for PU to reserve this particular slot and for SU to perform channel sensing. If a PU intends to use a slot it will simply remain active, in which case a SU will find the channel busy after the sensing period and not attempt to use it for the remainder of the slot. If the primary user is inactive during this sensing period, then the remainder of the slot is open to secondary users.

By taking into consideration the fact that in practical scenarios sensing errors are inevitable, spectrum sensing should generally perform to meet certain target Probability of False Alarm (PFA) and Probability of Miss Detection (PMD) [17]. Emphasis is laid on the impact of sensing errors on the MAC performance in terms of throughput for SUs and packet collision at PUs. In specific, a small packet collision at PUs can only be obtained with smaller performance in PMD. At the same time, a small performance in PFA yields larger throughput for SUs. Hence, the design of spectrum sensing faces a trade off between maximum throughput for SUs and packet loss at PUs, for which a balance between the PMD and PFA has to be reached [18, 19].

To do so, one approach involves to allocate a fixed portion of each time slot of SU as channel sensing period, and the duration of the sensing period needs to be long enough so as to minimize the PMD at PUs under the constraint that the PFA for SUs is kept below a prescribed threshold [20]. An alternative approach involves to optimize the sensing period in such a way that yields minimum PFA for SUs under the constraint that the PMD at PUs is kept below a certain threshold [21]. Also, it is possible to maintain a fixed portion of each time slot as SU packet size, and then one can optimize the SU packet size [22]. Equivalently, one may maintain both the sensing period and the transmission period fixed, and then optimize the probability of transmissions [23].

Having a certain structure imposed on both PUs and SUs introduces some degree of homogeneity between the SU system and the PU system. Consequently, if both PU and SU uses a common slotted communication protocol, then a PU is either present or absent during the whole duration of a time slot. Besides, if a SU correctly senses the channel to be idle, this obviates the need for protection demanded by PUs. Removing such a constraint complicates the problem even further because a PU may start or stop occupying the channel at any time [24]. This means that, regardless of perfect sensing, there is always a chance that the SU transmissions interfere that of PU. In this case, constraining the interference caused to PUs is non trivial [25].

If we look purely at the effect of PUs traffic, the above problem simplifies. One simple way is to assume that the PUs activities on a particular channel follow an alternating idle-busy pattern [26]. This obviates other sources of channel error, except in the event that a PU returns after the SU has already sensed the channel to be idle and started a transmission, which may occur at the beginning of each busy-idle cycle.

To select a model for a channel in such a situation, a slotted protocol is devised in such a way that the throughput to SUs is maximized under the constraint that the PU packet collision is no greater than some threshold [27]. The slots correspond to the transmission time of the packets used in the SU system when the channel is idle. The protocol may include an additional parameter β, the probability that a packet persists, and with $(1-\beta)$ being the probability of delaying transmission [28]. Alternatively, one can allocate a *vacation period* that consists of a fixed number K of time slots before each transmission slot [29]. Again one can also fix the packet length in each idle-busy cycle to L time slots [30].

One simple way to handle such a protocol is to assume that the PU idle-busy cycle (i.e., the channel law) follows exponential distributions. In this case the protocol parameter can be optimized so that the resulting capacity depends on simple average values only [31]. In a more practical scenario, the PU idle-busy cycle are generally distributed. In this case, the protocol parameter β [32] and K [33] as well as their respective channel capacities depend on additional details about the nature traffic such as the distribution of idle period. If we consider the distributed nature of cognitive channels, the problem become more complex in the sense that the optimization results also depend on the transmission protocol of PUs and SUs and the environmental uncertainties over which the transmissions of SUs take places.

1.6.3 Multi-channel CR Communications

In the following, we focus on the work related to the use of multiple channels capability to improve the capacity of the channels over which secondary transmissions take place. Multi-channel protocols for ad hoc networks present themselves as a first step to develop a cognitive MAC protocol. They operate in multi-channel context [34]. The number of channels over which a host has access is limited by the number of interfaces at the host [35]. A CR, however, is augmented with the set of mechanisms: a) to sense the operating radio environment, b) to distinguish between SUs and PUs, and c) to provide protection to PUs [36].

In selecting a protocol for a SU communication over multiple channels of PUs, one simple way is to solve the problem of opportunistic communication where both the transmitter and the receiver have complete knowledge of the channels model over which transmissions take place. As such, the channel selection process and access can be optimized accordingly [37]. Complete knowledge about the model governing the multiple channels arises when both the transmitter and receiver can permanently sense the spectral activity in all frequency bands at one time, and communicate opportunistically by selecting the channel whose transmission quality is relatively favorable.

In practical scenarios, however, continuous sensing of the frequency bands is challenging because it is hardware demanding and it consumes energy, especially for low-cost battery-powered handheld devices with bursty traffic. Instead, one should treat the problem of a communication situation where the transmitter and receiver have limited information about the channels over which transmissions occur [38]. The problem is further complicated when the constraint of imposing the same protocol between SUs and PUs is removed. In this case, one faces a trade off of performance for SUs against protection demanded by PUs, and that the trade off needs to adapt to the spectrum sensing deficiencies [39].

1.6.4 Admission Control in Cognitive Radio Channels

In the following, we consider the works that address issues related to traffic-performance in CR networks. Emphasis is laid on the traffic issues that affect the operation of a CR network and techniques that are used to analyze such a network. The traffic capacity is therefore the main element that encapsulates the effects of the traffic behavior of users in the operation of the network.

Methods that are used in the analysis of classic cellular networks provide the basis of modeling the operation of CR networks. These include modeling tools such as M/G/N loss model for a circuit switched link and M/G/1 single server queue model for a packet-switched link [40]. They also include issues of channel assignment, access, handoff control, traffic modeling and traffic management [41]. However, the operational features of CR networks are significantly different from those of classical cellular networks, and we briefly elaborate on those features in order to emphasize that the traffic behavior of CR networks are expected to differ substantially from those of traditional cellular networks.

Spectrum sensing is a very important feature in the operation of CR networks. SUs are required to pause any ongoing communication in order to detect the presence of PUs on a particular channel. In addition, SUs must execute spectrum sensing frequently to react quickly whenever a PU returns. To handle these requirements, transmission and spectrum sensing phases are typically interleaved in a cyclic manner. Also essential to the operation of CR network is the fact that PUs always have a higher priority over SUs. As such, if a PU returns and occupies a channel, every SU using this particular channel must halt its transmission, and thereafter it needs to hand over to another unused channel to continue its communication.

Hierarchically overlaid layout schemes have been suggested as one way to handle the operation of CR networks [42–46]. A system that employs SU network with overlaying PUs network is considered [42, 43]. [44] considers a M/M/N loss model with finite user population adapted to the tiered SU-PU architecture to analyze the performance of the channel assignment operation in CR network. In [45] a M/M/N queue model of CR network with infinite user population is considered. [46] considers that new calls and handoff calls enter at both the PU network and the SU network.

1.7 Thesis Outline

Much of the interest in RSA comes from the need to match increasing wireless demands of today's societies and the recognition that the current gridlock approach to spectrum use results in inefficient use of the radio spectrum. Our goal in this thesis is to address the problems encountered in alleviating the spectrum gridlocks by using CR techniques, and an alternative approach to RSA is advanced to solve these problems.

More specifically, three classes of communication models that describe several situations arising in a cognitive communication process are chosen considering the nature of information available about the channels over which transmissions of SUs take place. These are: with dynamic, distributed channels information; with dynamic, parallel channels information; and under a dynamic sub-channels allocation scheme. Related to this, emphasis is laid on the MAC protocol capabilities to determine the efficiency of the secondary sharing of spectrum.

The various channel models are studied in-depth in order to obtain and to evaluate the performance measures. Detailed theoretical studies of modeling and performance analysis are conducted. The performance aspects of the MAC protocols that enable reliable communications over these channels are evaluated with the purpose of investigating the limits and advantage of the DSA approach.

In Chapter 2, the first DSA model considered in the dissertation is presented. We investigate the capacity of opportunistic communication for a single SU that uses a listen-before-talk protocol as access scheme, in the presence of distributed, dynamic channels occupied by PUs. A two-mode timeout switch modeling of the channel occupancy characteristics by PUs is assumed at both SU end stations (i.e., the SU transmitter and the SU receiver). A statistical characterization of this channel model is developed and important properties are described. We obtain the protocol capacity for a single user cognitive channel, and show that the system capacity is robust with reference to uncertainty associated with the distributed, dynamic channels occupied by PUs.

In Chapter 3, the second DSA model considered in the dissertation is presented. Efficient transmission strategies for SUs that use a slotted protocol as access scheme in the presence of parallel, dynamic channels already occupied by PUs traffic are studied. In every time slot, a channel is selected and the probability of transmission in the selected channel is estimated by using local information about the quality of these channels. A SU transmits in a probabilistic manner in the selected channel based on this estimate. A Markovian decision model that embeds such a problem in its control actions is formulated. Optimal decision rules are found. The protocol capability in terms of throughput to SUs and packet error rate for PUs is expressed as reward and cost tradeoffs of the resulting Markovian decision processes. Numerical results on the performance of the suggested control access procedures are shown for nominal system.

In Chapter 4, the third DSA model considered in the dissertation is presented. We study the admission control of SUs to achieve a certain degree of QoS. A multi-dimensional Markov chain whose size increases quickly with the number of channels of PUs is assumed to characterize the known system. In the presence of uncertainty associated with the system size reduction, we consider a one dimension Markov chain approximation to characterize the uncertain system. Appropriate analytical models are developed and used to derive the performance characteristics of the secondary admission strategies. Numerical results are reported and they show a robust capacity in terms of blocking probability, dropping probability and delay performance as functions of spectrum utilization by PUs.

Finally, the Chapter 5 concludes the dissertation.

1.8 Contributions of the Thesis

In this thesis, we address the problem encountered in alleviating the spectrum gridlocks by using CR techniques, and an alternative approach to RSA is advanced to solve this problem. The major contributions of this thesis are:

- We introduce a two-mode timeout switch that is flexible to capture the channel occupancy characteristics by PUs as a mathematical model of cognitive channel. A statistical characterization of this channel model is developed and important properties such as the gain and cost of channel use in terms of PUs traffic are presented. We study the capacity of this channel where an opportunistic communication of SU that uses a listen-before-talk protocol as access scheme is assumed. We show that the capacity is robust, in the sense that the protocol achieves throughput while the interference tolerance is guaranteed to be satisfactory, in the presence of channels uncertainty.

- We study control procedures for multi-channel opportunistic spectrum access networks. A Markovian model characterization of cognitive channels is presented first. Three classes of control of secondary access to these channels are described. A general Markovian decision model is next formulated by including the suggested classes of access control procedure as special cases to the model. The protocol capacity, expressed in terms of throughput for SUs and packet error rates to PUs, is obtained from the resulting decision processes as cost and reward tradeoffs.

- We generalize the single user problem of access control to a multiuser access setting, where multiple SUs contend to access the channels. A multi-user access control strategy is described leading to a Markov chain model of the system. It is shown that if the exact knowledge of the system states is known to all CR hosts, it is possible to tune the strategy to an optimal value from which we obtain maximum throughput and minimum delay performance by solving the induced Markov chains.

- We study admission control of SUs to achieve a certain degree of QoS.

A multi-dimensional Markov chain whose size increases quickly with the number of channels of PUs is assumed to characterize the known system. In the presence of uncertainty associated with the system size reduction, we consider a one dimensional Markov chain approximation to characterize the uncertain system. Appropriate analytical models are developed and used to derive the performance characteristics of the secondary admission strategies. Numerical results are reported and show that the capacity expressed in terms of blocking probability, dropping probability and delay performance as functions of spectrum utilization by PUs is robust.

The work presented in this thesis was partly published in [47], [48] and [49].

Chapter 2

Spectrum Opportunity Exploration

Abstract

In this chapter, the first DSA model in the dissertation is presented. We investigate the capacity of opportunistic communication for a single SU that uses a listen-before-talk protocol as access scheme, in the presence of distributed, dynamic channels occupied by PUs. A two-mode timeout switch modeling of the channel occupancy characteristics by PUs is assumed at both SU end stations (i.e., the SU transmitter and the SU receiver). A statistical characterization of this channel model is developed and important properties are described. We obtain the protocol capacity for a single user cognitive channel, and show that the system capacity is robust with reference to uncertainty associated with the distributed, dynamic channels occupied by PUs.

2.1 Introduction

2.1.1 Overview

Cognitive radio networks have emerged as a system to facilitate the use of DSA for increased spectral efficiency of wireless networks [50]. Contrary to the exclusive access model, a block of radio spectrum in the DSA model can support a primary−secondary sharing arrangement [51]. One user has the right to operate as the PU, that is with higher priority of usage over the spectrum. The other user operates as the SU with access to this spectrum as long as PU is not using it and provided that the PU can be properly protected. DSA as a design methodology plays an enabling role for dynamically reconfiguring cognitive radio (CR) networks so that a radio spectrum allocated to PUs can be reused in an opportunistic manner by SUs while protection to PUs is guaranteed [52].

2.1.2 Motivation

The fundamental premise for the applicability of DSA lays on the fact that recent measurement studies on the actual spectrum utilization have revealed that a large portion of the radio spectrum experiences multiple idle periods with sparse periods of heavy utilization and congestion [53]. Other assumption is that cognitive radio can be developed to support DSA: which means to extend the capability of wireless users to sense and gather information from their surroundings as well as to optimally adapt their operating parameters according to local information obtained by this sensing [54].

A spectrum opportunity is a local concept defined with reference to the local spectrum occupancy by PUs perceived at each end of a cognitive channel. This implies that, because of the physical separation between a

SU transmitter and a SU receiver, the PU spectrum occupancy information at one end of this particular channel is different compared to the other end. This effect arises because of the random location of PUs in a given area. Therefore, one of the prominent characteristics of a cognitive channel is its *distributed* nature. We use the term "distributed" by referring to the spatial distribution of PU hosts.

Another important characteristic of a cognitive channel is its *dynamic* nature. Analysis of spectrum occupancy data shows multiple idle periods with sparse periods of heavy utilization and congestion [55]. Therefore, we use the term *dynamic* with reference to time variations inherent of spectrum occupancy by PUs.

In the presence of *distributed* and *dynamic* nature of cognitive radio channels, a SU transmitter senses the channels to detect the absence of PUs by using spectrum sensing. If so, it queries the *readiness* of the SU receiver by transmitting a control packet, and it decides whether to use the channel or not depending on its transmission protocol, and the result of the query. As access scheme, we suggest a slotted ALOHA protocol in which a SU transmitter decides how much and when to use the channel in such a way as to maximize throughput to SUs subject to throughput and packet loss constraints at PUs.

We show that a transmitter can attain the above performance objectives by using a simple timeout based switch, in which a SU transmitter is allowed to use the channel during the idle period only until the timeout reaches a certain threshold. Such a threshold timeout-based switch is simple to implement provided that a SU transmitter knows the optimal PU packet loss threshold, as the SU transmitter needs not to have complete knowledge about the spatial distribution of PUs and the load configurations of PUs, i.e., the exact number of PUs and the length of the messages transmitted on the channel by PUs.

2.1.3 Contributions

The main contributions of this chapter are summarized as follows.

- We introduce a two-mode timeout switch that is flexible to capture the channel occupancy characteristics by PUs as a mathematical model of cognitive channel.

- A statistical characterization of this channel model is developed and the important properties such as the gain and cost of channel use in terms of PUs traffic are presented.

- We study the capacity of this channel where an opportunistic communication of SU that uses a slotted protocol as access scheme is assumed.

- We show that the capacity is robust, in the sense that the protocol achieves throughput while the inference tolerance is guaranteed to be satisfactory, in the presence of channels uncertainty.

2.2 System Model

In this section, the notion of cognitive radio channel in the context of DSA is presented. After that, the cognitive channel model is described.

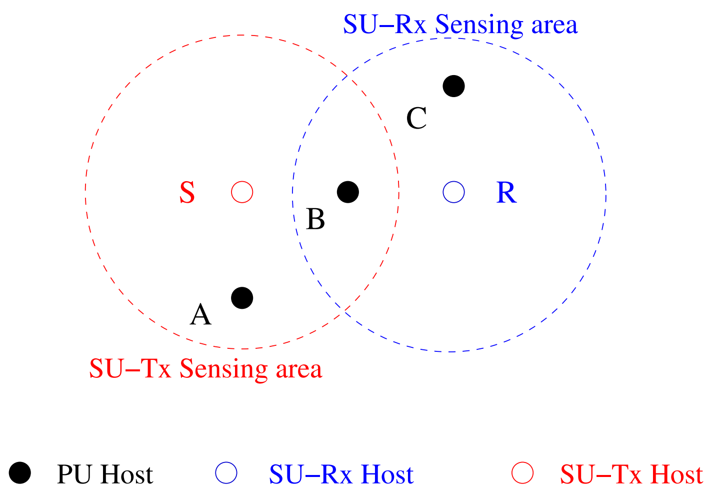

Figure 2.1: *Illustration of a cognitive radio channel*

2.2.1 System Operation

Figure 2.1 illustrates a hierarchical layout overlaid with a wireless network. The layout provides coverage for a pair of SUs at the lower level in the hierarchy. For clarity purposes, we use the open circles marked S and R to represent a cognitive radio transmitter and a cognitive radio receiver of the pair of SUs, respectively. The overlaying wireless network, on the other hand, covers a set of PU hosts that are randomly located, and provides idle channels for secondary communications whenever there is no PU present. The random location of PU hosts can be thought of as representing an arbitrary network or a snapshot of a mobile one. For simplicity purposes, we consider only three PU hosts represented by filled circles A, B and C.

We consider the system to be operated in such a way that PU communications occurring at the higher level in hierarchy alternate between *idle* state and *busy* state. The time scale for a change of channel state is much longer compared to the length of PU packet. Such a pattern in channel utilization is typical of wireless LANs, where all PUs are assumed to share a common wideband radio channel and to be in range and in line-of-sight of each other. During a busy period, the network adjusts itself by means of a transmission probability so as to give the best throughput. Consequently, each PU host with a packet ready for transmission has to wait until it is permitted to send a transmission, the restraint being the transmission probability. Hence, the channel maintains its busy state whenever there is at least one packet transmission using the channel. When all packets of PUs have been transmitted, the channel becomes idle.

A cognitive radio specific challenge is that a cognitive radio transmitter may sense *some* but *not all* PU activities in its neighborhood. For example in Figure 2.1, the dotted circles around a cognitive radio transmitter and receiver represent their respective sensing localities, this means that PU activities can only be sensed within this area. Hence, the SU transmitter $\boldsymbol{S}$ can detect the presence or the absence of the PUs $\boldsymbol{A}$ and $\boldsymbol{B}$, but not $\boldsymbol{C}$. Similarly, the receiver $\boldsymbol{R}$ can detect the presence of absence of the PUs $\boldsymbol{B}$ and $\boldsymbol{C}$, but not $\boldsymbol{A}$.

In such a case, the effect perceived on secondary communications is that a cognitive radio receiver is not always *ready* to receive. For example in Figure 2.1, when the PU $\boldsymbol{C}$ is transmitting, the receiver $\boldsymbol{R}$ cannot receive the transmission from the transmitter $\boldsymbol{S}$, though the channels at $\boldsymbol{A}$ and $\boldsymbol{B}$ are idle, as both transmissions will destructively interfere at $\boldsymbol{C}$.

2.2.2 MAC layer Cognitive Radio Channel

The objective is to design efficient transmission strategies for a wireless communication with a MAC layer cognitive radio channel. A single cognitive radio channel is shared by a group of PUs and a pair of SUs. A transmission strategy of SUs depends on the access protocol used by PUs and the idle period statistics characterization of idle periods. The key point is that such a strategy allows PUs and SUs to share simultaneously the same channel; it offers a solution which handles the geographical dispersion of PUs and which at the same time takes advantage of the channel idleness to improve the spectral efficiency.

We consider a single cognitive radio channel shared by N PU hosts and one pair of SUs. As access mode for SUs, we assume a time slotted protocol. The slots correspond to the transmission time of a SU packet, which is much smaller compared to that of PUs. Smaller values of the SU packet length allow more freedom for designing the SU access strategy. Each SU time slot consists of a sensing phase, followed by a data transmission phase.

Irrespective of the channel state, a SU transmitter and receiver sense the channels at PUs during the sensing phase. If the channel is deemed busy, no transmission is attempted during the data transmission phase, and the SU transmitter and receiver will sense the channels again in the sensing phase of the next time slot. The channel maintains its busy state whenever there is a PU packet transmission on this particular channel. If the channels at a SU transmitter are deemed idle, the SU transmitter queries the readiness state of the SU receiver by using a control packet, and it decides whether to transmit or not depending on some transmission strategies, and the result of the query. Similarly, the channel maintains its idle state until the first packet of PU arrives to occupy the channel.

Figure 2.2: *Illustration of a cognitive radio channel model*

We consider three QoS metrics in this particular channel to measure the transmission efficiency of the secondary sharing of channel: a) throughput to PUs, b) throughput to SUs, and c) packet loss to PUs. Our objective is to find a transmission strategy for which throughput to SUs is maximized subject to throughput and PU packet loss to PU constraints.

2.3 Mathematical Model

In this section, the mathematical model of a cognitive radio channel is presented as well as its important statistical properties. We use a *timeout switched* channel model as illustrated in Figure 2.2 to model the conceptual cognitive radio channel of Figure 2.1.

Two switches that model the local spectral activity as shown in Figure 2.2 are assumed at both SU tr and receiver ends. When the switch at the SU transmitter is open, we say the switch is in OFF state. This happens if a busy channel (i.e., the receiving activity of PUs) is detected in the neighborhood of a SU transmitter. On the other hand, the switch is said to be in state ON, when this switch is closed. This happens when the channel is experienced idle at the SU transmitter. The switch at the receiver is used to introduce the channel uncertainty that results from the distributed nature of the channel.

One approach to resolve this uncertainty is to have control messages exchanged between transmitter and receiver upon idle sensing result at the transmitter. A transmitter queries the readiness state of the receiver by using a request-to-send (RTS) message, which indicates an idle channel experience at the transmitter side. A clear-to-send (CTS) message successfully received at the SU transmitter indicates that the channel is also available at the receiver side, which is then followed by data transmission. Although, the RTS-CTS protocol obviates the channel uncertainty by introducing channel overheads, it also serves to indicate the readiness states of the SU receiver. Therefore, we consider only the two-mode switch at the receiver to model the distributed and dynamic nature of the considered channel.

In particular, the dynamic nature of the channels is contained in the sojourn times in the ON and OFF states of the switch, respectively. In other words, the rate at which the switch changes from ON to OFF, which reflects the idle periods of the channels at PUs, is directly related to the sojourn time in ON state. Similarly, the rate at which the switch changes from OFF to ON reflects the busy periods of the channels at PUs, and it is directly related to the sojourn time in OFF state.

Since the sojourn time in OFF state is solely affected by the transmission protocol used by PUs, we analytically estimate the sojourn time in the OFF state by developing a model with a finite number N of PUs. Our model is based on the assumption that for each transmission attempted in the PU network, a PU host uses a backoff interval sampled from a geometric distribution with parameter p. It follows that during the busy period of the channel all the processes that define the occupancy pattern of the channel by PUs traffic (e.g. backoff slots, collided packets, successful packet transmissions) are regenerative with respect to the sequence of moments corresponding to the completion of a successful packet transmission. Therefore, if we let a renewal period Z to define the time interval between two successful transmissions of PU packets, then Z and $E[Z]$ correspond to the sojourn time random variable and expectation in the OFF state, respectively. Z is also related to the theoretical throughput limit of the protocol used by PUs as follows:

$$\mathcal{P}_B = \frac{L_p}{E[Z]} \tag{2.1}$$

where L_p is the average packet payload in bits, i.e., the average time interval in a renewal period Z in which the channel is used by a successful transmission, $E[Z]$ is the average length of the renewal period, and $\mathcal{P}_B$ is the maximum achievable throughput for PUs which also represents the throughput constraint for SUs.

On the other hand, since the idle channels at the PUs can be arbitrary, we use a random variable Y to define the time interval from the moment a SU receiver is deemed ready to receive until the moment of the first packet of PUs arrives to occupy the channel. This corresponds to the sojourn time in ON state, and it is completely specified by the probability density function of Y, $f_Y(\cdot)$, and the cumulative distribution function of Y, $F_Y(\cdot)$.

By exploiting the expression in Equation 2.1 along with the statistical properties of the idle distribution $F_Y(\cdot)$, the analysis of SUs transmission protocol capacity, which indicates the efficiency of spectrum sharing in the context of DSA, can be performed by studying system behavior in a generic renewal period Z and a generic idle period Y.

2.3.1 PU Network and Spectrum

In the following, we study system behavior in a generic renewal period Z. We find the capacity of PU network under study, which is measured by the maximum achievable throughput $\mathcal{P}_B$ during busy period.

Load Configuration

To perform the capacity analysis of PUs access protocol, it is useful to denote by L the time it takes to complete a successful transmission by PUs. L corresponds to the time interval between the start of a transmission that does not experience a collision and the reception of corresponding acknowledgement (ACK) plus a Distributed Inter Frame Space (DIFS):

$$L = L_{ov} + L_{tr} \tag{2.2}$$

where L_{tr} denotes the packet transmission time at a given data rate and L_{ov} being the total time of packet overheads.

In many existing wireless LANs, devices are designed with the capability of transmitting using multiple data rates in order to improve spectrum efficiency. Depending on the quality of radio transmissions experienced by users, they are given the option to set a transmitting data rate that is tailored to prevailing channel conditions. For example, the basic rate set in IEEE 802.11b products includes 1, 2, 5.5, and 11 Mbps.

Considering the geographic dispersion of PU hosts, some hosts may be located far away from their access point so that the quality of their radio transmissions is low. In response to bad channel conditions experienced by these hosts, they reduce the data rate to some lower value, say 1 Mbps. We refer to such a PU host as *slow* PU, because it captures the channel for a long time by transmitting at lower data rate. Similarly, we refer to a PU host as *fast* PU when a PU is transmitting at the highest feasible data rate, i.e., 11 Mbps because of the proximity to the access point. In our context, every PU that is inside the sensing area of a SU transmitter corresponds to a fast PU, while every PU that is located outside the sensing area of the SU transmitter corresponds to a slow PU.

For simplicity purposes, a population of N PU hosts is assumed: one slow PU and $(N-1)$ fast PU. We define by L_f the time it takes to complete a successful transmission by a fast PU at the highest feasible data rate R:

$$L_f = L_{ov}^R + L_{tr}^R \tag{2.3}$$

where $L_{tr}^R = L_p/R$, and L_p being the packet payload. Similarly, we define by L_s the time it takes to complete a successful transmission by a slow PU that is transmitting at a lower rate r

$$L_s = L_{ov}^r + L_{tr}^r \tag{2.4}$$

where $L_{tr}^r = L_p/r$ and L_p being the packet payload.

Because of the fair access provided to all PU hosts, we can assume that the transmissions of all PU hosts, regardless of their respective channel conditions, alternate so that the average transmission time $E[L]$ to complete a successful transmission can be written as follows:

$$E[L] = \left(\frac{N-1}{N}\right) L_f + \frac{1}{N} L_s \tag{2.5}$$

PU Traffic

To perform the capacity analysis of PUs access protocol it is also useful to divide the time into generic time slots G whose time durations depend on events that may occur in a particular slot. An event in a generic time slot can be either a collision, a successful transmission, or a backoff slot. Thus, a renewal period Z can be thought of as a sequence of generic time slots that may contain several collisions and backoff slots, followed by an ending generic slot containing one successful packet transmission from a tagged PU.

We are interested in counting the number of successful transmissions from a tagged PU in any slot. This is a measure of throughput to PUs during busy period, of which the maximum attainable value corresponds to the network capacity. We start by studying the number of transmissions from any PU that can be handled per slot by the PU network. This is a measure of channel utilization by PUs in a slot.

Specifically, we count each transmission from any PU as a contribution to channel utilization. We define W as a renewal period associated with the sequence of moments corresponding to the completion of packet transmissions from any PU. Since N PU hosts are contending with one another to access the channel, and that each transmission is attempted with a common transmission probability p, the probability P_t of having a PU transmission during a generic time slot is the same as the probability of having at least one PU transmission in a randomly chosen slot:

$$P_t = 1 - (1 - p)^N \tag{2.6}$$

Therefore, a renewal period W follows a geometric distribution with parameter P_t. The average length of a renewal period W is obtained as:

$$E[W] = \frac{E[G]}{P_t} \tag{2.7}$$

where

$$E[G] = P_t E[L] + (1 - P_t)\sigma \tag{2.8}$$

σ represents the backoff slot and $E[L]$ represents the average length of a PU transmission.

Thus, the channel utilization U is related to $E[W]$ as follows:

$$\begin{aligned} U &= \frac{E[L]}{E[W]} \\ &= \frac{P_t E[L]}{P_t E[L] + \sigma(1 - P_t)} \end{aligned} \tag{2.9}$$

Now, we obtain the number of successful packet transmissions from a tagged PU in a generic slot as follows. The probability P_{succ} that a generic time slot contains a successful packet transmission from a tagged PU is the same as the probability of having the tagged PU attempting transmission in a randomly chosen time slot and that none of the $(N-1)$ remaining PUs transmit in this particular time slot

$$P_{succ} = p\,(1-p)^{N-1} \tag{2.10}$$

Thus, a renewal period Z follows a geometric distribution with parameter P_{succ}, and the average length of the renewal period Z can be obtained as

$$E[Z] = \frac{E[G]}{P_{succ}} = \frac{E[G]}{p\,(1-p)^{N-1}} \tag{2.11}$$

Similarly, the throughput of the PU network is related to $E[Z]$ as follows:

$$\mathcal{P}_B = \frac{L_p}{E[Z]} = p\,(1-p)^{N-1}\frac{L_p}{E[G]} \tag{2.12}$$

and the aggregated throughput for this network is given by

$$\mathcal{P}_{agg} = N\,\mathcal{P}_B \tag{2.13}$$

Once we rearrange Equations 2.6 and 2.9 so as to express P_t in terms of U:

$$P_t = \frac{\sigma U}{(1-U)E[L]+U\sigma} \tag{2.14}$$

It follows that the stationary probability p of having a PU transmission attempted in a time slot can be obtained as

$$p = 1 - f(U)^{\frac{1}{N}} \tag{2.15}$$

where

$$f(U) = \frac{(1-U)E[L]}{(1-U)E[L]+U\sigma} \tag{2.16}$$

In order to maximize the throughput $\mathcal{P}_B$, it is convenient to rearrange Equations 2.8, 2.11, and 2.12 so as to obtain $E[G]$, $E[Z]$, and $\mathcal{P}_B$ in terms of U:

$$\begin{aligned} E[G] &= \frac{E[L]\sigma}{(1-U)E[L]+U\sigma} \\ &= \frac{f(U)}{1-U}\sigma \end{aligned} \tag{2.17}$$

$$\begin{aligned} E[Z] &= \sigma\,\frac{f(U)}{1-U}\left[\left(1-f(U)^{\frac{1}{N}}\right)f(U)^{\frac{N-1}{N}}\right]^{-1} \\ &= \frac{\sigma}{1-U}\left[\left(f(U)^{-\frac{1}{N}}-1\right)\right]^{-1} \end{aligned} \tag{2.18}$$

$$\mathcal{P}_B = \frac{L_p(1-U)}{\sigma}\left(f(U)^{-\frac{1}{N}}-1\right) \tag{2.19}$$

The value of U for which $\mathcal{P}_B$ is maximized, is denoted by U^* and written as:

$$U^* = \arg\max_U \mathcal{P}_{agg}(U) \tag{2.20}$$

and it can be obtained after differentiating $\mathcal{P}_{agg}$ with respect to U:

$$\begin{aligned}\mathcal{P}_{agg}(U)' &= \frac{\partial}{\partial U}\frac{N\,L_p\,(1-U)}{\sigma}\Big(f(U)^{-\frac{1}{N}}-1\Big) \\ &= \frac{N\,L_p}{\sigma}\Big(\frac{\sigma\, f(U)^{\frac{N-1}{N}}}{N(1-U)E[L]}-f(U)^{-\frac{1}{N}}+1\Big)\end{aligned} \tag{2.21}$$

so that by equating $\mathcal{P}_{agg}(U)'$ to zero, one finds the solution as:

$$f(U)-\Big(1-\frac{\sigma}{N}\frac{f(U)}{(1-U)E[L]}\Big)^{N}=0 \tag{2.22}$$

2.4 Secondary Access Scheme

In the following, we study the system behavior in a generic idle period Y. Specifically, we find the capacity of transmission strategies of SUs by developing an analytical model which depends on the access protocol used by PUs and the statistical properties of idle periods distribution. In the model, a slotted Carrier Sense Multiple Access (CSMA)-like protocol is adopted in which the time is slotted into segments of Δ seconds, whose duration correspond exactly to the transmission time of a single packet, see Figure 2.3.

To simplify the problem, we assume that the time required for spectrum sensing is negligible, which means the spectrum sensing is perfect. All packets are of constant transmission duration Δ that is assumed close to zero ($\Delta \to 0$). Every overlapping transmission between the PU system and the SU system destroys all packets that are involved in the collision. To

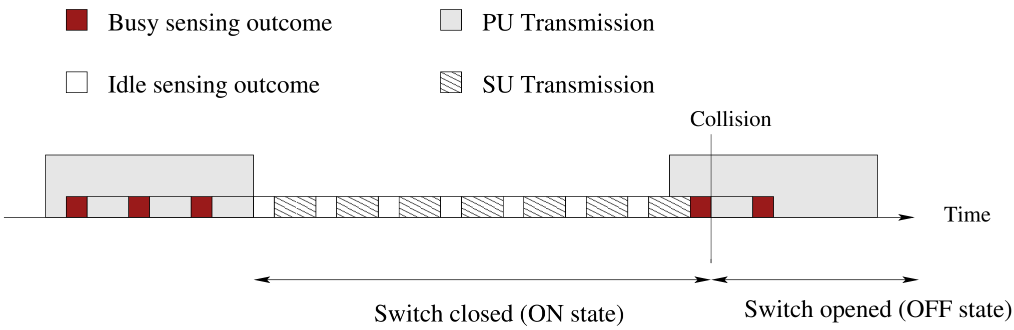

Figure 2.3: *Illustration of SU modes of operation*

transmit a SU packet, a SU sender senses the channel, and based on the channel sensing outcome it operates as follows:

(1) If the channel is sensed busy, i.e., the sensing outcome declares PU present, then it refrains its transmission until the next slot.

(2) If the channel is sensed idle, a SU sender has three options:

   (i) No transmission occurs, because a transmission strategy restrains access to the channel though the receiver is ready to receive;

   (ii) No transmission occurs, because the SU receiver is not ready to receive;

   (iii) Packet transmission takes place, because the SU receiver is ready to receive and that a transmission strategy allows access.

This is illustrated in Figure 2.3. In this case, the SU receiver readiness state always starts the beginning of idle periods, i.e, the departure of PUs. Also, we note that there is at most one PU packet interruption, which may occur at the beginning of every busy period.

We are interested in counting the number of SU packets that are successfully received during a generic idle period Y. Every time slot Δ that contains a successfully received SU packet is counted as a contribution to throughput for SUs. At the same time, every time slot Δ that contains a collision with PUs is counted as an instance of violating the protection of PUs. The time scale for a change of channel state is much longer compared to the length of a SU packet.

The use of the family of *timeout* based techniques is one way to approach the realization of such an access scheme. A timeout scheme assumes that each service is of constant length requiring X seconds of transmissions to complete satisfactory. To do so, let a generic idle period Y correspond to a renewal interval, we say that the switch is ON at time t if the renewal interval Y containing t is of length greater than or equal to X (i.e., $\min\{Y,X\}$), and we say that the switch is OFF at time t otherwise. In other words, the channel switch is always closed, if the idle period exceeds X or is always open if the idle period is less than X. Thus, the capacity of SU access protocol can be obtained from the long run proportion $\mathcal{P}_S$ of the switch ON time:

$$\mathcal{P}_S = \frac{E[\text{switch ON time}]}{E[\text{idle periods}]} \tag{2.23}$$

By exploiting Equation 2.23, the analysis of the protocol capacity for SUs, which indicates the efficiency of SU transmission strategy, can be performed by studying the optimal length of the timeout X, based on the renewal period Z and the renewal period Y. This can be obtained provided that the optimal packet loss for PUs, which specifies a threshold on PU packet loss, is known a priori.

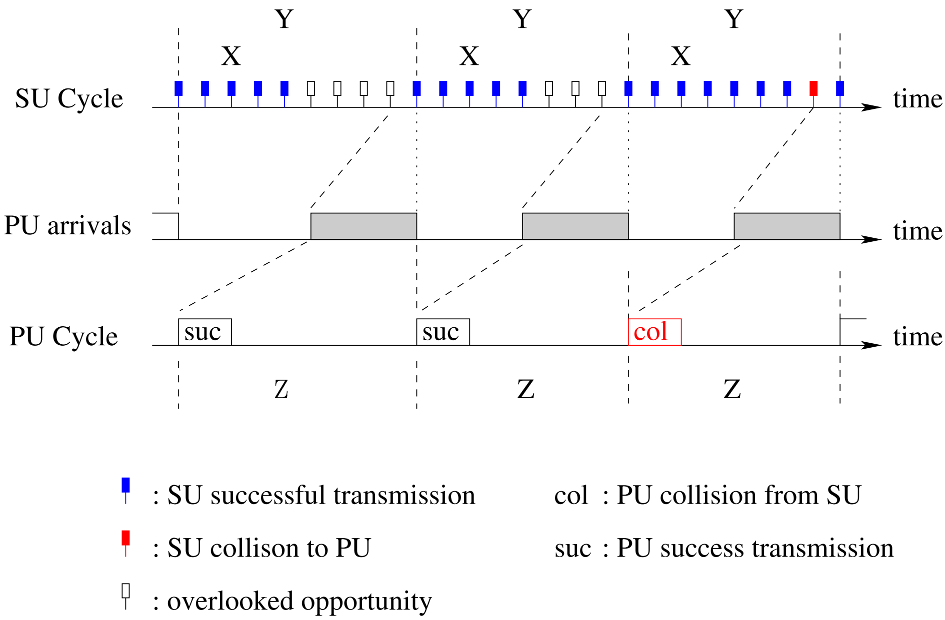

Figure 2.4: *Illustration of the analytical framework*

2.4.1 Performance Modeling Framework

In essence, a secondary access scheme aims at exploiting the idleness of PUs by using a channel that has a relatively low loss effect on PUs when a SU is transmitting. We assume that we have:

(A1) Perfect knowledge about the entire idle period trace;

(A2) Perfect knowledge about the threshold $\bar{P}_I$ on PU packet loss supported by the services provided to PUs.

Given the idle period statistics characterization of the idle period distribution $F_Y(y)$, an analytical model is developed from which the optimal X-value is obtained. This value is used in turn to characterize the capability of the access scheme to maximize throughput to SU while the packet loss tolerance for PUs protection is guaranteed satisfactory.

From the SU viewpoint, consider the time axis as marked SU-cycle (Figure 2.4), let the idle period Y correspond to a renewal, and say that the channel switch is ON at time t if the renewal Y containing t is greater than or equal to X. In other words, whenever a renewal occurs, we say that SU transmissions complete satisfactory after the channel switch has been closed for the first X seconds of the renewal. In this case, any arriving PU experiences successful packet arrival and a SU has all its packets successfully received.

On the other hand, we say that the channel switch is OFF at time t whenever the renewal Y at time t is less than X. In other words, whenever $(Y < X)$ the channel switch ON time is always zero. In this case, the arriving PU packet occurs before X, in which case a collision to PUs happens. Consequently, the SU has no successful packet reception and the arriving PU experiences packet drop-out.

Therefore, the throughput to SUs is given by the long run proportion $\mathcal{P}_S$ of switch ON time:

$$\begin{aligned}
\mathcal{P}_S &= \frac{E[\text{switch ON time}]}{E[\text{idle periods of PU in idle state}]} \\
&= \frac{E[\min(Y,X)]}{E[Y]} \\
&= \frac{1}{E[Y]} \int_0^\infty P(\min(Y,X) > t)dt \\
&= \frac{1}{E[Y]} \int_0^X \left(1 - F_Y(t)\right)dt
\end{aligned} \tag{2.24}$$

From the PU viewpoint, consider the time axis marked as PU-cycle (Figure 2.4), let a generic busy period of PUs correspond to a renewal Z, which is defined by the sequence of moments corresponding to successful packet transmission from a tagged PU. In other words, a renewal Z contains exactly one transmission from a tagged PU with probability $\mathcal{P}_B$. Let $\{Y_1, \cdots, Y_m\}$ be a set of random variables, with the same distribution as Y, and specifying the arrival times associated with a successful packet of a tagged PU in the renewal Z. If $Y_1 > X$, an arriving packet of a tagged PU at time 1 has no collision with SU, and hence it successfully completes transmission, and so on. Otherwise $Y_m < X$ the arriving packet of a tagged PU at time m is dropped out upon arrival because of SU transmissions. Then:

$$M = \inf \{m \mid Y_m < X\} \tag{2.25}$$

That is, M is associated with a renewal that represents the sequence of moments corresponding to a packet loss experienced at a tagged PU, because a SU is transmitting. We note that the probability distribution of

Y, $F_Y(t) = P(Y \leq t)$, is equivalent to the probability q that a SU packet is transmitting in X:

$$P(Y < t) = q \tag{2.26}$$

Therefore, the packet loss experienced by a tagged PU when a SU is transmitting can be obtain from:

$$\begin{aligned} \mathcal{P}_I &= \sum_{m=1}^{\infty} P(\text{m-th arrival dropped}) \Big(P(\text{tagged PU in } Z)\Big)^m \\ &= \sum_{m=1}^{\infty} P\big(M(t) = m\big)\, \mathcal{P}_B^m \\ &= \sum_{m=1}^{\infty} q(1-q)^{m-1}\, \mathcal{P}_B^m \\ &= \frac{q\,\mathcal{P}_B}{1-(1-q)\,\mathcal{P}_B} \end{aligned} \tag{2.27}$$

Hence, the timeout-based scheme for a secondary access of spectrum is designed in such a way that:

(C1) it holds because of prior knowledge about idle periods distribution:

$$F_Y(X) = q \tag{2.28}$$

(C2) it holds because of priori knowledge about packet loss threshold $\bar{P}_I$ imposed by PUs

$$\mathcal{P}_I \leq \bar{P}_I \tag{2.29}$$

2.4.2 Robustness to Distribution Uncertainty

In selecting a distribution for Y to characterize the idle period statistics, it is appropriated to require that the properties of this particular distribution are consistent with the ones present in the operational environment of a SU deployment.

We consider distributions that have well-known probability density functions for describing the idle periods of the PU system in idle state. Hence, the characteristics and properties of these distributions have significant impact on the performance of timeout based strategies, and furthermore affect the implementation of secondary access scheme.

Conditional Idle Probabilities

Given the idle period at some time t, the probability that the idle period of the PU does not last an additional time x, i.e., the idle period terminates before the time $(t+x)$ given that it has been idle during the interval $(0,t)$, can be written as [56]:

$$\begin{aligned} G_Y(X|t) &= \frac{P(t < Y < t+X)}{P(Y > t)} \\ &= \frac{F_Y(t+X) - F_Y(t))}{1 - F_Y(t)} \end{aligned} \tag{2.30}$$

From the PU system viewpoint, this is the conditional probability that a PU returns during the interval $(t+x)$, given that it has been absent until time t. Similarly, from the SU system viewpoint, this is the conditional probability that a SU transmission collides that of a PU.

Let $R_Y(x|t)$ denote the probability that the idle period lasts an additional time x, given that it has been idle until t, i.e, the idle period survives an additional time $(t+x)$ given that it has survived during the interval $(0,t)$. We have [56]:

$$R_Y(X|t) = \frac{1-F(t+X)}{1-F(t)} \tag{2.31}$$

This represents the conditional probability that the idle period remains idle during an additional time $(t+x)$.

Intensity Function

A function denoted $h(t)$, whose type and shape is useful to determine the distribution of idle periods, is called intensity function. It is used to describe the relationship between the distribution under consideration and time. This is given as follows [56]:

$$\begin{aligned} h(t) &= \lim_{x\to 0} \frac{1}{x} \frac{F_Y(t+x)-F_Y(t))}{1-F_Y(t)} \\ &= \lim_{x\to 0} \frac{1}{x} \Big[1-R_Y(x|t)\Big] \\ &= \frac{f(t)}{1-F(t)} \end{aligned} \tag{2.32}$$

Figure 2.5 shows the connection between the conditional probability $G_Y(X|t)$ for interrupting a PU, and the conditional probability $R_Y(X|t)$ for SU access opportunity using the intensity function $h(t)$. One can observe three relationships, namely A, B and C. A is constant, B is monotonically increasing and C is monotonically decreasing.

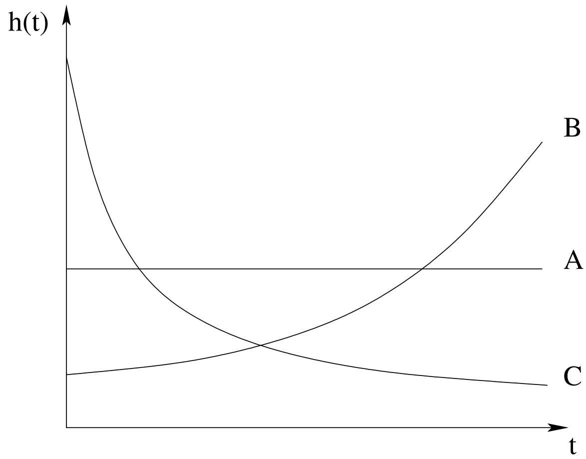

Figure 2.5: *Illustration of the intensity function $h(t)$*

Relationship B means that the process generating idle periods is subject to *aging*. In other words, the conditional probability $R_Y(x|t)$ is a decreasing function of time, i.e., $R_Y(x|t)$ is decreasing in $0 \leq t < \infty$ for all $x \geq 0$. Consequently, $h(t)$ is an increasing function of time. From the PU system viewpoint, this means that as the idle periods age it is highly likely that a PU returns. This scenario is marked B in Figure 2.5.

Relationship A means that the process generating idle periods is subject to *randomization*. This means that, conditional on the time t, the probability $R_Y(x|t)$ over an additional time period x is the same regardless of the present age t. This is marked with A in Figure 2.5. Hence, $h(t)$ is constant in $0 \leq t < \infty$ for all $x \geq 0$.

Relationship C means that the process generating idle periods is subject to *excess*. This means that, conditional on the time t, the probability $R_Y(x|t)$ over an additional time period x increases in $0 \leq t < \infty$ for all $x > 0$.

As a result, $h(t)$ decreases and so does the probability of interruptions (Figure 2.5).

Therefore, a timeout based strategy that has the distribution of the idle periods as a parameter model should also be sensitive to these properties of monotone decreasing, monotone increasing and constant. Specifically, the consideration of the distribution for the idle period is taken from the viewpoint of the intensity function $h(t)$. Hereafter, the intensity function $h(t)$ is viewed as a sufficient statistic for describing the idle periods for the entire range of time.

2.5 Adaptive Control Techniques

This section is dedicated to the study of practical schemes for controlling SU access with respect to their sensitivity to monotone increasing, monotone decreasing and constant properties of the idle periods. These includes uniform distribution, generalized Pareto distribution, and exponential distribution. Based on the assumption that a spectrum manager has complete a priori knowledge about the entire channel utilization trace, the availability of such schemes enables the actual implementation of an ideal secondary access protocol.

2.5.1 Exponential based Timeouts

From the viewpoint of constant intensity function of the idle periods $h(t)$, the exponential distribution can be regarded as the only distribution with constant intensity function [57]. Typically, the exponential distribution of idle periods arises naturally, when the arrival process of a PU flow in the channel can be assimilated to a Poisson process [57]. The constant intensity function $h(t)$ means that the conditional probability of a new PU arrival in a short period $(t+dt)$, taken as $h(t)dt$, is the same regardless of the present age t.

Let $E[Y]$ denote the parameter of the idle period exponentials with the following properties:

$$f_Y(t) = \frac{1}{E[Y]}\exp\Big(-\frac{t}{E[Y]}\Big), \quad F_Y(t) = 1-\exp\big(-\frac{t}{E[Y]}\big), \quad E[Y] \geq 0 \quad (2.33)$$

$$\begin{aligned} h(t) &= \frac{f(t)}{1-F(t)} \\ &= E[Y] \end{aligned} \quad (2.34)$$

The secondary access in this case is such that as soon as the idle period starts, a SU is immediately put into ON state by closing its switch. After a timeout of X_{EX} seconds, it can be transitioned into OFF state to wait until the beginning of detected idle period. The algorithm described above is then repeated. The timeout in this case is obtained as follows:

$$\begin{aligned} F_Y(X_{EX}) &= \int_0^{X_{EX}} \frac{1}{E[Y]}\exp\big(-\frac{t}{E[Y]}\big)dt \\ &= q \\ X_{EX} &= -E[Y]\ln(1-q) \end{aligned} \quad (2.35)$$

The capacity $\mathcal{P}_S$ of the access scheme using this algorithm is therefore given by:

$$\mathcal{P}_S \equiv F_Y(X_{EX}) = q \tag{2.36}$$

2.5.2 Uniform based Timeouts

From the viewpoint of increasing intensity function $h(t)$, the conditional probability of a new PU arrival in a short period $(t+dt)$, taken as $h(t)dt$, is an increasing function of time t, and therefore aging is forced. The uniform distribution is one example in the class of distributions that have this type of intensity function [57].

Let the interval $[0,a]$ be the parameter characteristic of the idle periods with uniform distribution. We have the following properties:

$$f_Y(t) = \begin{cases} \frac{1}{a} & \text{if} \quad 0 \leq t \leq a & \text{(2.37a)} \\ 0 & \text{otherwise} & \text{(2.37b)} \end{cases}$$

$$F_Y(t) = \begin{cases} 0 & \text{if} \quad t < 0 & \text{(2.38a)} \\ \frac{t}{a} & \text{if} \quad 0 \leq t \leq a & \text{(2.38b)} \\ 0 & \text{if} \quad t > 0 & \text{(2.38c)} \end{cases}$$

$$E[Y] = \frac{a}{2} \tag{2.39}$$

$$h(t) = \frac{f(t)}{1-F(t)} \tag{2.40}$$

$$= \frac{1}{a-t}$$

The intensity function is therefore an increasing function of time.

The secondary access in this case is such that, as soon as the idle period starts, immediately a SU is put in ON state by closing the channel switch for X_{UN} seconds. When the timeout expires, it can be transitioned into the OFF state by opening the switch to wait until the beginning of idle period is detected, and the algorithm described above is repeated. The timeout in this case is obtained as follows:

$$\begin{aligned} P(Y < X_{UN}) &= q \\ \int_0^{X_{UN}} f_Y(t)dt &= q \\ X_{UN} &= q\,a \end{aligned} \tag{2.41}$$

The capacity $\mathcal{P}_S$ of the access scheme when using this algorithm has the following expression:

$$\begin{aligned} \mathcal{P}_S &= \frac{E\big[\min(X_{UN}, Y)\big]}{E[Y]} \\ &= \frac{2}{a}\int_0^{\infty} P\big(\min(X_{UN}, Y) > t\big)dt \\ &= \frac{2}{a}\int_0^{X_{UN}} P(Y > t)dt \\ &= \frac{2}{a}\int_0^{X_{UN}} \big(1 - F_Y(t)\big)dt \\ &= \frac{2\,X_{UN}}{a} - \Big(\frac{X_{UN}}{a}\Big)^2 \\ &= q\,(2-q) \end{aligned} \tag{2.42}$$

2.5.3 Generalized Pareto based Timeouts

From the viewpoint of the decreasing intensity function $h(t)$, the conditional probability of a new PU arrival in a short period $(t+dt)$, taken as $h(t)dt$, is a decreasing function of time t, and therefore excess is in force. The idle periods satisfying such properly have the so-called heavy tailed distribution property [58]. The generalized Pareto distribution is one example in the class of heavy-tailed distributions.

Let the shape index $k \neq 0$ and the scale factor σ be the parameters characteristic of the idle periods with generalized Pareto distribution. We have the following properties:

$$F_Y(t,k,\sigma) = 1-\left(1+\frac{kt}{\sigma}\right)^{-\frac{1}{k}}; \qquad k \neq 0, \quad \sigma \geq 0 \tag{2.43}$$

$$f_Y(t,k,\sigma) = \frac{1}{\sigma}\left(1+\frac{kt}{\sigma}\right)^{-1-\frac{1}{k}}; \qquad k \neq 0, \quad \sigma \geq 0 \tag{2.44}$$

$$E[Y] = \frac{\sigma}{1-k}; \qquad 0 \leq k < 1 \tag{2.45}$$

$$h(t) = \frac{\sigma^2}{\sigma+kt} \tag{2.46}$$

The secondary access in this case is such that as soon as the idle period starts, immediately a SU is put in OFF state by opening its switch for X_{GN} seconds. When the timeout expires, it can be transitioned into the ON state until an interruption to PU is detected on the channel. The timeout in this case is obtained as follows:

$$\begin{aligned} P(Y < X_{GEN}) &= 1-q \\ \int_0^{X_{wafe}} f_Y(t)dt &= 1-q \\ X_{GEN} &= \frac{\sigma}{k}\left(q^{-k}-1\right) \end{aligned} \tag{2.47}$$

The performance characteristic $\mathcal{P}_S$ of this scheme is such that:

$$\begin{aligned} \mathcal{P}_S &= 1-\frac{E\left[\min(X_{GEN},Y)\right]}{E[Y]} \\ &= 1-\frac{1}{E[Y]}\int_0^{\infty} P\big(\min(X_{GEN},Y)>t\big)dt \\ &= 1-\frac{1}{E[Y]}\int_0^{X_{GEN}} \big(1-F_Y(t)\big)dt \\ &= q^{1-k} \end{aligned} \tag{2.48}$$

2.6 Robust Control Techniques

In this section, we study robust control techniques with reference to the uncertainty introduced by considering the fact that in most real world scenarios there is little knowledge about the distribution i.e., either unknown distribution or non-stationary. Hence, it is highly desirable to implement timeout controls that are insensitive to precise distribution information about the idle periods, also referred to as robust timeout controls. The availability of such schemes would enable the implementation of secondary access protocol capable of online adaptation.

2.6.1 Problem Description

We denote by ε a near future event that we want to predict, i.e., $\varepsilon = \{Y \leq X_{pred}\}$. Also, we denote by υ the past event whose occurrence is used to make prediction about the near future event ε. By taking into account uncertain predictions, the conditional probability $P(\varepsilon|\upsilon)$ represents the uncertain probability $\mathcal{P}_S$ after a history of past observations denoted the event υ.

We call *ideal* predictor the one with complete a priori knowledge about the entire utilization trace:

$$\mathcal{P}_S = q \tag{2.49}$$

The predictor of practical interest is defined by comparing its performance to that of the ideal predictor so that a zero coverage error condition holds:

$$\mathcal{P}_S(\upsilon|\varepsilon) = q + \text{ zero coverage error} \tag{2.50}$$

This is hereafter referred to as a predictor with zero coverage error. We pose the problem of finding such a predictor whose availability would enable the implementation of zero coverage error timeouts for secondary access management.

2.6.2 The Family of Exponential based Techniques

By taking into consideration the memoryless property of the exponential distribution, the case of having limited information about idle periods distribution of the channel occupancy by PU can be approximated to the family of exponential distributions with unknown parameter λ.

Let $\{Y_1,...,Y_n\}$ denote a random sample of size n from an exponential population with unknown parameter λ and density $f_{Y_i}(t)$. We use subscripts to indicate that the past observed idle periods occur in sequence, where n indicates the latest observed idle period. Let λ and n be respectively the scale parameter and the shape parameter characteristic of the gamma distribution [57]. It follows that a period V from the onset of the history of idle periods to the nth occurrence is Gamma distributed with density $f_V(t)$ and parameter n and λ. That means, we have:

$$V = \sum_0^n Y_i, \quad f_{Y_i}(t) = \frac{1}{\lambda}\exp(-\frac{t}{\lambda}), \quad f_V(t) = \frac{t^{n-1}}{\Gamma(n)\,\lambda^{n-1}}\exp(-\frac{t}{\lambda}), \quad t \geq 0 \tag{2.51}$$

According to our notation, the observed event υ can be written as:

$$\upsilon = \{Y_n, Y_{n-1}, ..., Y_1\} \tag{2.52}$$

so that the predicted timeout X_{pred} can be obtained from the past history by a regression equation:

$$X_{pred} = r_n(q)\sum_0^n Y_i \tag{2.53}$$

Given a new exponential idle period Y, the near future event ε that we want to predict can be written as:

$$\varepsilon = \left\{Y \leq r_n(q)\sum_0^n Y_i\right\} \tag{2.54}$$

Therefore the probability $\mathcal{P}_S(\varepsilon|\upsilon)$ can be written as:

$$\begin{aligned}
\mathcal{P}_S(\varepsilon|\upsilon) &= P\Big(Y \leq r_n(q)\,V\Big) \qquad (2.55)\\
&= 1-\int_o^{\infty}(1-F_{Y_i})\,f_V(t)\,dt\\
&= 1-\int_o^{\infty}\frac{t^{n-1}}{\Gamma(n)\,\lambda^n}\exp\Big(-\frac{r_n(q)\,t}{\lambda}\Big)\exp(-\frac{t}{\lambda})\\
&= 1-\int_o^{\infty}\frac{t^{n-1}}{\Gamma(n)\,\lambda^n}\exp\Big(-\big(1+r_n(q)\big)\,\frac{t}{\lambda}\Big)\\
&= 1-\big(1+r_n(q)\big)^{-n}\\
&\quad\times\int_0^{\infty}\frac{t^{n-1}\,\big(1+r_n(q)\big)^n}{\Gamma(n)\,\lambda^n}\exp\Big(-\big(1+r_n(q)\big)\,\frac{t}{\lambda}\Big)\\
&= 1-\Big(1+r_n(q)\Big)^{-n}
\end{aligned}$$

The performance characteristic can be obtained by using the equivalence relation Equation 2.50 such that a timeout that takes on a value

$$X_{pred} = \Big((1-q)^{-\frac{1}{n}}-1\Big)\sum_0^n Y_i \qquad (2.56)$$

also satisfies a zero coverage error prediction, and hereafter it is referred to as zero coverage error timeout.

2.7 Numerical Results

In this section, we evaluate the channel capacity $\mathcal{P}_S$ for the studied timeout-based schemes of secondary access. We start with the experimental setup from which we derive the system parameters used to evaluate the secondary access. Following, numerical results are reported to examine the analytical capacity of the studied schemes.

2.7.1 Experimental Network Setup

We evaluate the analytical throughput $\mathcal{P}_B$ for a 11 Mbps **WLAN!** (**WLAN!**) cell shared by N PU hosts, operating at heavy utilization and congestion state.

If we neglect the propagation times, Table 2.1 shows the overall transmission time of a single packet being sent by a single PU host over a WLAN cell of size $N = 1$. The selection of the parameter values was made according

Table 2.1: *Setting of a WLAN Packet,* $r = 1, 2, 5.5$ *;* $R = 11$ *Mbps*

SIFS	10 μs
DIFS	50 μs
σ-slot	20 μs
PHY header(short)	96 μs
PHY header(long)	192 μs
MAC header	272 bits
ACK	112 bits
packet payload	12000 bits
overhead_r	10 + 50 + 2 × 192 + (272 + 112)/r
overhead_R	10 + 50 + 2 × 96 + (272 + 112)/R
packet time (L_r)	overhead_r + payload/r
packet time (L_R)	overhead_R + payload/R

to the IEEE 802.11b MAC protocol for the CSMA/CA base access model. The possible data rates for IEEE 802.11b products are: 1, 2, 5.5 and 11 Mbps. For the load configuration L, a pair $(L_R : L_r)$ is taken to denote a class of two WLAN packet sizes constructed as follows: (i) the lowest feasible data rate r varies in the range 1 , 2, 5.5, 11 Mbps, and (ii) the highest feasible data rate R is fixed to 11 Mbps. Hence, the possible packet size classes are (11:11), (11:5.5), (11:2), and (11:1). Similarly, a pair $(N_f - N_s)$ is taken to denote a PU network of size N, made of $(N-1)$ fast PU hosts transmitting at R and one slow PU host transmitting at r.

Assuming a packet payload of size 12000 bits being transmitted, the actual busy time $E[Z]$, the analytical throughput $\mathcal{P}_B$ along with the cumulative throughput $N \times \mathcal{P}_B$ are reported as the estimate of the MAC protocol capacity in Tables 2.2, 2.3, 2.4, and 2.5, respectively. This is for each packet size class $(L_R : L_r)$ and the PU network size ranging from $N = 1$ to $N = 10$. Also, the channel utilization U and the minimum p-value are reported as protocol parameters.

It is observed that all PU hosts have the same throughput $\mathcal{P}_B$ that decreases with the network size N. Also, it is observed that the PU throughput $\mathcal{P}_B$ is heavily affected by the presence of at least one slow PU host that captures the channel for long times because its bit rate is low. Yet, the network size ranked by majority is made of PU hosts using the highest feasible data rate. Such behavior is inherent of the CSMA protocol used by PUs and thus inherently affects the design of secondary access schemes.

Table 2.2: *Protocol with* $(L_R : L_r) = (11 : 11)$ *Mbps*

Network size		Protocol parameter			Protocol capacity Mbps	
N	$N_f - N_s$	p_{min}	U	$E[Z]$ ms	$\mathcal{P}_B$	$N \times \mathcal{P}_B$
1	0−1	1.0000	1.0000	1.377	8.71	8.71
2	1−1	0.1075	0.9462	3.492	3.89	7.77
3	2−1	0.0637	0.9377	4.715	2.55	7.64
4	3−1	0.0455	0.9338	6.337	1.89	7.57
5	4−1	0.0354	0.9316	7.958	1.51	7.54
6	5−1	0.0290	0.9302	9.578	1.25	7.52
7	6−1	0.0246	0.9292	11.198	1.07	7.50
8	7−1	0.0213	0.9284	12.818	0.94	7.49
9	8−1	0.0188	0.9279	14.438	0.83	7.48
10	9−1	0.0169	0.9274	16.057	0.75	7.47

Table 2.3: *Protocol with* $(L_R : L_r) = (11 : 5.5)$ *Mbps*

Network size		Protocol parameter		Protocol capacity Mbps		
N	$N_f - N_s$	p_{min}	U	$E[Z]$ ms	$\mathcal{P}_B$	$N \times \mathcal{P}_B$
1	0−1	1.0000	1.0000	2.695	4.45	4.45
2	1−1	0.0902	0.9549	4.477	2.68	5.36
3	2−1	0.0561	0.9450	6.118	1.96	5.88
4	3−1	0.0412	0.9399	7.747	1.55	6.20
5	4−1	0.0327	0.9368	9.372	1.28	6.40
6	5−1	0.0271	0.9347	10.995	1.09	6.55
7	6−1	0.0232	0.9331	12.617	0.95	6.66
8	7−1	0.0202	0.9320	14.238	0.84	6.74
9	8−1	0.0180	0.9311	15.859	0.76	6.81
10	9−1	0.0162	0.9303	17.479	0.69	6.87

Table 2.4: *Protocol with* $(L_R : L_r) = (11 : 2)$ *Mbps*

Network size		Protocol parameter			Protocol capacity Mbps	
N	$N_f - N_s$	p_{min}	U	$E[Z]$ ms	$\mathcal{P}_B$	$N \times \mathcal{P}_B$
1	0−1	1.0	1.0	6.636	1.81	1.81
2	1−1	0.066	0.967	8.579	1.4	2.8
3	2−1	0.0435	0.9571	10.265	1.17	3.51
4	3−1	0.0333	0.9512	11.919	1.01	4.03
5	4−1	0.0272	0.9472	13.561	0.88	4.42
6	5−1	0.023	0.9442	15.196	0.79	4.74
7	6−1	0.02	0.9419	16.827	0.71	4.99
8	7−1	0.0178	0.94	18.455	0.65	5.2
9	8−1	0.0159	0.9385	20.081	0.6	5.38
10	9−1	0.0145	0.9373	21.706	0.55	5.53

Table 2.5: *Protocol with* $(L_R : L_r) = (11 : 1)$ *Mbps*

Network size		Protocol parameter		Protocol capacity Mbps		
N	$N_f - N_s$	p_{min}	U	$E[Z]$ ms	$\mathcal{P}_B$	$N \times \mathcal{P}_B$
1	0−1	1.0	1.0	12.828	0.94	0.94
2	1−1	0.0504	0.9748	14.959	0.8	1.6
3	2−1	0.0342	0.9662	16.706	0.72	2.15
4	3−1	0.0268	0.9605	18.401	0.65	2.61
5	4−1	0.0223	0.9563	20.072	0.6	2.99
6	5−1	0.0192	0.9531	21.729	0.55	3.31
7	6−1	0.017	0.9505	23.377	0.51	3.59
8	7−1	0.0152	0.9483	25.020	0.48	3.84
9	8−1	0.0138	0.9465	26.659	0.45	4.05
10	9−1	0.0127	0.9449	28.294	0.42	4.24

Table 2.6: *q-values at different packet size classes*

	q-values vs normalized $\mathcal{P}_B$ for various $(L_R : L_r)$ Mbps; $\bar{P}_I = 0.042$							
	(11:11)		(11:5.5)		(11:2)		(11:1)	
N	$\mathcal{P}_B$	q	$\mathcal{P}_B$	q	$\mathcal{P}_B$	q	$\mathcal{P}_B$	q
1	0.792	0.011	0.809	0.01	0.904	0.0040	0.935	0.0030
2	0.353	0.076	0.325	0.087	0.215	0.152	0.134	0.27
3	0.231	0.138	0.214	0.153	0.146	0.243	0.094	0.403
4	0.172	0.2	0.161	0.281	0.115	0.32	0.077	0.501
5	0.137	0.262	0.129	0.343	0.096	0.392	0.066	0.586
6	0.114	0.324	0.108	0.406	0.083	0.46	0.059	0.663
7	0.097	0.386	0.093	0.468	0.073	0.526	0.054	0.735
8	0.085	0.448	0.082	0.468	0.066	0.591	0.049	0.805
9	0.076	0.51	0.073	0.53	0.06	0.656	0.046	0.874
10	0.068	0.572	0.066	0.593	0.055	0.72	0.042	0.941

Table 2.7: *X-values generating the timeouts ms*

X-values for various $(L_R : L_r)$ Mbps											
(11:11)			(11:5.5)			(11:2)			(11:1)		
EX	UN	GE	EX	UN	GE	EX	UN	GE	EX	UN	GE
0.4	0.9	305	0.4	0.8	321	0.2	0.4	464	0.1	0.2	563
3.2	6.1	108	3.6	6.9	99.7	6.6	12.2	67.5	12.6	21.6	41.3
6.0	11.1	72.3	6.6	12.2	67.1	11.2	19.5	45.6	20.6	32.2	26.3
8.9	16.0	54.1	9.8	17.4	50.5	15.5	25.6	34.6	27.8	40.1	19.1
12.2	21.0	42.5	13.2	22.4	39.8	19.9	31.3	27.3	35.2	46.8	14.3
19.5	25.9	34.2	16.8	27.5	32.0	24.6	36.8	21.9	43.5	53.0	10.7
23.8	30.9	27.8	20.8	32.5	26.1	29.9	42.1	17.6	53.2	58.8	7.9
23.8	35.8	22.7	25.3	37.5	21.3	35.8	47.3	14.0	65.5	64.4	5.4
28.5	40.8	18.6	30.2	42.4	17.3	42.6	52.4	11.0	82.8	69.9	3.3
33.9	45.7	15.0	35.9	47.4	14.0	50.9	57.6	8.4	113	75.3	1.5

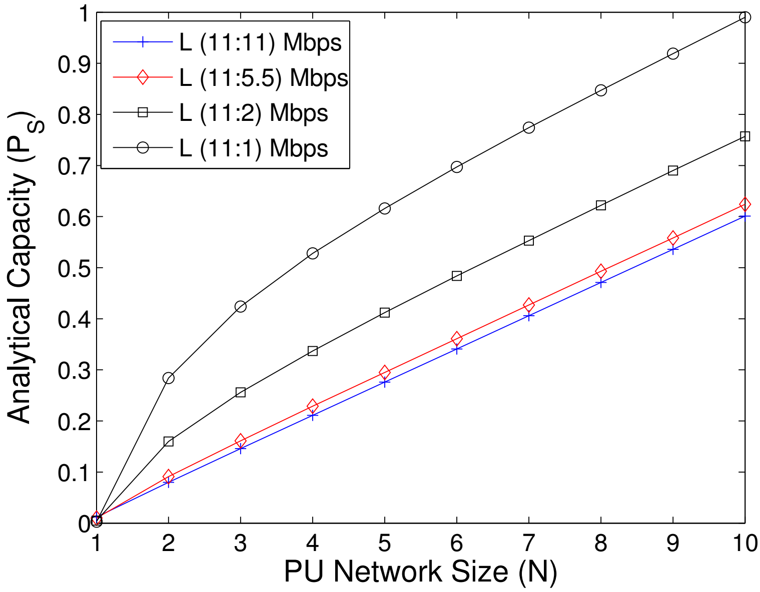

Figure 2.6: *Performance of exponential based timeouts*

2.7.2 The Impact of Idle Period Distributions

In the following, we examine the effect of the idle period distributions using the analytical capacity $\mathcal{P}_S$ for the studied timeout-based schemes controlling the secondary access. As reference distributions for idle period, we consider three distributions with the same mean $E[Y] = 30$ ms: the exponential distribution with parameter $E[Y]$, the uniform distribution over the interval $\left[0, 2 \times E[Y]\right]$, and the generalized Pareto distribution with param-

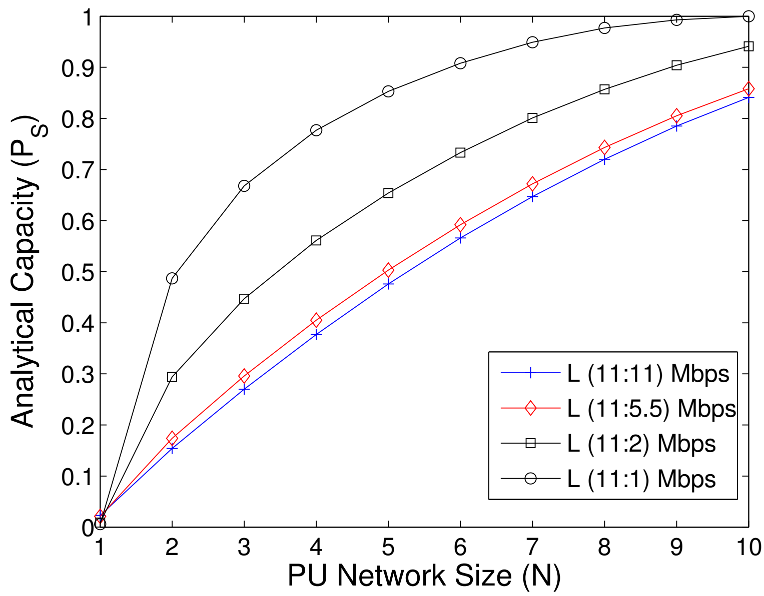

Figure 2.7: *Performance of uniform based timeouts*

eter $k = 0.01$ and $\sigma = 0.4$. In all cases studied, the threshold on the packet loss for the service provided by PUs is set to 0.042 packet/ms. The selection of these parameters are not crucial, they can be thought of as representative of a general behavior of the network.

To evaluate the analytical capacity $\mathcal{P}_S$, we start with the characterization of the idle distribution of Y at X, $F_Y(X)$ (Equation 2.28), computed using the q-value solving for $\mathcal{P}_I = \bar{P}_I$ (Equation 2.29). For each packet size class (L_R, L_r), together with the normalized values of $\mathcal{P}_B$, the q-values that are used for the evaluation of the analytical capacity $\mathcal{P}_S$ are reported in Table 2.6 by varying the network size N. The X-values generating the

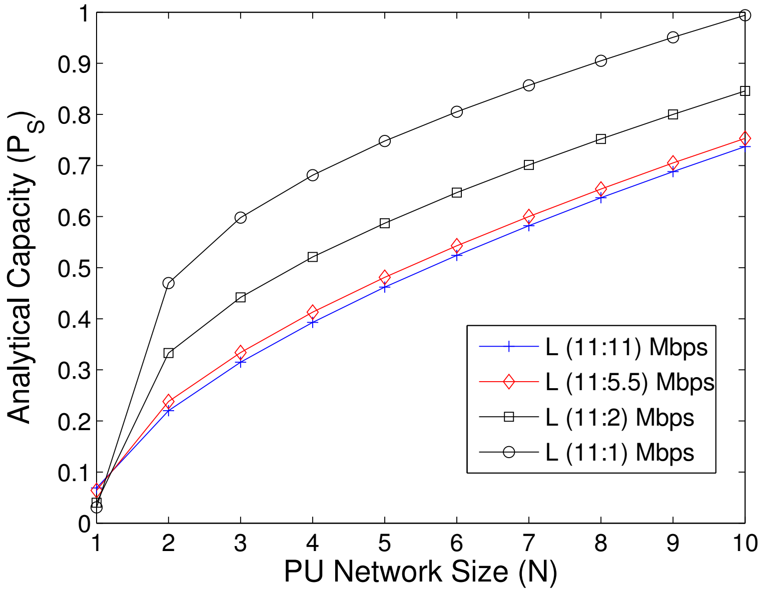

Figure 2.8: *Performance of generalized Pareto based timeouts*

timeouts for each idle period distribution are reported in Table 2.7 under the notations marked as EX−exponential, UN−uniform, GE−generalized Pareto.

Figure 2.6 shows the analytical capacity $\mathcal{P}_S$ for the exponential-based timeouts. According to Table 2.6, a maximum level of 0.042 laid on the PU packet loss $\mathcal{P}_I$, when SUs are transmitting, corresponds to the maximum level within which interference caused in the PU network by SUs can be tolerable. The effect perceived by PU hosts in this situation is that the loss in $\mathcal{P}_I$ is a lot less when all PUs have perfect transmission conditions than when there is at least one PU experiencing a poor transmission quality. In all cases the loss in $\mathcal{P}_I$ increases gradually with N.

On the other hand, Figure 2.6 represents the tradeoffs between maximum interference to a PU and maximum throughput obtained by SU. As loss tolerance of a PU increases with the number of contending hosts N, the effect is that throughput $\mathcal{P}_S$ for SUs increases with N. Similarly, as loss tolerance of a PU increases with the presence of a least one PU host experiencing bad channel conditions (i.e., slow PU), the effect is throughput increase in $\mathcal{P}_S$ for SUs as the data rate used by the slow PU decreases to lower values in response to the experience of bad channel conditions. For example, it is observed that the better performance in $\mathcal{P}_S$ has the consequence that the interruptions effect is greatest to the PU host transmitting with a lower rate, but this particular host has a maximum interruption probability that is tolerable.

Furthermore, Figure 2.7 shows the analytical capacity $\mathcal{P}_S$ for the uniform-based timeouts. It is observed that the extension to uniform idle period distribution gives a better performance in $\mathcal{P}_S$ as compared to that of the exponential based timeouts.

Figure 2.8 shows the analytical capacity $\mathcal{P}_S$ for the generalized Pareto based timeouts.ls It is observed that the extension to generalized Pareto idle period distribution also gives better performance in P_S as compared to that of the exponential based timeouts.

2.7.3 The Impact of Zero Coverage Error Timeouts

We examine the effect of zero coverage error timeouts for controlling secondary accesses on the channel by using a simulation example. In the simulation, we generate n consecutive idle periods. The length of each idle period is continuous and follows the exponential distribution with unknown

Table 2.8: *Zero coverage error timeouts*

	Sample	Zero Coverage Errors		
q-value	n	$\lambda = 10$	$\lambda = 20$	$\lambda = 30$
$q = 0.403$	50	-0.0109	-0.0110	-0.0108
	100	-0.0055	-0.0055	-0.0054
	250	-0.0022	-0.0022	-0.0022
	500	-0.0013	-0.0011	-0.0011
$q = 0.63$	50	-0.0124	-0.0124	-0.0124
	100	-0.0062	-0.0063	-0.0062
	250	-0.0025	-0.0025	-0.0025
	500	-0.0012	-0.0013	-0.0013
$q = 0.94$	50	-0.0124	-0.0124	-0.0123
	100	-0.0063	-0.0063	-0.0063
	250	-0.0025	-0.0025	-0.0025
	500	-0.0013	-0.0013	-0.0013

parameter λ. Taking the sample size as $n = 10, 50, 100, 200, 500$ and the q-values as $0.403, 0.63, 0.94$, table 2.8 shows the approximate zero coverage errors for different values of λ. In the table, each point is obtained by computing X_{pred} (Equation 2.56) and using the difference $\mathcal{P}_S(\varepsilon|\upsilon) - q = 0$ over 10000 trails.

It is observed that the zero coverage error timeouts are insensitive to the choice of the parameter λ. It is also observed that the accuracy of the coverage errors increases roughly by n orders of magnitude with the increase in sample size.

2.8 Summary

Efficient transmission strategies for a single SU that uses a listen-before-talk protocol as access scheme in the presence of distributed, dynamic channels occupied by primary user PUs has been studied. A statistical characterization of cognitive channel model has been developed, and the statistical properties in terms of PUs traffic are reported. The timeout-based techniques have been studied in two main classes. These are: (i) where the strategies are developed based on perfect knowledge about functional form of the distribution function for the random variable idle period, and (ii) where the strategy is developed based on limited information about the functional form of the distribution function for the random variable idle period. We have demonstrated the use of our technique on numerical examples with nominal system parameters.

Chapter 3

Spectrum Opportunity Exploitation

Abstract

In this chapter, the second DSA model considered in the dissertation is presented. Efficient transmission strategies for a single SU that uses a slotted protocol as access scheme are studied in the presence of parallel, dynamic channels already occupied by PUs traffic. In every time slot, a channel is selected and the probability of transmission in the selected channel is estimated by using local information about the quality of these channels. A SU transmits in a probabilistic manner in the selected channel based on this estimate. A Markovian decision model that embeds such a problem in its control actions is formulated. Optimal decision rules are found. The protocol capability in terms of throughput for SUs and loss at PUs is expressed as reward and cost tradeoffs of the resulting Markovian decision processes. Numerical results on the performance of the suggested access protocols are shown for nominal system parameters. An extension to

multiple access schemes is include to study multi-user efficient transmission strategies.

3.1 Introduction

3.1.1 Overview

The fundamental objective of spectrum management is to maximize the use of spectrum [59]. This has generally been taken to be the use of the spectrum by avoiding any incidents of interference [60]. Each station is equipped with a radio unit and a control unit. A procedure is invoked to perform the packet switching functions so as to establish the communication between geographically separated hosts or users. This procedure is known as MAC protocol [61]. Recent measurement studies on the actual spectrum utilization have revealed that this approach uses much more spectrum than actually needed [62].

Cognitive radio technology offers support to new users with access to the existing crowded spectrum but underutilized in practice [63]. In particular, the use of CR supported by the DSA policy has been considered in the context of secondary sharing of the spectrum by SUs with PUs entitled to protection [64]. The objective is that the sharing is transparent, i.e., without the constraints of common protocols, or any degree of information sharing, or even the awareness of other users. By doing so, the radio spectrum can be reused in an opportunistic manner for improved spectrum utilization.

Since reliability and performance can be traded off in communications systems, we can say that classical MAC techniques guarantee maximum safety by trading off spectrum utilization. In contrast, DSA techniques have the potential to maximize the spectrum utilization by trading off spectrum safety. Hence, the design of CR networks places much more emphasis on the confidence of no interference as the primary objective, leaving other performance metrics as secondary objectives.

3.1.2 Motivation

Additive inter-user interference at receivers is an important physical layer characteristic of wireless networks. Based on this, traditional wireless MAC treats simultaneous transmissions by two nearby wireless transmitters as noise, because it leads to collision, which in turn destroys collided transmissions. Orthogonal frequency division multiplexing (OFDM) is a form of transmission that uses a large number of close spaced channels [65]. The signals on these channels are expected to interfere with each other, but, by making the signals orthogonal to each other, there is no mutual interference. We refer to such a flexibility in the physical layer characteristics as *parallelization*.

Diverse traffic behavior of PUs in various channels results in diverse channel opportunities for SUs. Therefore, parallelization is an important feature of cognitive radio channels, where the channels are divided according to the different behavior of PUs. Since the transmission quality of a cognitive radio channel varies in time according to the occupancy of this particular channel by a flow of PU packets, the channel variations correspond to the occupancy of multiple parallel flows of PU packets traveling on their respective channels. This means that a packet flow of PUs using one channel does not affect any other flow using another channel. Hereafter, we use the term *parallel, dynamic channels* by referring to spectrum occupancy variations across time and channels.

In such a system, a SU could possibly utilize the diversity of these channels by selecting the best one to use for transmission [66, 67] provided that the loss at PUs is kept below some threshold. Transmitting on the channel with a relatively favorable condition is referred to as *opportunistic access*. To do so, it is desirable for the transmitter and/or receiver to periodically obtain local information about the quality of the channels by using spectrum sensing.

However, the process of sensing the entire spectrum at one time is challenging, because it is hardware demanding and consumes energy, especially for battery powered users with bursty traffic. Instead, in many practical situations, *some* but *not all* channels can be sensed at one time. The effect to SU access schemes is increased uncertainty in the process of channel selection.

In the presence of channel selection uncertainty, a SU needs efficient transmission strategies that determine which channel to use and when to access it so as to maximize throughput for SUs subject to loss constraints at PUs.

3.1.3 Contributions

The main contributions of this chapter are summarized as follows:

- We study efficient transmission strategies of a single SU that employs a slotted protocol for opportunistic access of parallel, dynamic channels already occupied by PUs. Three classes of control of secondary access are described as efficient transmission schemes to exploit these channels.
- A general Markovian decision model is formulated by including the

suggested classes of access control procedure as special cases to the model. The protocol capacity, expressed in terms of throughput for SUs and loss at PUs, is obtained in terms of cost and reward trade-offs of the resulting decision processes.

- We generalize the single user problem of opportunistic access to a multi-user access setting, where multiple SUs contend to access the channels. A multi-user transmission strategy is described leading to a Markov chain model of the system.
- It is shown that if the exact knowledge of the system states in known to all CR hosts, it is possible to tune the multi-user strategy to an optimal value from which we obtain maximum throughput and minimum delay performance by solving the induced Markov chains.

3.2 System Model

In this section, the system model is described. The statistical model of cognitive radio channel and its properties are presented.

In the context of DSA, CR technology provides SUs with access to multiple channels that are orthogonal with each other, and already occupied by PUs. For simplicity purposes, in Figure 3.1, we use filled circles marked R_1, R_2, R_3, R_4, R_5, R_6 and R_7, respectively, to represent the receiving activities of PU hosts in a certain communication layout. The layout provides seven channels, with a different PU in each. The open circles marked S_1 and D_1, on the other hand, are the cognitive radio transmitter and receiver of one SU pair, respectively.

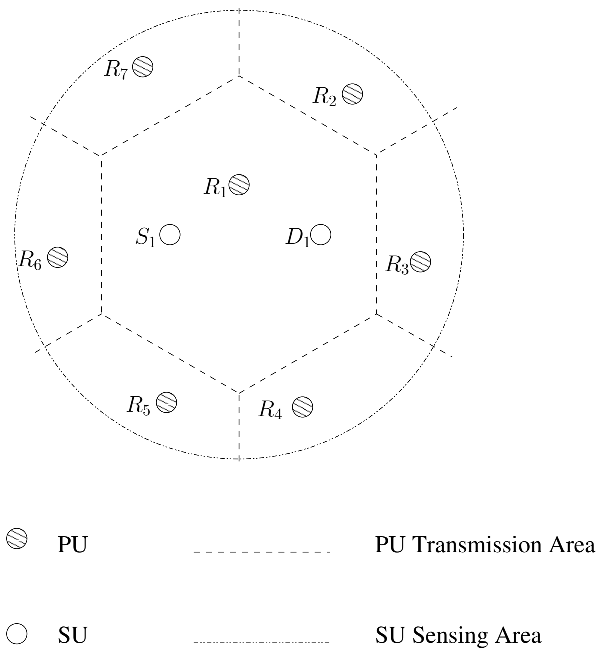

Figure 3.1: *Cognitive radio channel*

Thus, it suffices for a transmitter to utilize the time-varying nature of the multiple channels provision by opportunistically selecting the best one to use for transmission, provided that the loss at PUs on the selected channel is kept below a certain threshold. To do so, it is desirable for the transmitter to periodically obtain information about the quality of these channels by using spectrum sensing. For example, in Figure 3.1, we use the dotted circle around the transmitter S_1 to represent the sensing locality associated with a CR transmitter. This means that, within this area, the transmitter S_1 can detect the transmissions of every PU and, at the same time, every PU is sufficiently protected.

In many practical situations, however, the process of sensing the entire spectrum at one time is challenging, because it is hardware demanding and consumes energy, especially for battery powered users with bursty traffic. Consequently, *some* but *not all* channels can be sensed at one time.

For example, in Figure 3.1, the channels of PUs, within the sensing locality of a transmitter S_1, are numbered according to the order in which they are sensed. The effect to SU access schemes is increased uncertainty in the process of channel selection. Therefore, a system designer is confronted with the problem of searching efficient transmission means to balance the aspirations for maximum performance for SUs and minimal loss at PUs against immunity from uncertainties associated with spectrum sensing deficiencies.

Our objective is, therefore, to design efficient transmission strategies that determine which channel to use and when to access it so as to maximize throughput for SUs subject to loss constraints at PUs, in the presence of channel selection uncertainty.

3.2.1 Cognitive Radio Channel Model

We consider a typical ON-OFF cognitive radio channel associated with the packet train model characteristic of PU traffic passing through a channel [68]. In a packet train model, multiple PU packets, possibly with various lengths, are transmitted within a busy period of a queue from which the packet train is departing. When all PU packets in the queue have been transmitted, the channel becomes idle. The channel remains idle until the arrival of the next packet, which starts the service time of the next packet train.

Assume that the lengths of successive occurrences of busy period are independent and exponentially distributed random variable with mean $1/\mu$, and also that the lengths of successive occurrences of idle period are independent and exponentially distributed random variable with mean $1/\lambda$. It follows that the model of channel occupancy by PUs reduces to a continuous time Markov chain with state space $\{0 \text{ (idle)}, 1 \text{ (busy)}\}$. The generator matrix is [56]:

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \tag{3.1}$$

Hence

$$\eta(0) = \frac{1}{1-\rho}, \qquad \eta(1) = \frac{\rho}{1+\rho}, \qquad \rho = \frac{\lambda}{\mu} \tag{3.2}$$

where the steady-state probability $\eta(c)$ in state $c \in \{0,\ 1\}$ completely specifies the channel whose occupancy state by PUs is observed to be $c \in \{\text{idle}\,(0),\ \text{busy}\,(1)\}$. Equivalently, it determines the channel whose transmission quality for SUs is $c \in \{\text{ON}\ (0),\ \text{OFF}\ (1)\}$. $\rho < 1$ is the traffic intensity on the channel.

3.2.2 Channel Throughput

We assume that SUs employ a slotted access protocol with the probability 1 of transmitting in the ON channel and the probability 0 otherwise. The time slots correspond to the transmission time of a packet used in the system of SUs, whose duration is x seconds.

We are interested in counting the number of transmitted SU packets that are successfully received at the SU receiver per time slot. This quantity measures the protocol throughput for a single SU transmitter-receiver pair. In particular, since the SU transmits only in the channel whose state is ON, it follows that the throughput of the OFF channel is 0 packet per time slot.

Moreover, we assume that a transmission attempted in the ON channel is subject to interruption with a constant interruption rate λ, due to PU returns to occupy the channel. Consequently, the following situations happen: (i) When a SU transmitter starts transmission in a time slot, and it does not collide with the PU during its transmission, this event results in one packet successfully received by the SU receiver; (ii) When a secondary transmission that started with the PU absent experiences a transmission overlap before it finishes transmission, because of PU return, it follows that this particular packet is received in error at the SU receiver. The packet error arises because of the assumption that any overlapping transmissions among SUs and PUs results in the packets being destroyed.

Assuming stationarity conditions, we define the throughput in the channel whose state is c:

$$B(c) = \begin{cases} \exp(-\lambda x) & \text{ON channel (c = 0)} \qquad (3.3a) \\ 0 & \text{OFF channel (c = 1)} \qquad (3.3b) \end{cases}$$

where $\exp(-\lambda x)$ is the same as the probability of non-overlapping transmission during a time interval x. Equivalently, it represents the probability of a successful packet reception on a cognitive channel.

For such a single SU access protocol and a single channel occupied by PUs, the average throughput $\mathcal{U}$ is given by:

$$\mathcal{U} = \sum_c B(c)\, \eta(c) \qquad (3.4)$$

3.2.3 Channel Loss

In the following, we are interested in counting the number of transmitted PU packets that are received in error at the PU receiver in a time slot, as a result of overlapping transmissions between a SU and a PU. This quantity measures the protocol *loss* at PUs caused by SUs, and it serves to quantify the effect of the secondary sharing on the protection of PUs. Since we assume perfect sensing, which means that the SU can always detect the presence or absence of PUs before the execution of a transmission, and that no transmission of SUs is attempted in the OFF channel, it follows that the loss at PUs in the OFF channel is 0 packet error per slot. However, in the ON channel, because of PU returns, the loss of 0 packet error per slot can not be attained.

Let E be a random variable representing the number of slots before the occurrence of an interference caused by a secondary transmission and P_e the distribution of E. The random variable number of slots E is geometrically distributed with parameter $\big(1-\exp(-\lambda x)\big)$, the probability that a packet error occurs. Also, let F denote a random variable representing the number of slots until the next PU packet is successfully transmitted. Hence, the loss at PUs in the ON channel is given by the probability that the event E occurs before the event F:

$$\begin{aligned}
\text{Prob}(F > E) &= \sum_{e=0}^{\infty} P_e \, \text{Prob}(F > e) \\
&= \sum_{e=0}^{\infty} P_e \, \text{Prob}(\text{at least } e \text{ idle slots before first busy slot}) \\
&= \sum_{e=0}^{\infty} \Big[\big(1-\exp(-\lambda x)\big) \, \exp(-\lambda x)^e\Big] \, \eta(0)^e \\
&= \frac{1-\exp(-\lambda x)}{1-\exp(-\lambda x)\, \eta(0)} \qquad (3.5)
\end{aligned}$$

Assuming stationarity conditions, we define the loss at PUs in the channel whose state is c:

$$A(c) = \begin{cases} \dfrac{1-\exp(-\lambda x)}{1-\exp(-\lambda x)\,\eta(0)} & \text{ON channel (c = 0)} \quad (3.6a) \\ 0 & \text{OFF channel (c = 1)} \quad (3.6b) \end{cases}$$

and the average system loss at PUs is given by

$$\mathcal{V} = \sum_{c} A(c)\,\eta(c) \tag{3.7}$$

3.3 Mathematical Model

The details of the parallel nature of channels are hidden in the above model. In the following, we consider such a channel model in the setting of multiple channels. We consider a spectrum consisting of N parallel communication channels for transmissions by PUs. By "parallel channels" we mean that the use of one channel by a particular flow of PUs traffic does not affect the use of any other flow on the channel. As access mode for cognitive radio, we use a slotted transmission protocol.

3.4 Markov Decision Processes

Let $\mathscr{C} = \{1, \cdots, 2^N\}$ be an index set created from possible configurations from which we define a continuous-time Markov chain modeling of spectrum occupancy by PUs. In particular, a spectrum occupancy model consists of N independent Markov chains. Each chain models the PU channel occupancies 0(idle) and 1(busy) according to the pattern of PUs traffic flowing the particular channel.

Equivalently, the two states model of channel occupancy determines the transmission qualities for accessing the particular channel by SU during a time slot. Thus, such a process models the time varying transmission conditions of the various channels at the flow-level where some channels have good conditions (i.e., PU absents) and other have bad conditions (i.e., PU presents).

Consider a discrete time Markov chain $\mathcal{C}_j(x) = [C_1,\, C_2, \cdots, C_N]$ of sensing results indexed by $j = 1,\, 2, \cdots,\, 2^N$, which are observed at time points $x = 0,\, 1\,, 2\,, \cdots$, to be in one of possible transmission qualities of the various channels. We refer to a generic random variable transmission quality of a channel C_i as the *channel state information.* Let $\mathscr{D} = \{d_1,\, d_2, \cdots,\, d_N\}$ be a set of possible actions that a SU host can execute by using a slotted transmission protocol based on the knowledge of the channel states information C_i. We assume perfect sensing by a SU host, i.e., one can always detect the presence or absence of PUs on a channel, and the sensing time is negligible. Therefore, the action $d_i(k) = i$ taken at the beginning of time slot k indicates that a channel i, whose state information takes on an idle outcome, is selected by a SU host to deliver its transmission.

Given prior knowledge of the channel states information $\mathcal{C}_j$ and the action d_i to be realized, $B : \mathscr{C} \to [0,1]$ represents the channel output rate function. A value B_i summarizes the level of performance for SUs following the action $d_i = i$, which is a random variable as well. The randomness comes from the time varying conditions of the channel i. Usually, the better the channel condition, the larger the value of B_i. Typically, $B_i(c)$ represents the probability of transmission success attained on a channel whose state information is known as $(C_i = c)$. Similarly, $A_i(c)$ represents the loss at PUs on channel i, whose state information is known as $(C_i = c)$.

The controlled stochastic process $\{\mathcal{C}_j(x), d_i(x)\}$, with channel states information $C_j(x)$ and control variables $d_i(x)$, both function of time $x = 0,\, 1\,, \cdots$, forms the Markov decision model for a multi-channel opportunistic spectrum access network.

3.4.1 Dynamic Access Control Procedures

By dynamic access control procedure we mean a rule that prescribes the decision d_i based on the knowledge of channel states information C_i, for each $i = 1, 2, \cdots, N$. We denote such a rule π, and it is referred to as the *channel selection strategy* of the Markovian decision model for multichannel opportunistic access networks. Specifically, a strategy π is an algorithm, where for each possible configuration of idle-busy channels, the control action $d_i(\pi)$ identifies the channel i that may be used to deliver a secondary transmission. When π decides using channel i to send a transmission, the channel output rate is B_i. Thus, a fixed strategy π governs the channel selection in each slot, and the corresponding control variable $d_i(\pi)$ denotes the decision one would make knowing the channel state information C_i by following π.

3.4.2 Performance Criterion

The two important dimensions along which a strategy π for opportunistic access of PU channels can be evaluated are: the throughput for SUs that measures the efficiency of secondary sharing of the channels and the loss constraint at PUs that quantifies the protection guarantee provided to PU hosts on their respective channels. The *throughput* $\mathcal{U}(\pi)$ for SUs is the average fraction of transmitted packets that are successfully received at the SU receiver. This serves to indicate the efficiency of secondary sharing of the channels.

$$\mathcal{U}(\pi) = \lim_{K\to\infty} \frac{E_\pi\left[\sum_{k=1}^{K} B_\pi(k)\right]}{K} \tag{3.8}$$

where the expectation $E_\pi[\cdot]$ is taken over the probability distribution induced by π, and $B_\pi(k)$ is the channel output rate of π in time slot k.

At the same time, we define the *loss* $\mathcal{V}(\pi)$ at PUs by the average fraction of transmitted PU packets in a given channel that are received in error at the PU receiver as a result of collisions among a SU and PUs. The *loss constraint* is a maximum level α laid on the loss at PUs on a given channel.

In the absence of the channel selection uncertainty, we say that a strategy π is the optimum strategy or simple *performance-maximizer*, denoted by π^o, when the system with this strategy attempts to maximize the throughput for SUs subject to the loss constraint at PUs:

$$\pi^o = \arg\max_{\pi} \mathcal{U}(\pi) \quad \text{subject to} \quad \mathcal{V}(\pi) \leq \alpha \tag{3.9}$$

3.4.3 Positive Robustness

The notion of *robust-satisfying* strategy consists of finding a strategy that enhances the system immunity to uncertainty arising from the channel selection process [69]. Under the optimal strategy π^o, let $\mathcal{U}^*$ be the optimum throughput for a cognitive radio attained by using the perfect knowledge about the channel states information. Similarly, let $\mathcal{U}(\pi)$ be an unknown actual throughput for a cognitive radio for these channels. We are aware that the actual throughput $\mathcal{U}(\pi)$ may deviate from the optimum throughput $\mathcal{U}^*$. We denote by $\mathcal{A}(\mathcal{R})$ a set of all available designs from which we must choose the strategy π. $\mathcal{A}(\mathcal{R})$ represents a family of sets of $\mathcal{U}$-values, all containing the optimum-throughput $\mathcal{U}^*$ and whose deviation from the optimum-throughput is bounded by $\mathcal{R}$.

$$\mathcal{A}(\mathcal{R}) = \left\{ \mathcal{U} \,:\, |\mathcal{U}(\boldsymbol{\pi}) - \mathcal{U}^*| \leq \mathcal{R} \right\}, \quad \mathcal{R} \geq 0 \tag{3.10}$$

If there is only one design specification for performance, then the performance-maximization option $\boldsymbol{\pi}^o$ is the unique possibility. Thus:

$$\mathcal{R} = 0 \quad \text{means that} \quad \mathcal{A}(0) = \left\{ \mathcal{U}^* \right\} \tag{3.11}$$

As $\mathcal{R}$ increases, the set of available designs becomes more inclusive, that is:

$$\mathcal{R} > \mathcal{R}' \quad \text{means that} \quad \mathcal{A}(\mathcal{R}) \subseteq \mathcal{A}(\mathcal{R}') \tag{3.12}$$

We refer to a control design as being *robust-satisfying* for which a strategy $\boldsymbol{\pi}^\Delta$ is selected satisfying a feasible system-performance, that is, with a positive robustness.

3.5 Perfect Information State Procedures

In this section, we analyze the mathematical model of the channels based on the assumption that a cognitive radio has complete knowledge about channel states information, i.e., *perfect information state*. This may arise in the case the sensing front-end of a cognitive radio supports broadband sensing. A cognitive radio is then provided with the capability to simultaneously observe all N channels at the beginning of every time slot.

Consider N parallel channels already occupied and underutilized by PUs. Each of them evolves as an independent and identically distributed, continuous-time Markov chain with state space $\{0(\text{idle}), 1(\text{busy})\}$. For a channel i, the mean holding times in state 0 and state 1 are λ_i and μ_i, respectively, with steady-state probability of state 0 and 1 being $\eta_i(0)$ and $\eta_i(1)$, respectively.

As access mode, we use a time slotted protocol with parameter $\pi = [d_1, d_2, \cdots, d_N]$. We assume that the time it takes to go through the sensing-transmission process is within the duration of one time slot. In a randomly chosen time slot, the channel throughput $B_i(c)$ is the probability that the considered time slot contains one successful transmission for a SU host, delivered by the channel i whose state information is $C_i = c$. The channel loss $A_i(c)$ is the probability that the considered time slot contains one packet error at PUs in the channel i whose state information is also $C_i = c$.

When acting according to π, a SU host emits one flow using the N channels that are available to it. The problem facing a SU host is to choose a value for π so as to attain maximum throughput $\mathcal{U}(\pi)$ and, at the same time, to ensure that the loss $\mathcal{V}_i(\pi)$ at PUs in channel i, generated by the strategy π, is kept below some threshold value α. In essence, the strategy π is such that associated with each channel i there is a positive number between 0 and 1, $p_i \in [0, 1]$, representing the long run fraction of time slots to be assigned to this particular channel.

We say that these N channels are *stochastically ordered* if there is a permutation $\{1, 2, \cdots, N\}$ of the N channels such that, for all i, j, if $i < j$, the throughput at channel i is greater than the throughput at channel j, $(B_i \geq B_j)$. The additional information, whether the channels are stochastically ordered or not, enables us to divide the access model into two subclasses, namely the *ordered channels information* and *independent channels information.*

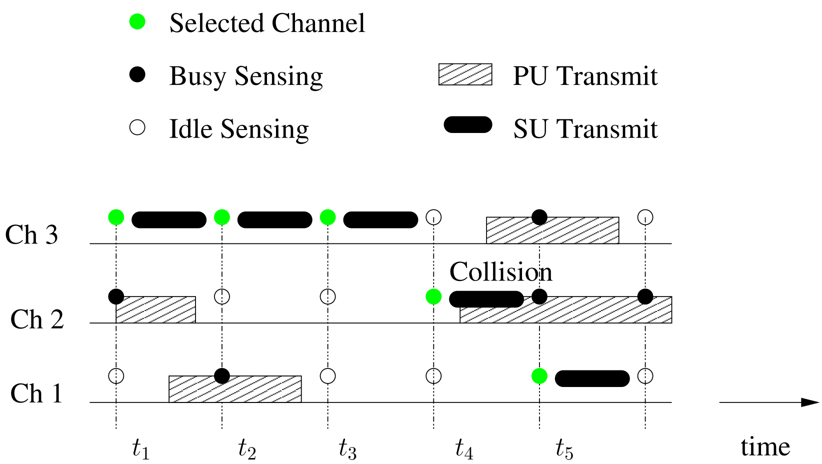

Figure 3.2: *Ordered channels information for multi-channel cognitive radio*

3.5.1 Ordered Channels Information (OCI)

Based on the ordered channels information, which, in our case, is the best model available about the channel selection process, it is of particular interest to choose a strategy π^o maximizing throughput for SU $\mathcal{U}(\pi)$ while the packet error rates of PU flows are satisfactory supported on their respective channels. We refer to such a strategy as throughput-maximizer or simply Ordered Channels Information (OCI).

Under an OCI procedure, a cognitive radio is provided with a list of N ordered channels. A sensing outcome under such a procedure is generated at time slots $x = 0,\ 1,\ 2, \cdots$, and it consists of the latest N-tuple $\left[C_1(x), C_2(x), \cdots, C_N(x)\right]$ of 0s and 1s. A value of 0 in the i-th position taken by C_i indicates that the channel i is deemed idle, i.e., the channel state information realizes in a good transmission quality. Since the system is i.i.d. across slots, maximizing the system throughput is equivalent to maximizing the expected throughput in each slot. For notation simplicity, we consider a generic time slot and drop the time x from the notation.

The rational of the OCI procedure is that, since a channel can deliver a transmission only after knowing the channel state information, provided that the channels are ordered, the system throughput is maximized by transmitting at the first channel deemed idle. We only need to maintain the first channel whose state information takes on an idle outcome.

Let us define $\mathcal{P}_i$ as a set of channels whose states information have been seen so far through sensing in a time slot, where i indicates that the system has already found out the states of i channels. Denote by u the best achievable channel condition seen among the channels in $\mathcal{P}_i$. We say of such a system to be in state $(\mathcal{P}_i, u)$. That is, we are to choose a probability of stopping at channel i, which is typically $\eta_i(0)$, $0 \leq \eta_i(0) \leq 1$, for each $i = 1, 2, \cdots, N$. For instance, in state $(\mathcal{P}_1, u)$, the channel in position 1 is the first to be sensed and given that the state of the first channel comes out idle, $\eta_1(0)$ represents the probability of stopping after the first sensing attempt, and so on.

Figure 3.2 gives a simple illustration of acting according to the OCI procedure. Thus, the OCI procedure has the following possible actions in state $(\mathcal{P}_i, u)$:

1) Select the best probed channel (i.e., a channel whose state is deemed idle, say i) and with probability d_i a transmission is done on the channel.

2) Probe one more channel among the remaining $(N - i)$ channels. In

this case, the system state $(\mathcal{P}_i, u)$ changes to $(\mathcal{P}_{i+1}, u \vee C_{i+1})$. C_{i+1} is the state of the sensed channel newly included and $\mathcal{P}_{i+1} = \mathcal{P}_i \cup \{i+1\}$. The factor $a \vee b$ denotes $\max(a,b)$.

3) Otherwise, the transmitter remains silent.

Hence, the OCI procedure determines the probability distribution over the respective channels i at which stopping occurs:

$$\Big[p_1(C_1),\ p_2(C_1,C_2), \cdots,\ p_N(C_1,C_2,\cdots,C_N)\Big]$$

where $p_i(C_1,C_2,\cdots,C_N)$ determines the probability mass of stopping at channel i, and it is given by:

$$\begin{aligned} p_1 &= \eta_1(0) \\ &\vdots \\ p_N &= \left(\prod_{i=1}^{N-1}\big(1-\eta_i(0)\big)\right)\eta_N(0) \end{aligned} \tag{3.13}$$

Let $\Delta(\alpha)$ denote the set of all admissible channel selection strategies, that is, the set of d_i for which the loss constraint α at PUs is satisfactory met. A fixed value α, which determines the maximum level on the loss $\mathcal{V}_i(\pi)$ at PUs in channel i, results in each channel i solving for the d_i satisfying:

$$\Delta(\alpha) = \Big\{d_i\ :\ \mathcal{V}_i(\pi) = \alpha\Big\}, \quad \forall i \tag{3.14}$$

It follows that, if $\mathcal{V}_i$ are uniquely specified for all i, one can find d_i by using a simple linear program:

$$d_i = \begin{cases} \dfrac{\alpha}{p_i A_i(0)} & \text{if} \quad \mathcal{V}_i > \alpha, \quad \forall i \qquad (3.15\text{a}) \\ 1 & \text{otherwise} \qquad (3.15\text{b}) \end{cases}$$

In response to the demand for maximum performance, one can find the maximum system throughput $\mathcal{U}_{\mathtt{OCI}}(\pi)$ obtainable with OCI to be:

$$\mathcal{U}_{\mathtt{OCI}}(\pi) = \sum_{i=1}^{N} p_i d_i B_i(0) \tag{3.16}$$

$$= \sum_{i=1}^{N} \frac{B_i(0)\,\hat{\alpha}_i}{A_i(0)}, \qquad \hat{\alpha}_i = \min\left\{\alpha, p_i\, A_i(0)\right\} \tag{3.17}$$

3.5.2 Independent Channels Information State

The statistical independent information of the channels entails the best model of the system in $\Delta(\alpha)$ as well. That is, one can find another channel selection strategy $\pi^o \in \Delta(\alpha)$ that maximizes the throughput performance criterion. We refer to such a strategy as Independent Channels Information (ICI).

Under an ICI procedure, a cognitive radio is provided with a list of N independent channels, as well. A sensing outcome under such a procedure is generated at time slots $x = 0,\, 1,\, 2, \cdots$, and consists of the latest N-tuple $[C_1(x), C_2(x), \cdots, C_N(x)]$ of 0s and 1s. A value of 0 in the i-th position taken by C_i indicates that the channel i is deemed idle, i.e., the channel state information realizes in a good channel condition. Since the system is i.i.d. across slots, maximizing the system throughput is equivalent to maximizing the expected throughput in every slot. For notation simplicity, we consider a generic time slot and drop the time x from the notation.

Let $\mathcal{P}_u$ be a set of channels whose state information C_i realizes in good transmission quality after sensing, where u indicates the number of channels

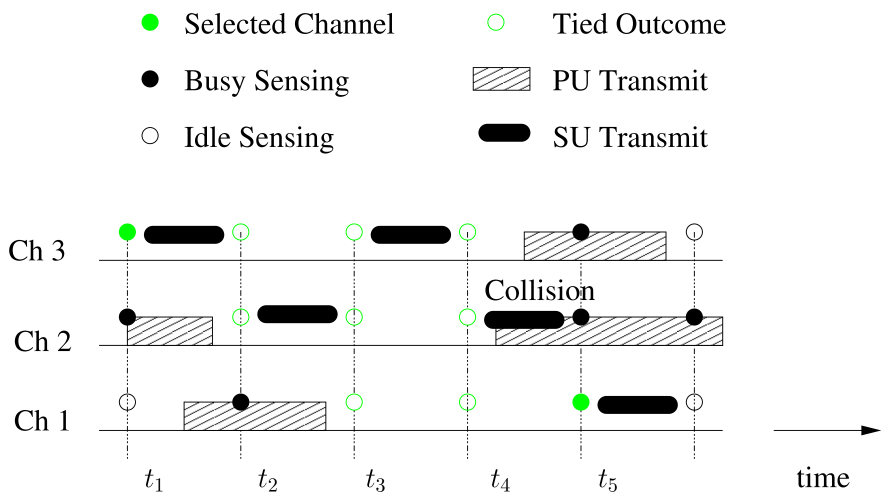

Figure 3.3: *Statistical independent channels information for multi-channel cognitive radio*

with good transmission quality. We denote by i the channel index whose state information results in the best channel condition among the channels in $\mathcal{P}_u$. Ties are broken arbitrary. We say of this system to be in state $(\mathcal{P}_u, i)$, that is, with probability $\eta_i(0)$, $0 \leq \eta_i(0) \leq 1$, the channel i is idle, for all $i = 1, 2, \cdots, N$.

Figure 3.2 gives a simple illustration of acting according to the ICI procedure. Thus, the ICI procedure has the following possible actions in state $(\mathcal{P}_u, i)$:

1) if $u = 1$, then state $(\mathcal{P}_1, i)$ contains only one element so that the strategy selects channel i and with, probability $d_{i,1}$, a transmission is done on the channel.

2) if $u > 1$, then state $\mathcal{P}_u$ contains more that one element and ties always

happens. In this case, tied outcomes need to be broken arbitrary to resolve the channel index i. The strategy then selects the channel i and with, probability $d_{i,u}$, a transmission is done on the channel.

2) Otherwise, the transmitter remains silent.

Hence, the ICI procedure determines the probability distribution over the respective channels i based on the number of idle sensing outcome, denoted by:

$$\Big[p_{1,u}(C_1,\cdots,C_N),\ p_{2,u}(C_1,\cdots,C_N),\cdots,\ p_{N,u}(C_1,\cdots,C_N)\Big],\quad u=1,\cdots,N$$

where $p_{i,u}(C_1,\cdots,C_N)$ determines the probability mass assignment to channel i given that u channels have been seen having idle outcomes. This is given by:

$$p_{i,u} = \frac{\prod_{w=1}^{N}\Big(\mathbb{1}_{[C_w=0]}\eta_w(0)+\mathbb{1}_{[C_w=1]}\big(1-\eta_w(0)\big)\Big)}{u} \tag{3.18}$$

where $\mathbb{1}_{[\cdot]}$ is a function that maps "true" to 1 and "false" to 0.

Since the ICI belongs to $\Delta(\alpha)$, one standard algorithm to solve equation (3.14) is based on simple control rules u. First, we enumerate possible number of channels having an idle outcome by $u = 1,\ 2,\cdots,\ N$, so that, corresponding to each channel i, the quantities $V_{i,1},\ V_{i,2},\cdots,\ V_{i,N}$ form the respective contributions to the loss at PUs in the channel i, $\mathcal{V}_i(\pi)=\sum_u V_{i,u}$.

For a fixed value of α, the control rule $1 \le u \le N$ on the number of idle channels results in each channel i solving, for the smallest index u such that:

$$V_{i,1}+V_{i,2}+\cdots+V_{i,u} \ge \alpha \tag{3.19}$$

It follows that, if $V_{i,u}$ are uniquely specified for all i and u, one can find $d_{i,u}$ by using a simple linear program:

$$d_{i,u} = \begin{cases} \dfrac{\alpha_i}{p_{i,u}A_i(0)} & \text{if} \quad u = 1, \forall i \qquad (3.20a) \\ \dfrac{\alpha_i - \sum_{w=1}^{u-1} p_{i,w}A_i(0)}{p_{i,u}A_i(0)} & \text{if} \quad 1 < u < N, \forall i \qquad (3.20b) \\ 1 & \text{otherwise} \qquad (3.20c) \end{cases}$$

In light of demand for maximum performance, one can find the maximum system throughput $\mathcal{U}_{\mathtt{ICI}}(\pi)$ obtainable with ICI to be:

$$\mathcal{U}_{\mathtt{ICI}}(\pi) = \sum_{i=1}^{N} \sum_{u=1}^{N} p_{i,u} d_{i,u} B_i(0) \tag{3.21}$$

$$= \sum_{i=1}^{N} \frac{B_i(0)\ \hat{\alpha}_i}{A_i(0)}, \qquad \hat{\alpha}_i = \min\left\{\alpha, \sum_u p_{i,u}\, A_i(0)\right\} \tag{3.22}$$

3.6 Imperfect Information State Procedures

As discussed in the previous sections, a channel selection strategy maximizing system performance hinges on the optimal use of the statistical information about the channels. In this section, we consider special cases where the resources available to a cognitive radio enable sensing of one channel within a time slot, also referred to as *Imperfect Information State Procedures.*

We are given N parallel channels already occupied and underutilized by PUs. Each of them evolves as an independent and identically distributed, two-state continuous-time Markov chain. For a channel i, the mean holding times in state 0 and state 1 are λ_i and μ_i, respectively, with steady-state probability of state 0 being $\eta_i(0)$.

As access mode we adopt a time slotted protocol with parameter $\pi = [d_1, d_2, \cdots, d_N]$. We assume that the time it takes to go through the sensing-transmission process is within the duration of one time slot. In a randomly chosen time slot, the channel throughput $B_i(c)$ is the probability that the considered time slot contains one successful transmission for a SU host, delivered by the channel i whose state information is $(C_i = c)$. The channel loss $A_i(c)$ at PUs is the probability that the considered time slot contains one packet error at PUs whose state information is also $(C_i = c)$ in the channel i.

At the beginning of each time slot, a SU has a time designated to perform channel sensing. The hardware limitations and energy cost of spectrum sensing put a limit on the time allocated to the channel sensing operation, so that a cognitive radio can sense only one channel during the sensing time. Thus, a sensing operation on a channel i does not provide any information about any other channel j, but only the outcome of the considered channel, i.e., the true state of the channel i. This has the implication that the internal states of the system are known imperfectly when making the channel selection decision, and thus we have the uncertainty about the channel selection process.

The deficiencies in the knowledge about the channel states information reduces to uncertainty about the channel selection process. This confronts us with the question on how to trade off the system performance against the immunity of the system to uncertainty arising from spectrum sensing deficiencies. This leads to the analysis of performance-sub-optimal procedures, which enhances the robustness to uncertainty while satisfying performance at levels not too much less than the performance-optimum.

3.6.1 Design of Robust Hopping Sequence

The imperfect information about the channels arises when a cognitive radio is equipped with one tunable half-duplex radio transceiver able to switch over different channels. To execute both sensing and access operations during a time slot, a SU host hops over the channels available to it based on a predefined hopping sequence. A hopping sequence determines which channel to use for sensing and access decision. If the channel is observed to have an idle outcome, then the channel is open for secondary access for the remainder of the time slot. Otherwise, the channel is having a busy outcome and no transmission is attempted in the remainder of this particular time slot.

To support opportunistic use of the multichannel capability, a hopping sequence is designed with multiple hopsets during a period of K time slots. In particular, a scheduling period of K slots represents the time slots assignment to the protocol for secondary access. To do so, during a scheduling period, we combine multiple hopsets to form a hopping sequence that covers all channels to which a user has access. That is, if a user has access to N channels one can form the respective hopsets by $n_1, n_2, \cdots, n_N$. A hopset n_i, associated with the channel i, is a set of time slots during which the sensing and access operations can execute on this particular channel. For instance, $n_1 = \{1, 4, 5\}$, means that channel 1 is assigned time slots 1, 4 and 5 for sensing and access decision.

We enumerate possible hopping channels by $i = 1, 2, \cdots, N$, with respective probabilities p_i that a time slot is allocated to channel i. The probability p_i of hopping to channel i is defined as the fraction of the time slots channel i is used for sensing and access decisions, where all probabilities are independent. It follows that a process $\mathcal{C}_i(x)$, which reports at time slot $x = 1, 2, \cdots$ the state information C_i associated with the hopping channels i, is completely specified by the hopping probabilities p_i and the entropy function H:

$$H(p) = -\sum_{i=1}^{N} p_i \log p_i \tag{3.23}$$

The entropy function can be viewed as the expected amount of uncertainty that exists before a channel is used for hopping. In absence of prior information, the entropy attains its maximum when the hopping probabilities are equally probable. In more general context, the principle of maximum entropy provides that prior probabilities can be constructed by maximizing the entropy under the constraints at hand [70]. The solution is considered to be the least biased possible as any other solution would imply lower entropy and thus lead to more predictable state than the one implied by the given information. Therefore, if a hopping sequence is viewed as a sample path of an underlying structure state process C_x, the corresponding channel selection strategy π determines the hopping probabilities p_i for which the entropy function H is maximum.

3.6.2 Skewed Hopping Channels Information

The skewed hopping information state applies in the situation where a SU wishes to meet a requirement for feasible performance, that is, it tries to find a channel selection strategy π for which the system throughput is guaranteed to be at least as large as some $\mathcal{U}_{\mathtt{SHI}}(\pi)$, $U_{\mathtt{SHI}}(\pi) \leq \mathcal{U}^*$. $\mathcal{U}^*$ is the optimum system throughput. This condition is necessary to enable positive robustness with reference to uncertainty about the channel selection process. We refer to such a strategy as *robust satisfaying* or simply Skewed Hopping Channels Information (SHI).

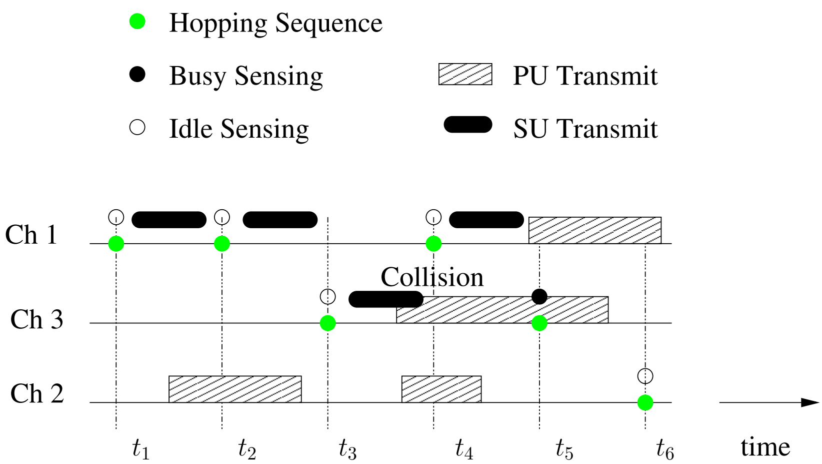

Figure 3.4: *Skewed hopping channels information for multi-channel cognitive radio*

Based on the magnitude of the average channel loss at PUs in a time slot $A_i(0)\eta_i(0)$, the strategy SHI is implemented by avoiding excessive loss in the channels. A channel i is classified as having excessive loss if the estimated average loss at PUs in the channel i is greater than a given threshold γ. Otherwise, average loss at PUs in the channel i is less than this threshold, and the corresponding channel is classified as having tolerable loss.

In this situation, every channel classified as having excessive loss is avoided by assigning to its hopset n_i few time slots, whereas every channel classified as having tolerable loss is favored by assigning to its hopset n_i more time slots. Equivalently, a skewed distribution is our way for doing this, so that a host will spend larger fraction of its time slots hopping over the channels having tolerable packet errors. Hence, a SHI maintains orderings of probability distribution over the hopping channels $\left[p_1,\, p_2, \cdots,\, p_N\right]$ that are decreasing.

Figure 3.4 provides a simple illustration of a hopping sequence generated under the SHI design by using three possible hopping channels over a period of six time slots. The hopping sequence in the example consists of three channels, each of which is associated with one hopset. In a cycle of 6 time slots, channel 1 is allocated the time slots in the hopset n_1 (i.e., time slots 1, 2, and 4), channel 3 is allocated the time slots in hopset n_3 (i.e., time slots 3 and 5) and channel 2 is allocated the time slot in hopset n_2 (i,e., time slot 6).

$$\begin{aligned}\text{SHI based hopping sequence} &= \{n_1,\, n_3,\, n_2\} \\ &= \Big\{\{1,\, 2, 4\},\, \{3,\, 5\},\, \{6\}\Big\}\end{aligned}$$

Let the available information about the hopping channel distributions $\{p_i\}$ be such that in a time slot the set of average losses at PUs in every channel i, $\{A_i(0)\eta_i(0)\}$, is uniquely specified. The principle of maximum entropy results in that, for all distributions satisfying the constraints supplied by this information, the minimally prejudiced hopping distribution is the one that maximizes the system entropy function $H(p)$ subject to the constraints:

$$\sum_{i=1}^{N} p_i = 1 \tag{3.24}$$

$$\sum_{i=1}^{N} A_i(0)\eta_i(0)p_i = \gamma \tag{3.25}$$

3.6.3 Finding a Maximum Entropy Distribution

The maximum entropy distribution is found by using the Lagrange method [70]. Using the constraints in Equations 3.24 and 3.25, where:

$$0 = \sum_{i=1}^{N} p_i - 1 \tag{3.26}$$

$$0 = \sum_{i=1}^{N} p_i A_i(0)\eta_i(0) - \gamma \tag{3.27}$$

one can form the functional entropy:

$$H(p) = -\sum_{i=1}^{N} p_i \log p_i + \beta_1 \Big(\gamma - \sum_{i=1}^{N} p_i A_i(0)\eta_i(0)\Big) + \beta_2 \Big(\sum_{i=1}^{N} p_i - 1\Big) \tag{3.28}$$

that is maximized by differentiating with respect to p_i:

$$\frac{\partial H(p)}{\partial p_i} = -\log p_i - 1 - \beta_1 A_i(0)\eta_i(0) - \beta_2 \tag{3.29}$$

Setting this to zero, we have the general form of the entropy maximizing the probability mass given by:

$$p_i = \frac{1}{\exp(\beta_2)} \exp\Big(-\beta_1 A_i(0)\eta_i(0)\Big) \tag{3.30}$$

where $H(p)$ is the Lagrange function, with β_1 and β_2 being the Lagrange multipliers.

To obtain β_2, if Equation 3.30 for p_i is summed over the probabilities, one can then obtain:

$$\beta_2 = \log\left(\sum_{i=1}^{N} \exp\Big(-\beta_1 A_i(0)\eta_i(0)\Big)\right) \tag{3.31}$$

Given the expression for β_2, we can find β_1. We multiply Equation 3.30 for p_i by $A_i(0)\eta_i(0)$ and by $\exp(\beta_2)$, and sum over the probabilities so that we can write:

$$0 = \sum_{i=1}^{N} \Big(\gamma - A_i(0)\eta_i(0)\Big) \exp\Big(-\beta_1 A_i(0)\eta_i(0)\Big) \tag{3.32}$$

If this equation is multiplied by $\exp(\beta_1 \gamma)$, the result is obtained by solving for β_1:

$$f(\beta_1) = 0 \tag{3.33}$$

where

$$f(\beta_1) = \sum_{i=1}^{N} \Big(\gamma - A_i(0)\eta_i(0)\Big) \exp\Big(-\beta_1(\gamma - A_i(0)\eta_i(0))\Big) \tag{3.34}$$

For a fixed value γ, the value of β_1^* that satisfies the maximum entropy distribution is the value for which $f(\beta_1^*) = 0$. Since the SHI design applies whenever γ lies between the smallest value of $A_i(0)\eta_i(0)$ and the largest one, there exists at least one value for which the quantity $\Big(\gamma - A_i(0)\eta_i(0)\Big)$ is positive and at least one for which this quantity is negative.

Let $\hat{\Delta}(\alpha)$ be the set of all feasible channel selection strategies π, that is, the set of d_i for which the loss constraint α at PUs is satisfactory met. This means that, if a hopping channel i is observed to have a idle outcome, with probability d_i a transmission executes in the considered channel. Otherwise, the hopping channel i is observed to have a busy outcome in which case no transmission is attempted until the next hop. In this case, a fixed value α, which determines the maximum level on the loss $\mathcal{V}_i(\pi)$ at PUs in a channel i with the SU transmitting according to π, results in every channel i solving for the d_i satisfying:

$$\hat{\Delta}(\alpha) = \Big\{d_i \, : \mathcal{V}_i(\pi) = \alpha\Big\} \tag{3.35}$$

where $\mathcal{V}_i(\pi) = p_i A_i(0)\eta_i(0)$.

It follows that, if p_i are uniquely specified for all i, one can find d_i by using a simple linear program:

$$d_i = \begin{cases} \dfrac{\alpha}{p_i\,\eta_i(0)\,A_i(0)} & \text{if} \quad A_i\,\eta_i(0)\,p_i > \alpha & (3.36a) \\ 1 & \text{otherwise} & (3.36b) \end{cases}$$

Thus, the throughput of the system under the SHI strategy is given by:

$$\begin{aligned} \mathcal{U}_{\mathtt{SHI}}(\pi) &= \sum_{i=1}^{N} p_i\,\eta_i(0)\,d_i\,B_i(0) \qquad (3.37) \\ &= \sum_{i=1}^{N} \frac{B_i(0)\,\tilde{\alpha}_i}{A_i(0)}, \qquad \hat{\alpha}_i = \min\{\alpha, p_i\,\eta(0)\,A_i(0)\} \end{aligned}$$

3.6.4 Round Robin Hopping Information

The round robin hopping channels [27] is a special case of skewed hopping channels where the hopping probabilities p_i are fixed to $1/N$. We refer to the round robin hopping based strategy as simply Round Robin Hopping (RRH).

Thus, the throughput of the system under the RRH stategy is given by

$$\begin{aligned} \mathcal{U}_{\mathtt{RRH}}(\pi) &= \sum_{i=1}^{N} \frac{d_i\,\eta_0\,B_i(0)}{N} \qquad (3.38) \\ &= \sum_{i=1}^{N} \frac{B_i(0)\,\tilde{\alpha}_i}{A_i(0)}, \qquad \hat{\alpha}_i = \min\{\alpha, \frac{1}{N}\eta_0\,A_i(0)\} \end{aligned}$$

where η_0 is the probability that channel i is idle and d_i is given by

$$d_i = \begin{cases} \dfrac{N\alpha}{\eta_0 A_i} & \text{if } \ \frac{1}{N}\eta_0 A_i(0) > \alpha \quad (3.39a) \\ 1 & \text{otherwise} \quad (3.39b) \end{cases}$$

3.7 Extension to Multi-user Access

The parallel nature of cognitive radio channels yields a natural extension to the case with a finite user population of M hosts contending to access N channels of PUs. In such a system, each SU host can be in one of two states: *blocked* or *active*. In the active state, a SU host can accept from its application or generates one packet with a constant probability g. From the time moment a host receives a packet until the time moment the packet is successfully transmitted, we say that this particular host is blocked, in the sense that it can not generate or accept a new packet to send. A blocked SU host returns to the active state after a successful transmission.

Since the network of SUs operates with time-varying channels, a centralized protocol is typically the best option to handle the channels variations. A central controller, i.e., a base station or a suitable distributed protocol, is assumed to synchronize the slot-starts of every SU hosts. This has the consequence that every SU host always starts the process of sensing and transmission at the beginning of a fixed-length time slot. The operation of such a protocol follows a two stage slot delivery process: contention resolution phase and interference resolution phase.

Stage 1 – When a time slot begins, only SU hosts with a packet enter the contention phase. Based on full users state information, i.e, the information about which SU hosts are blocked and which ones are active, the system maintains a number u, $0 \leq u \leq M$, of contending hosts. All blocked SU hosts simultaneously select from the channel poll one channel by sending a Request for Transmission Opportunity (RTO) message with a probability $h_u \in \{h_0, h_1, \cdots, h_M\}$ adjusted to the available information. The base station delivers the feedback message for each channel where the contention-free RTO message has been successfully collected. A host that does not receive a positive acknowledgement for its RTO message keeps the packet for the next slot.

Stage 2 – After the contention phase, the SU transmissions are completely orthogonalized. Following the grant of contention-free hopping channels, there are fewer contention-free SU hosts than the number of channels so that these hosts can be fit into orthogonal hopping sequences. By doing so, one obviates collisions among SUs during the transmission phase. Based on full channel state information, i.e., the information about which channels are idle and which ones are busy, a SU host can send a packet with a probability $d(\alpha)$ adjusted to meet the loss constraint α at PUs on a given channel. The base station delivers at the end of the time slot the feedback message for each channel where the inference-free packet has been successfully received. Only a host that receives a positive acknowledgement for packet reception can remove the packet from the buffer.

We operate the protocol under the following assumptions: (1) Each SU host has a buffer to store only one packet; (2) When a SU host generates a new packet, it is treated the same as an existing blocked host.

3.7.1 Multi-user Access Control Strategy

Let W_x be a random variable number of blocked hosts that represents the number of hosts with a stored packet to be transmitted at the beginning of a time slot x. This value decreases when a transmission is attempted with probability h_u for a successful transmission and it increases when a new packet is generated with probability g. The quantity W_x takes on one of the $(M+1)$ possible values of $u \in \{0, 1, \cdots, M\}$, and it serves as the state variable description of the system. Because of the memoryless assumption, the process W_x is a finite-state Markov chain with the state transition probability matrix $[q_{uw}]$ and stationary probabilities $\{Q_u : u = 0, 1, \cdots M\}$, where $q_{uw} = \text{Prob}(W_{x+1} = w | W_x = u)$ and $Q_u = \text{Prob}(W_x = u)$.

A multi-user access control strategy is a rule for choosing the transmission probability h_u as the system evolves. Since the selection of h_u depends only on the state of the process W_x observed at a time slot x, a multi-user access strategy has the form $[h_0(x), h_1(x), \cdots, h_M(x)]$, where h_u is used when the system is observed in state u at time slot x. When such a strategy is embedded with the process W_x, the corresponding controlled system is still a finite-state Markov chain but the transition matrix $[q_{uw}(h_u)]$ is now a function of h_u. Hence, the stationary probabilities $Q_u(h_u)$ exist and satisfy the following set of linear equations:

$$
\begin{aligned}
Q_w(h_u) &= \sum_{u=0}^{M} Q_u(h_u)\, q_{uw}(h_u), \quad w = 0\, 1, \cdots, M \qquad (3.40)\\
1 &= \sum_{u=0}^{M} Q_u(h_u)
\end{aligned}
$$

3.7.2 Capacity of Cognitive Channels

We are interested in counting the number of successful transmissions in any slot as well as the number of blocked transmissions that continue to attempt in the next slot. These slots correspond to the transmission time of a packet used in the system. The maximum throughput is one packet per slot for a single channel and one packet per slot times the number of channels N for the whole system. However, in the presence of parallel, dynamic channels, we need to determine the capacity of cognitive radio channels as a measure of channel transmission qualities.

For simplicity purposes, we study a special case with homogeneous channels that are statistically identical with one another. We consider N parallel channels already occupied and underutilized by PUs. Each of the channels evolves as a homogeneous, two-state $\{\text{idle}(0), \text{busy}(1)\}$ continuous-time Markov chain. The mean holding times in state 0 and state 1 are λ and μ, respectively. The loss at PUs on a given channel A and the channel rate B are also given.

Typically, if we let the number $C = 0,\ 1,\ ,2, \cdots,\ N$ of idle channels be the channel state information in a particular time slot, a $M/M/N$ queuing system with finite user population can, therefore, be constructed as a statistical model for the channels. Hence, the channel state information that determines the channel transmission qualities is completely specified by the steady state probabilities:

$$p(c) = \frac{\binom{N}{c}\left(\frac{\lambda}{\mu}\right)^{c}}{\sum\limits_{v=0}^{N}\binom{N}{v}\left(\frac{\lambda}{\mu}\right)^{v}} \tag{3.41}$$

We find the throughput of the channels in terms of the rate R of interference-free transmissions by SUs in a time slot. To do so, when a transmission phase starts, we use the protocol parameter the transmission probability d and the number S of contention-free SU hosts. In particular, at the beginning of a time slot, a number $W = 0, 1, 2, \cdots, M$ of SU hosts start contending with each other for access over N channels. At the end of this contention phase, there is a number $S = 0, 1, 2, \cdots, \min\{W, N\}$ of contention-free SU hosts orthogonal to each other that have been allocated hopping channels. The randomness in R arises from the random variables S and W.

Following the grant of contention-free hopping channels, two situations may happen: (i) Every SU hosts that experiences a busy channel keeps the packet for the next slot. (ii) Every SU hosts that experiences an idle channel sends the packet with probability d adjusted to the current number $(S = s)$ of SU hosts.

By taking into consideration the loss constraint α at PUs on a channel that is now imposed on the transmissions of S SUs, we obtain the transmission probability d for which the loss at PUs in a given channel, contributed by the transmissions of $(S = s)$ SUs, satisfies:

$$\sum_{c=1}^{N} \frac{s}{N}\, p(c)\, A(c)\, d \leq \alpha \tag{3.42}$$

Therefore, considering the number S of contention-free SUs in a N channels system with loss constraint α at PUs on every channel, a SU transmits in a channel with the transmission probability d given by:

$$d(s) = \min\left\{ \frac{\alpha\, N}{\sum\limits_{c=1}^{N} s p(c)\, A(c)}, 1 \right\} \tag{3.43}$$

We obtain the throughput of the channels, which represents the random variable $R(s)$ rate of successful transmissions, as follows:

$$R(s) = \frac{1}{N}\sum_{c=1}^{N} p(c)\, B(c)\, d(s) \tag{3.44}$$

3.7.3 Performance Modeling

Two objective functions from which any multi-user access strategy can be evaluated are the system throughput $\mathcal{S}$ and the packet delay $\mathcal{D}$. At the end of a time slot x, we define the random variable $(S_x = s)$ number of orthogonal transmitting hosts and the rate of successful transmissions $R(s)$ associated with s. Thus, the average number of successful transmission for the system can be obtained from:

$$\mathcal{S} = \sum_{s=0}^{N} s\,\mathrm{Prob}(S_x = s) R(s) \tag{3.45}$$

where

$$\mathrm{Prob}(S_x = s) = \sum_{u=s}^{M} \mathrm{Prob}(S_x = s | W_x = u)\mathrm{Prob}(W_x = u) \tag{3.46}$$

and $\mathrm{Prob}(S_x = s|W_x = u)$ refers to the conditional probability of having s orthogonal transmitting hosts given u contending hosts. $\mathrm{Prob}(W_x = u)$ refers to the steady state probability of having u contending hosts.

At the same time, we define the length of time for which a packet is blocked until the packet is sent successfully by the random variable packet delay $\mathcal{D}$. The average delay of the system can be obtained from:

$$\mathcal{D} = \frac{\mathcal{S}}{\mathcal{W}} \tag{3.47}$$

where $\mathcal{W}$ is the average number of blocked hosts in a time slot, which is given by:

$$\mathcal{W} = \sum_{u=0}^{M} u \, \mathrm{Prob}(W_x = w) \tag{3.48}$$

Based on the perfect knowledge of the system state, we choose the optimal value of h_u to be:

$$h_u = \min \left\{ 1, \frac{N}{u} \right\} \tag{3.49}$$

resulting so in a transition probability matrix, which in turn results in a state probability vector that minimized the average number $\mathcal{W}$ in equation (3.48) and, at the same time, the system throughput $\mathcal{S}$ in Equation (3.45) maximized, and with the system packet delay $\mathcal{D}$ in Equation (3.47) minimized.

3.7.4 Induced Markov Chain Derivations

The derivation of the state transition probability matrix $[q_{uw}]$ follows a similar approaches as in [71] and [72], but it uses a simple recursive counting algorithm suggested in [73] to obtain the numerical results. The key contribution to the related analytical frameworks is the application of a backoff scheme that spreads randomness in the frequency domain to resolve contention effects.

We are given an homogeneous, aperiodic and irreducible Markov chain number of blocked SU hosts $\{W_x : x = 1, 2, \cdots\}$ defined with state space $\{1, 2, \cdots, M\}$ and transition matrix $[q_{uw}]$. The random variable W_x denotes the number of contending SU hosts in stage 1 at the beginning of a generic time slot x. Conditioned on this number, let the random variable S_x, with $0 \leq S_x \leq \min(N, W_x)$, denote the number of orthogonal transmitting SU hosts in time slot x.

Let V_{x+1}, with $0 \leq V_{x+1} \leq M - W_x$, be the random variable number of active hosts having a new packet arrival in time slot x. That is, V_{x+1} represents the number of hosts that join the existing blocked hosts at time slot $(x+1)$, so that the following equation holds:

$$W_{x+1} = W_x - S_x + V_{x+1} \tag{3.50}$$

Or, equivalently:

$$V_{x+1} = w - u + s, \tag{3.51}$$

where $W_{x+1} = w$, $W_x = u$, $S_x = s$, and $V_{x+1} = v$

Derivation of the State Transition Matrix $[q_{uw}]$

The following rules of probability theory applies:

$$\begin{aligned} q_{uw} &= \text{Prob}(W_{x+1} = w | W_x = u) \\ &= \sum_{s=0}^{\min(N,u)} \text{Prob}(W_{x+1} = w,\ S_x = s | W_x = u) \end{aligned} \tag{3.52}$$

Derivation of Prob$(w,s|u)$

The expression for $\text{Prob}(w,s|u)$ uses the following property:

$$\begin{aligned}\text{Prob}(w,\ s|u) &= \text{Prob}(W_{x+1} = w|S_x = s,\ W_x = u)\ \text{Prob}(S_x = s|W_x = u)\\ &= \text{Prob}(V_{x+1} = w-u+s|s,\ u)\ \text{Prob}(s|u) \qquad (3.53)\end{aligned}$$

By using Equation 3.53, and based on the fact that the hosts in active mode generate new packets with a constant probability g, we have

$$\text{Prob}(w|s,u) = \begin{cases} \text{Prob}(w-u+s|s,u) & u \leq w+s \leq M \qquad (3.54a)\\ 0 & \text{otherwise} \qquad (3.54b)\end{cases}$$

where

$$\begin{aligned}\text{Prob}(w-u+s|s,u) &= \text{Prob}(w-u+s|u) \qquad (3.55)\\ &= \binom{M-u}{w-u+s} g^{w-u+s}\,(1-g)^{M-w-s}\end{aligned}$$

Derivation of Prob$(s|u)$

Since a transmission in blocked state is sent with probability h, let T_x denotes the random variable number of transmitting hosts in time slot x among the blocked hosts. Similarly, the expression for $\text{Prob}(s|u)$ uses the following property:

$$\begin{aligned}
\text{Prob}(s|u) &= \sum_{t=s}^{u} \text{Prob}(S_x = s, T_x = t | U_x = u) \\
&= \sum_{t=s}^{u} \text{Prob}(S_x = s | T_x = t, U_x = u)\text{Prob}(T_x = t | U_x = u) \\
&= \sum_{t=s}^{u} \text{Prob}(S_x = s | T_x = t)\text{Prob}(T_x = t | U_x = u)
\end{aligned} \tag{3.56}$$

where

$$\text{Prob}(t|u) = \binom{u}{t} h^t \, (1-h)^{u-t} \tag{3.57}$$

Derivation of Prob$(s|t)$

Given that Prob$(s|t)$ is unique, it follows that the state transition matrix can finally be obtained from Equations 3.52, 3.53 and 3.56, which in turn results in steady state probabilities $[Q_0, Q_1, \cdots, Q_M]$ by using the set a linear equations given by Equation 3.40.

In the following, we derive the conditional probability Prob$(R_x = r | T_x = t)$ as a parameter for some backoff scheme in which contention effects among these hosts are resolved in the frequency domain. In a generic time slot x, a poll of N channels is assumed available to serve t hosts. Each host can randomly select one of these channels at a time with a probability $1/N$. If there is exactly one host that selects a specific channel, the channel is said to be contention-free, otherwise a collision event happens. Since the randomness is spread across channels for a fixed time slot, we refer to this as frequency domain backoff scheme. We use the probability Prob$(R_x = r | T_x = r)$ of contention-free channel to evaluate such a multiaccess scheme.

Having t transmissions of u blocked hosts results in u^t different ways. We are interested in the number of contention free transmissions r. Let us define $\mathrm{Sum}(r|t,u)$ so that:

$$\mathrm{Prob}(r|t,u) = \frac{\mathrm{Sum}(r|t,u)}{u^t} \tag{3.58}$$

We use the following algorithm to obtain the argument $\mathrm{Sum}(r|t,u)$ [73]:

- For the case that $t \leq n$:

$$\mathrm{Sum}(r|t,u) = \begin{cases} 0 & \text{if } t < r \leq u \quad (3.59a) \\ \binom{r}{u} r! & \text{if} \quad r = t \quad (3.59b) \\ \binom{r}{t}\binom{r}{u} r! \Big[(u-r)^{(t-r)} - & (3.59c) \\ - \sum_{i=1}^{\min(t-r,u-r)} \mathrm{Sum}(i|t-r,u-r) \Big] & \text{if} \quad 0 \leq r \leq t \end{cases}$$

- For the case that $t > n$:

$$\mathrm{Sum}(r|t,u) = \begin{cases} 0 & \text{if } r = n \quad (3.60a) \\ \binom{r}{t}\binom{r}{u} r! \Big[(u-r)^{(t-r)} - & (3.60b) \\ - \sum_{i=1}^{\min(t-r,u-r)} \mathrm{Sum}(i|t-r,u-r) \Big] & \text{if} \quad 0 \leq r \leq n \end{cases}$$

3.8 Numerical Results

3.8.1 Effect of Controlled Single User Access

Parameters of the Channels Occupancy by PUs

Table 3.1 shows the numerical parameters corresponding to a N independent, two-state Continuous Time Markov Chain (CTMC)s model of spectrum occupancy by PUs. This model is an approximation of some traffic patterns of various existing wireless access applications [26]. The selection of parameter values is not crucial, but can be thought of as representing a general behavior of a wireless LAN.

Table 3.1: *System parameters*

Number of Parallel Channels N	3		
Duration of a time slot x [μs]	680		
Respective traffic loads:			
ρ_1, ρ_2, ρ_3	0.27,	0.44,	0.68
CTMC Approximations:			
λ_1^{-1}, λ_2^{-1}, λ_3^{-1} [ms]	9.13,	4.81,	1.50
μ_1^{-1}, μ_2^{-1}, μ_3^{-1} [ms]	3.45,	1.90,	1.01
Stationary probability distributions:			
Idle Probabilities $\eta(0)$	0.79,	0.70,	0.60
Busy Probabilities $\eta(1)$	0.21,	0.30,	0.40

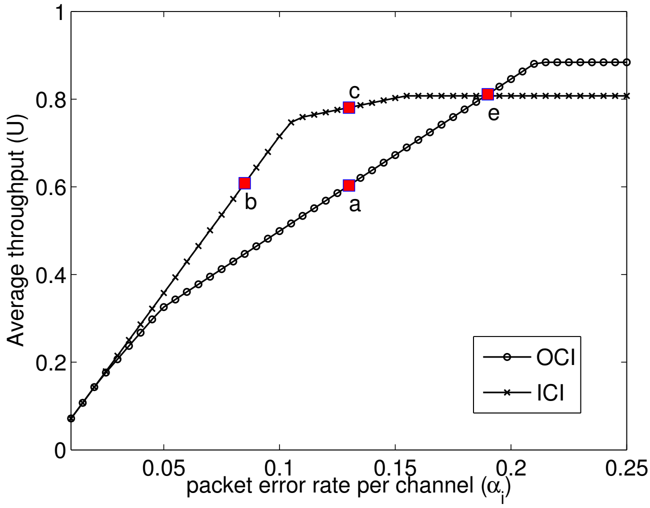

Figure 3.5: *Average throughput for SUs vs system loss constraint at PUs for a system with $N = 4$ channels*

Performance of Controlled CR Access Networks

Figure 3.5 shows the trade-off between the throughput $\mathcal{U}$ for SUs and the loss $\mathcal{V}$ at PUs caused by SUs on a given channel. As performance maximizing design options, we evaluate the maximum throughput $\mathcal{U}^*$ of the OCI and ICI under the condition that the loss $\mathcal{V}_1$ at PUs on say channel 1 is kept below a threshold value α_1. The threshold α_1 is then varied in the range [0.01, 0.25].

Observe that for both OCI and ICI access controls, a small tolerance in

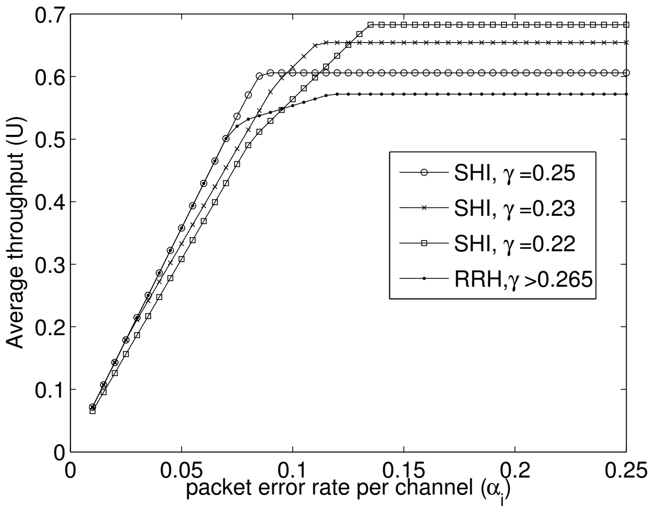

Figure 3.6: *Average throughput for SUs vs loss constraint at PUs for a system with $N = 4$ channels*

the loss α_1 at PUs significantly increases the throughput $\mathcal{U}$ for SUs. Large throughput performance for SUs is obtained only at the expense of small packet errors support by PUs. Let the mark e in the Figure 3.5 be the breaking point along a union curve made of the ICI curve until e and after e the curve of OCI. Every point along this curve is optimal in the sense that points above the optimal curve are inaccessible: no design can realize such $(\mathcal{U}, \mathcal{V})$ combinations. Whereas, points below the optimal curve correspond to sub-optimal design options.

For instance, the point marked b is throughput optimal, this means that 0.085 is the least tolerance in the loss α_1 on channel 1 that provides a secondary throughput as large as 0.60. In the Figure 3.5, the point marked a seen to the right of b is sub-optimal, in the sense that it represents the excessive loss 0.13 in α_1 that needs to be tolerable at PUs on channel 1 to achieve the 0.60 throughput for SUs. Likewise, the point marked c seen above a in the Figure 3.5 is optimal: 0.78 is the greatest throughput performance for SUs achievable with the loss tolerance 0.13 in α_1 at PUs on channel 1. So again the point marked a is sub-optimal.

Therefore, we say that any sub-optimal design option is either loss excessive for its throughput (increases loss caused on the channels while holding the throughput fixed) or throughput reduction for its loss (throughput is reduced while holding fixed loss caused on the channels).

Robustness of Controlled CR Access Networks

The trade-off of CR throughput performance is evaluated against robustness to uncertainty arising from the spectrum sensing deficiencies. As a robust satisfying design option, we evaluate the robustness of the SHI access control for any throughput requirement $\mathcal{U}_c$.

Consider the loss constraint α_1 at PUs on say channel 1. Let $\mathcal{U}_c$ be a satisfactory level of performance achievable under SHI access control. That is, the $(\alpha_1, \mathcal{U}_c)$ settings do not necessary fall on the optimal-design curve in Figure 3.5.

To be reliable (or feasible), $\mathcal{U}_c$ needs to be less that the throughput-optimum $\mathcal{U}^*$ for SUs obtainable with the least tolerance in loss α_1 at PUs on channel 1, namely $\mathcal{U}_c < \mathcal{U}^*$. This assures that the robustness of the SHI access control is positive. Observe that the value of $\mathcal{U}_c$ at which the robustness becomes zero falls on the optimal-design curve, that is, the value $\mathcal{U}_c$ for which the SHI throughput matches the maximum throughput ($\mathcal{U}_c = \mathcal{U}^*$).

For instance, to obtain a throughput with positive robustness, SHI should accept the total loss γ at the SU on all channels it uses. Figure 3.6 shows the throughput with positive robustness for different values of the total loss γ as a function of α_1. It is observed that, when the constraint γ on the total loss tolerable at SU is strict (i.e., γ is small), a positive robustness is obtained with excessive loss in α_1 at PUs on channel 1 for the throughput. However, when the constraint γ on the total loss tolerable at SU is relaxed (i.e., γ is large), a positive robustness is obtained with a reduction in the throughput for the loss α_1 at PUs on channel 1. Besides, Figure 3.6 shows the RRH access scheme a special case of SHI access scheme with γ large (i.e., $\gamma \geq 0.27$) and throughput reduced too far below the throughput optimum but with more improved robustness.

Therefore, for any access design controlling the opportunistic use of the multichannel capability in CR networks, a high performance requirement is accompanied with no immunity to uncertainty arising from spectrum sensing deficiencies. Immunity can be obtained only by giving up the requirement for maximum performance for both the PU system and the SU system.

3.8.2 Effect of Controlling the Multi-user Access Network

Numerical results are reported to evaluate the effect of controlling secondary access of the channels of PUs by using the suggested multi-user control access scheme.

Table 3.2: *Parameters for the multi-user access network*

Number of Parallel Channels	6
Duration of a time slot x [μs]	167
Traffic load:	
ρ	0.1639
CTMC Approximation:	
λ^{-1} [ms]	6.1
μ^{-1} [ms]	1
Probability distributions	
in idle and busy states	
$[\eta_0, \eta_1]$	$[0.8, 0.2]$

Parameters of the Channels Occupancy by PUs

The following numerical constants are assumed in Table 3.2, correspond to the occupancy of N parallel, statistically identical channels of PUs.

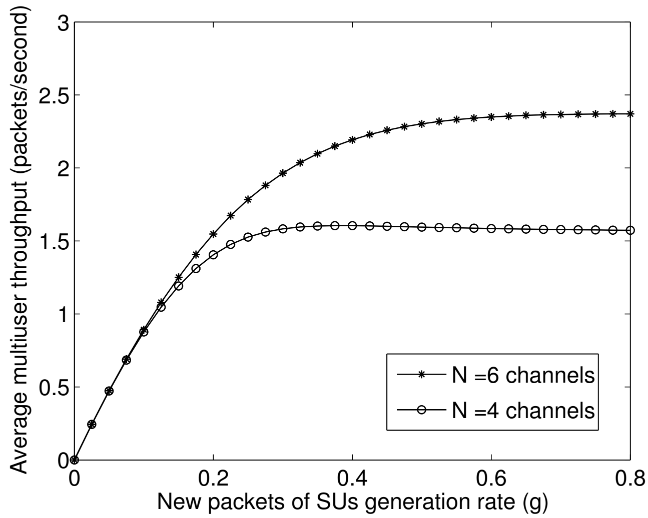

Figure 3.7: *Average number of contention-free transmissions when the number of users is* $M = 10$

Network Operation Points

Figure 3.7 shows the average number of contention-free transmissions as a function of new packet generation rate g. We assume a network in which the number of users is fixed to $M = 10$ operating with different numbers N of ideal channels, satisfying $N < M$. By ideal channels we mean channels with PUs completely absent. A technique optimized to resolve contentions effects among SUs is applied.

Observe that the number of contention-free transmissions increases sharply at light traffic load g, and then a relatively steady level of this number is maintained for heavy traffic load. When the new packet generation rate is small ($g < 0.125$), the network hosts spend more time in active mode,

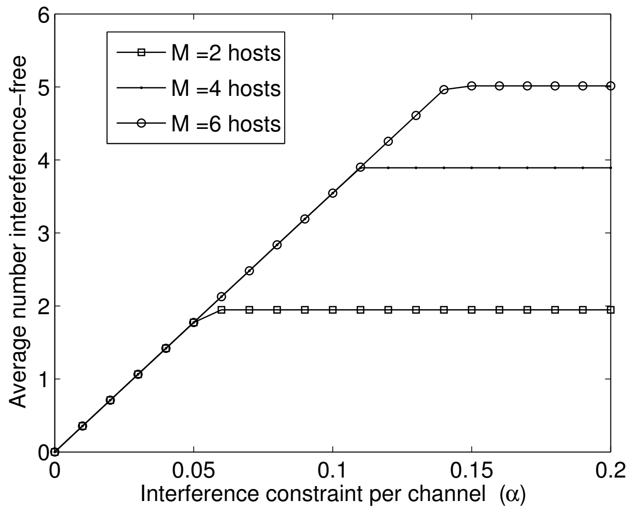

Figure 3.8: *Average number of interference-free transmissions when the number of channels is* $N = 6$

which in turn obviates contention effects among hosts. Thus the number of contention-free transmissions increases only with the traffic load g. As the new packet generation rate increases ($g > 0.125$), the network hosts spend more time in blocking mode. While this resolves contention effects it also keeps the throughput at relatively steady levels. Hence the number of contention-free transmissions can be improved only by increasing the number of channels.

Figure 3.8 shows the total rate of interference-free transmissions for SU hosts as a function of loss constraint α at PUs on a given channel. The network operates with a fixed number of time-varying channels $N = 6$ and it serves different numbers M of hosts, satisfying $M \leq N$. The new packet generation rate is $g = 1$ so that these host always have a packet to send. A

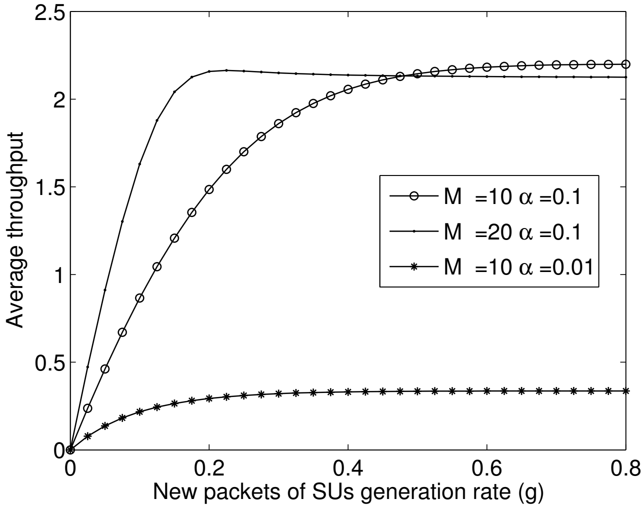

Figure 3.9: *Average multi-user system throughput when the number of channels is* $N = 6$

technique optimized to resolve interference effects caused to PUs is applied.

It is observed that a small tolerance in the loss α_1 at PUs on say channel 1 increases the total interference-free transmission rate for SU hosts by multiple. For instance, consider the loss tolerance 0.05 in α at PUs on a given channel. This represents the least tolerable loss at PUs for which the total rate in interference-free transmissions increases up to almost 2 packets per slot, when $M = 2$, 4 and 6 hosts. Alternatively, consider the loss tolerance ($\alpha = 0.15$) at PUs on a given channel. This represents the greatest level of increase of the total rate in interference-free transmissions (i.e., up to 5 packets per slot) provided by a loss tolerance of only ($\alpha = 0.15$), when $M = 2$, 4 and 6 hosts.

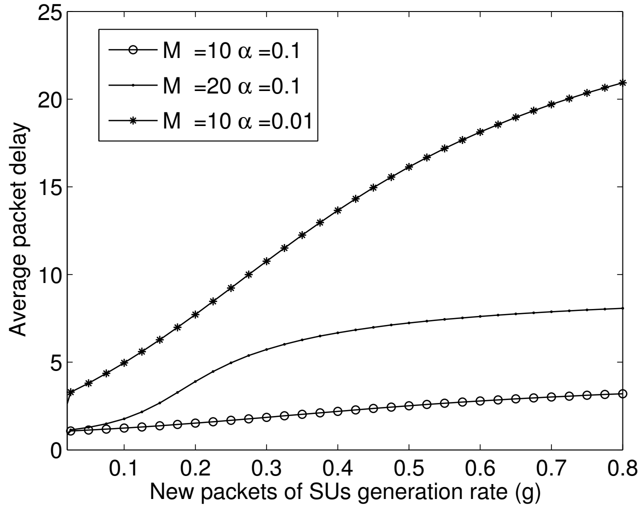

Figure 3.10: *Average packet delay when the number of channels is* $N = 6$

Throughput and Delay Performance

Figure 3.9 shows the throughput performance of SUs as a function of new packet generation rate g of the SUs. The multi-user access scheme operates over $N = 6$ channels of PUs with different parameters: number of SU hosts M and interference constraint α. It is observed that the throughput $\mathcal{S}$ increases at low packet generation rate ($g < 0.125$), and then it maintains a relative steady level as the packet generation rate increases ($g > 0.125$). Increasing the number of SU hosts at low traffic load removes the channel idleness, which in turn results in sharp increase of throughput. This reflects the effect of using a contention mitigation scheme optimized with the perfect knowledge of the system parameter. It is also observed that the system of SUs gives a large throughput for SUs as a result of having a larger loss tolerance at PUs. The effect of having a strict loss constraint at PUs (small α) laid on the transmissions of SUs causes the throughput for SU to drop too far below from the one available for secondary sharing.

Figure 3.10 shows the delay performance for different network parameters M and α. Also, the system of SUs gives a large packet delay (i.e., $\mathcal{D}$ becomes large) when the loss constraint is strict (i.e., small α). In this case SU hosts spend more time being blocked. But when this constraint is relaxed (i.e., α becomes large) the system gives an improved packet delay performance.

3.9 Summary

Efficient transmission strategies for SUs that use a slotted protocol as access scheme in the presence of parallel, dynamic channels already occupied by PUs traffic has been studied. In every time slot, a channel is selected and the probability of transmission in the selected channel is estimated by using local information about the quality of these channels. A SU transmits in a probabilistic manner in the selected channel based on this estimate. A Markovian decision model that embeds such a problem in its control actions is formulated. Optimal decision rules are found. The protocol capability in terms of throughput to SUs and packet error rate for PUs is expressed as reward and cost tradeoffs of the resulting Markovian decision processes. Numerical results on the performance of the suggested control access procedures are shown for nominal system parameters. An extension of the suggested strategy is included to support multi-user efficient transmission strategies.

Chapter 4

Spectrum Opportunity Sharing

Abstract

In this chapter, the third DSA model considered in the dissertation is presented. We study admission control of SUs to achieve a certain degree of QoS. A multiple dimension Markov chain whose size increases quickly with the number of channels of PUs is assumed to characterize the known system. In the presence of uncertainty associated with the system size reduction, we consider a one dimension Markov chain approximation to characterize the uncertain system. Appropriate analytical models are developed and used to derive the performance characteristics of the secondary admission strategies. Numerical results are reported and show a robust capacity in terms of blocking probability, dropping probability and delay performance as functions of spectrum utilization by PUs.

4.1 Introduction

4.1.1 Overview

The management of interference is one of the critical challenges towards practical realization of DSA networks [74]. The DSA approach consists of dividing users of a system into a multi-tiered hierarchy with PUs and SUs [75]. PUs are given priority of interrupting a SU already in service. SUs on the other hand are cognitive radio capable. This means that a SU with an ongoing communication must perform spectrum sensing and it uses the local information about the channels, obtained by sensing, to switch to the channel where a PU is detected [76] as absent. Such a flexibility of switching frequency bands is referred to as *spectrum mobility* or *spectrum handover*. In response to the dynamic nature of the channels occupied by PUs, a SU is required to frequently execute spectrum sensing, which in turn causes frequent pauses of any ongoing communications [77].

To combat the channel fluctuations, the principle of Orthogonal Frequency Division Multiplexing (OFDM) divides the total channel bandwidth into parallel sub-channels to be shared in an opportunistic manner. The transmissions on these sub-channels are expected to interfere with each other but, since the transmissions are made orthogonal to each other, there is no mutual interference [78]. Such a flexibility in the physical layer characteristics can by deployed in DSA networks to support adaptive resource allocation by using *sub-channel assignment* schemes.

By dynamically allocating the sub-channels to SUs according to their channel conditions, one can ensure that each sub-channel is allocated to SUs with a relatively better channel condition (i.e., PU absent). By doing so, the spectral efficiency can be effectively improved while the requirement of minimum interference is met.

4.1.2 Motivation

In the following, the focus is laid on the study of the capacity of secondary admission control for SUs and interference guaranteed to PUs from a traffic theory viewpoint. We assume that the spectrum consists of N parallel channels, each of which being a channel for a PU to access. Each channel of PUs is divided into K sub-channels for opportunistic communications by SUs. The primary network follows a predefined pattern to assign channels to a finite population M_p of PUs. The secondary network senses and assigns the sub-channels deemed idle to a finite population M_c of SUs. The secondary network is aware of the rule followed by the primary network and uses this knowledge as local information state for the protection of PUs.

When a channel of PUs is deemed idle, the corresponding sub-channels are open for secondary communications. If a PU host arrives in the case the channel is deemed busy, the secondary communications on each of the sub-channels associated to this particular channel are halted. An interrupted communication of a SU host is allowed to be handed over to another sub-channel, and it may continue if this particular sub-channel is seen as having an idle outcome. Otherwise (i.e., in case there is no sub-channel deemed idle) a halted communication of a SU has to be queued and waiting to continue with its transmission later on.

Consequently, the average holding time of a sub-channel by a secondary service becomes less than the average service duration it requires until completion. In addition to the new service attempts, handoff attempts are generated. If all sub-channels are in use, new setup request of service are denied access resulting in blocking effects. For the services whose transmissions has been halted, random delay effects are introduced. Some services may be badly affected in which case handoff blocking effects occur. Some fraction of handoff attempts will be unsuccessful at any time in which case force termination or dropping out effects happen.

Traffic theory has proved to be a feasible, alternative tool to assess these effects in the context of CR networks [42, 43]. By traffic theory we mean *traffic-performance* formulas like the Erlang formula linking traffic intensity offered to specific capacity with the attainable performance. Most commonly used modeling techniques involve multiple dimension Markov chains. A Finite State Machine (FSM) is first generated, which in turn is converted into a CTMC, where state transitions are assigned transition rates, so that it can be solved using numerical methods [79]. However, the underlying CTMC has the size of its state space growing much faster. This also applies to storage and solution models, that is, the approximate data structures for sparse matrix and sparsity preserving solution methods must be used. Due to energy or hardware constraints model reduction techniques are equally important, but at a cost of increased uncertainty about the system model.

4.1.3 Contributions

The main contributions of this chapter are summarized as follows:

- A multiple dimension Markov chain whose size increases quickly with the number of channels of PUs is assumed to characterize the known system.
- In the presence of uncertainty associated with the system size reduction, we consider a one dimension Markov chain approximation to

characterize the uncertain system.

- Appropriate analytical models are developed and used to derive the performance characteristics of the secondary admission strategies.

- Numerical results are reported and show a robust capacity in terms of blocking probability, dropping probability and delay performance as functions of spectrum utilization by PUs.

4.2 System Model

In this section, the network model, the traffic model, the traffic performance measures and the sub-channel assignment schemes for different systems are presented.

4.2.1 Network Model

The basic system model assumes a wireless system with a finite population M_p of PUs and a finite population M_c of SUs. The wireless spectrum consists of N channels available for transmissions of PUs. Each channel of PUs in the spectrum is divided into K sub-channels for SUs. SUs communicate via sub-channels to a multiple access receiver (i.e., a Base Station (BS)). Here we broadly use the term channel as a unit of resource, where the physical meaning is dependent on the specific technological implementation of the radio interface.

We assume that channels available for primary network operation are numbered according to the order in which they are assigned by the primary network. This means that, to setup a new PU request the system searches the channels in increasing order starting from the lowest index say 1 until a channel available is found to serve the new setup request.

The secondary network operates under the control of a scheduler at the BS. This monitors the occupancy state information over the channels occupied by PUs using spectrum sensing−which channels have PU presence and which ones have PU absence. Based on this information, new SU setup requests are admitted according the sub-channels corresponding to the available channels, starting from low channel indexes.

4.2.2 Traffic Model

We assume that the occupancy of the N channels by PUs follows an alternating idle-busy process each, according to the idle and busy PU traffic patterns. When a PU host is absent (i.e., a channel seen as idle), then the time it takes for that particular PU host to arrive is a random variable with exponential distribution with mean $1/\lambda$. All PUs act independently, so that if there are i PUs that are actually occupying their respective channels, then exactly $(M_p - i)$ PU hosts are absent, in which case the effective arrival rate is $(M_p - i)\lambda$. Also, when a PU host occupies a channel (i.e., a channel seen as busy), then the time it takes for that particular host to complete its transmissions is a random variable exponentially distributed with mean $1/\mu$. This is well known as a $M/M/N//M_p$ queue.

Similarly, the traffic sources of SU hosts emit each a on-off traffic pattern that can be modeled using a $M/M/NK//M_c$ queuing system with exponentially distributed on and off periods. The mean of the on and off periods are $1/\alpha$ and $1/\beta$, respectively.

To analyze the queuing system characteristic of spectrum occupancy of PUs and traffic sources of SUs, we enumerate possible number of busy channels by $i = 0,\ 1,\ 2, \cdots,\ N$ and possible number of sub-channels used by SU traffic by $j = 0,\ 1,\ 2, \cdots,\ N \times K$. We denote by $B(\mathbf{x}) = (i, j)$ the system state vector, and we use $B(\mathbf{x})$ as channel state information in state $\mathbf{x}$. $B(\mathbf{x}) = iK + j$ and it represents the amount of resource occupied in state $\mathbf{x}$. In this case, the dynamic behavior in time of such a system can be modeled as a multidimensional Markov process whose set of feasible states is given by:

$$\mathcal{S} = \left\{ \mathbf{x} = (i, j) \ :\ B(\mathbf{x}) \leq NK \right\} \tag{4.1}$$

4.2.3 One Dimension Markov Process Approximation

Since the primary network is oblivious of the existence of the secondary network, when they use the channels, the performance analysis of the primary users can be independently evaluated as a set of states i, environmental specifications for the system, each of which being a complete configuration to operate SUs. The underlying stochastic process is a homogeneous CTMC with state space $\{0,\ 1,\ 2, \cdots,\ N\}$ and it is completely specified by the steady-state probabilities $\{\pi_0,\ \pi_1,\ \pi_2, \cdots,\ \pi_N\}$ across the states i. By attaching a suitable chosen set of weights $\{r_i\}$ across the states of the CTMC model of spectrum occupancy by PUs, one can get most traffic performance measures characteristic of PU traffic and expressed in form of weighted average as follows:

$$R = \sum_i r_i \pi_i \tag{4.2}$$

Based on this information, the performance analysis of SUs can be approximated as a one dimension Markov process conditional on i. The index

i is the state of spectrum occupancy environment to the system where exactly i channels are used by PUs communications. On the other hand, the index j defines a state system object where exactly j sub-channels are being used by SU traffics. The underlying stochastic process is a homogeneous CTMC conditional on i with state space $\{0, 1, 2, \cdots, M_c\}$ and steady-state probabilities $\{\pi_0, \pi_1, \pi_2, \cdots, \pi_{M_c}\}$. Similarly, by attaching a suitable chosen set of weights $\{w_j\}$, one can get most traffic performances characteristic of the system in form of weighted average that depends on i:

$$W_i = \sum_j w_j \pi_j, \qquad i = 0, 1, \cdots, N-1 \tag{4.3}$$

we ignore the case where $i = N$, since in this state the system has complete outage, i.e., $W_N = 0$.

W_i can be thought of as being the traffic performance changes associated with i. In the presence of system model uncertainty, it follows that a robust traffic capacity of such a system can be obtained as:

$$S = \sum_i W_i \pi_i \tag{4.4}$$

4.2.4 Performance Measures

The probability of blocking of new SU connection request, also referred to as *new call blocking*, is of particular interest just as it is in traditional traffic theory. The new call blocking probability P_b is defined to be the probability that a new connection request of SU is denied access to a sub-channel, either because the channels of PUs are all occupied or because all sub-channels are also occupied by other SUs.

A distinctive feature of CR radio is inter handoff, also called spectrum mobility, where PUs occupy their respective channels oblivious to the existence of SUs. A PU is given priority to interrupt a SU in service and takes the channel. At the same time, an interrupted SU needs to opportunistically switch to another sub-channel. The probability that a sub-channel reallocation fails is another important measure, and we refer to it as the *handoff* blocking P_h.

The probability that a SU call in service is dropped out is a third measure of traffic performance. This may arise following the fact that during a SU call a number of PU arrivals happen. A SU may see its call handed over as many times as the number of interruptions before reaching the handoff where it is blocked. The dropout probability P_d is therefore the probability that a call is dropped out at any point during service. Whereas the handoff blocking probability P_h is simply a measure of sub-channel switching failure.

With the assumption of an infinite buffer at the BS, when all sub-channels are occupied, any arriving SU call is queued locally until such sub-channels become available again. In this case, the response time of SUs in such a system is also an important as traffic performance measure. Since a new arriving call of SU must wait while those already in the queue, including the one in service, are served, we define the response time D to be the time interval from the moment of a call at the system till the moment of completion. This includes the time needed to complete the services of earlier calls plus the time needed to serve the entering call.

We are now in position to specify the mathematical analysis required to determine the traffic capacity measures. These quantities depend on the scheme used to manage the sub-channel assignment.

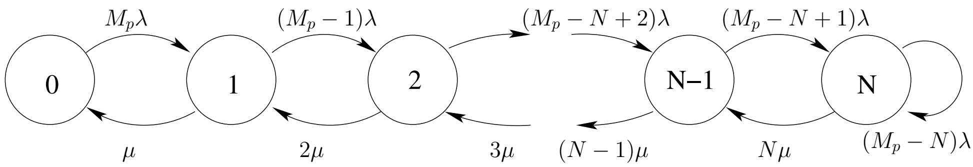

Figure 4.1: *The state diagram of a finite population occupancy of the channels by PUs*

4.3 Sub-channel Assignment Scheme

In this section, the channels state information is given by the channel capacity C of PUs as a measure of protection guarantee to PUs. Given the knowledge about the channels state information, two systems are studied based on whether the assumption of an infinite buffer is supported or not by the deployed admission strategy.

4.3.1 Channels State Information

The channels state information consists of a loss system with parallel channels modeling the spectrum occupancy by PUs. In particular, we are interested in the knowledge of the traffic throughput C of these channels. We denote by $i = 0, 1, 2, \cdots, N$ the number of channels used by PUs to represent the channels state information. The state diagram is shown in Figure 4.1. The expressions for steady state probabilities $\{\pi_0, \pi_1, \cdots, \pi_N\}$ are obtained by using the following steady state balance equations:

$$(M_p - i + 1)\,\lambda\,\pi_i = i\,\mu\,\pi_{i-i} \qquad 0 \le i \le N \tag{4.5}$$

so that

$$\begin{aligned}\pi_i &= \frac{(M_p - i + 1)\,\lambda}{n\,\mu}\pi_{i-1} \qquad (4.6)\\ &= \frac{(M_p - i + 1)\,(M_p - i + 2)\cdots M_p}{i\,(i-1)\cdots\,1)}\left(\frac{\lambda}{\mu}\right)^i \pi_0\\ &= \frac{M_p!}{(M_p - i)!\,i!}\left(\frac{\lambda}{\mu}\right)^i \pi_0\\ &= \binom{M_p}{i}\left(\frac{\lambda}{\mu}\right)^i \pi_0, \qquad 0 \le i \le N\end{aligned}$$

By normalization, we have

$$\pi_i = \frac{\binom{M_p}{i}\left(\frac{\lambda}{\mu}\right)^i}{\sum_{k=0}^{N}\binom{M_p}{k}\left(\frac{\lambda}{\mu}\right)^k}, \qquad 0 \le i \le N \qquad (4.7)$$

By attaching to states i a weight of

$$r_i = (M_p - i)\,\frac{\lambda}{\mu} \qquad i = 0,\,1,\,2,\cdots,\,N \qquad (4.8)$$

the total load offered by PUs can be obtained as

$$\begin{aligned}\hat{\lambda} &= \sum_{i=0}^{N}(M_p - i)\,\frac{\lambda}{\mu}\pi_i \qquad (4.9)\\ &= \frac{M_p \sum_{i=0}^{N}\binom{M_p-1}{i}\left(\frac{\lambda}{\mu}\right)^i}{\sum_{k=0}^{N}\binom{M_p}{k}\left(\frac{\lambda}{\mu}\right)^k}\end{aligned}$$

Similarly, attaching to states i weights of

$$r_i = i \qquad i = 0,\ 1,\ 2, \cdots, N \tag{4.10}$$

one can write the total traffic of PUs entitled to protection as follows:

$$\begin{aligned} C &= \sum_{i=0}^{N} i\pi_i \qquad (4.11) \\ &= \frac{\sum\limits_{i=0}^{N} \binom{M_p}{i}\, i \left(\frac{\lambda}{\mu}\right)^i}{\sum\limits_{k=0}^{N} \binom{M_p}{k} \left(\frac{\lambda}{\mu}\right)^k} \end{aligned}$$

Equivalently, the traffic capacity C is given by

$$C = \hat{\lambda}(1 - B) \tag{4.12}$$

where B is the blocking probability of PU calls being served by theses channels and it is given by

$$\begin{aligned} B &= 1 - \frac{C}{\hat{\lambda}} \\ &= \frac{M_p \binom{M_p - 1}{N} \left(\frac{\lambda}{\mu}\right)^{N+1}}{\hat{\lambda} \sum_{k=0}^{N} \binom{M_p}{k} \left(\frac{\lambda}{\mu}\right)^{k}} \\ &= \frac{\binom{M_p - 1}{N} \left(\frac{\lambda}{\mu}\right)^{N}}{\sum_{k=0}^{N} \binom{M_p - 1}{k} \left(\frac{\lambda}{\mu}\right)^{k}} \end{aligned} \tag{4.13}$$

4.3.2 Sub-Channel Assignment without Buffer

In this case, the system is characterized by a secondary admission strategy employing a random sub-channel allocation scheme without buffering. This means that all SU requests, new and handoff, are accepted and served from a poll of available sub-channels associated with the idle channels. When a SU call gets a sub-channel, it keeps it until it is either completed satisfactory or interrupted by a PU arrival. When a SU call is completed, the sub-channel becomes available to serve another SU call.

Thus, when a new SU setup request arrives and finds the system in state $\mathbf{x}$, an admission decision is made based on the following rule:

$$\text{setup decision in } \mathbf{x} = \begin{cases} \text{accepted} & \text{if} \quad (NK - B(\mathbf{x})) \geq 0 \quad \text{(4.14a)} \\ \text{blocked} & \text{if} \quad (NK - B(\mathbf{x})) < 0 \quad \text{(4.14b)} \end{cases}$$

A PU arrival in state $\mathbf{x}$ causes all SUs in this particular state to halt their ongoing communications in order to vacate the channel where the PU arrival is detected. Based on handoffs demand from state $\mathbf{x}$, two situation arises. If the system state is such that the condition $(NK - B(\mathbf{x})) \leq K$ is satisfied, SU handoff setup requests are accepted. Thus, the halted communications of SUs continue to other sub-channels, which are guaranteed to exist given that the above condition is met. Otherwise, for the case $(NK - B(\mathbf{x})) > K$, the halted communications of SUs are forced to terminate before completion.

Let $Q(x)$ denote the number of handoffs blocking in state $\mathbf{x}$, $q_{\mathbf{x},\mathbf{y}}$ the transition rate from $\mathbf{x}$ to $\mathbf{y}$. $\mathbf{e}_i$ and $\mathbf{e}_j$ are unit vectors associated with the system state $\mathbf{x} = (i, j)$ such that $\mathbf{e}_i = (1,0)$ and $\mathbf{e}_j = (0,1)$, respectively.

$$Q(\mathbf{x}) = \min\{s \in \mathbb{N} : B(\mathbf{x} + \mathbf{e}_i - s\mathbf{e}_j) \leq KN\} \tag{4.15}$$

$$q_{\mathbf{x},\mathbf{y}} = \begin{cases} (M_p - i)\lambda & \text{if} \quad \mathbf{x} + \mathbf{e}_i - Q(\mathbf{x})\mathbf{e}_j & \text{(4.16a)} \\ (M_c - j)\alpha & \text{if} \quad \mathbf{x} + \mathbf{e}_j & \text{(4.16b)} \\ i\mu & \text{if} \quad \mathbf{x} - \mathbf{e}_i & \text{(4.16c)} \\ j\beta & \text{if} \quad \mathbf{x} - \mathbf{e}_j & \text{(4.16d)} \end{cases}$$

In the absence of system model uncertainty, the total balance equations can be expressed as follows:

$$\sum_y q_{\mathbf{x},\mathbf{y}} \pi(\mathbf{x}) = \sum_y q_{\mathbf{y},\mathbf{x}} \pi(\mathbf{y}) \qquad \forall \mathbf{x} \tag{4.17}$$

and

$$\sum_{\mathbf{x}} \pi(\mathbf{x}) = 1 \tag{4.18}$$

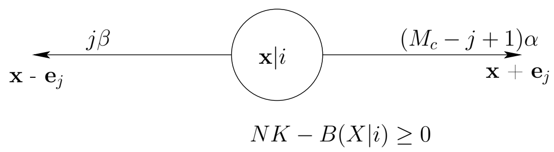

Figure 4.2: *The state diagram for a one dimension CTMC without buffering*

However, since we are interested to study the secondary admission in the presence of system model uncertainty, we resort to one dimension Markovian approximation methods. To do this, we consider the system state $\mathbf{x}|i$ conditional to the number i of channels occupied by PUs. We denote possible number of SUs in conditional system state $\mathbf{x}|i$ by $j = 0,\ 1, 2, \cdots,\ M_c$. Also, the conditional number of resources is given by $B(\mathbf{x}|i) = iK + j$. Thus, the state diagram is reduced to a one dimension CTMC as shown in Figure 4.2. The expressions for steady state probabilities $\{\pi(\mathbf{x}|i)\}$ are obtained by using the following steady state balance equations:

$$(M_c - j + 1)\,\alpha\,\pi(\mathbf{y}|i) \,= j\beta\,\pi(\mathbf{x}|i), \quad \mathbf{y} = \mathbf{x} - \mathbf{e}_j \tag{4.19}$$

Following Equation 4.7, we can write the steady state probabilities as

$$\pi(\mathbf{x}|i) = \frac{\binom{M_c}{j}\left(\frac{\alpha}{\beta}\right)^j}{\sum\limits_{k=0}^{N}\binom{M_c}{k}\left(\frac{\alpha}{\beta}\right)^k}, \qquad 0 \leq B(\mathbf{x}|i) \leq NK \tag{4.20}$$

for all $i = 0,\ 1, \cdots,\ N-1$. Following Equation 4.27 and given that a call blocking happens when all channels and sub-channels are occupied, we obtain the blocking probability $P_b(i)$ by attacking to the system states $\mathbf{x}|i$ the following weights:

$$w(\mathbf{x}|i) = \begin{cases} 0 & \text{if} \quad B(\mathbf{x}|i) < KN \quad \text{(4.21a)} \\ 1 & \text{if} \quad B(\mathbf{x}|i) = KN \quad \text{(4.21b)} \end{cases}$$

so that we can write

$$P_b(i) = \sum_{\mathbf{x}} w(\mathbf{x}|i)\pi(\mathbf{x}|i) \tag{4.22}$$
$$= \frac{\binom{M_c-1}{NK-iK}\left(\frac{\alpha}{\beta}\right)^{(NK-iK)}}{\sum\limits_{k=0}^{NK-iK}\binom{M_c-1}{k}\left(\frac{\alpha}{\beta}\right)^{k}}$$

for all $i = 0,\ 1, \cdots, N-1$.

The blocking probability $P_b(i)$ can be thought of as a weight in the channel state i so that the total blocking probability $\mathcal{P}_b$ to the system can be written as

$$\mathcal{P}_b = \sum_i P_b(i)\pi_i \tag{4.23}$$
$$= \sum_{i=0}^{N-1} \frac{\binom{M_c-1}{NK-iK}\left(\frac{\alpha}{\beta}\right)^{(NK-iK)}}{\sum\limits_{k=0}^{NK-iK}\binom{M_c-1}{k}\left(\frac{\alpha}{\beta}\right)^{k}}\pi_i$$

Since a PU arrival that finds the system in state $\mathbf{x}|i$ causes blocking of SU handoff setup requests with probability $\frac{Q(\mathbf{x})}{j}$, thus moving the system to state $\mathbf{y}|i = \mathbf{x}|i + \mathbf{e}_i - Q\mathbf{e}_j$. We obtain the handoff blocking probability $P_h(i)$ by attacking the following weights in state $\mathbf{x}|i$ as

$$w(\mathbf{x}|i) = \frac{Q(\mathbf{x})}{j}, \quad Q = 0,\ 1, \cdots, N \tag{4.24}$$

so that we can write

$$
\begin{aligned}
P_h(i) &= \sum_x w(\mathbf{x}|i)\pi(\mathbf{x}|i) \\
&= \sum_{j=1}^{KN-iK} \frac{Q(\mathbf{x})}{j}\pi(\mathbf{x}|i) \qquad i = 0,\ 1, \cdots, N-1
\end{aligned}
\tag{4.25}
$$

for each $i = 0,\ 1, \cdots, N-1$. Where $Q = 0$ if $B(\mathbf{x}|i + \mathbf{e}_i) = K$, also $Q = K$ if $B(\mathbf{x}|i + \mathbf{e}_i - K\mathbf{e}_j) = NK$.

When a SU call is assigned a sub-channel, this call will hold the sub-channel. Because of a PU return to occupies a channel, which can occur at any time, this call can be drouped out at any time as well. This happens if the following events occur: (i) a SU succeeds to handover its call in each of the $(l-1)$ handoff attempts that it requires, (ii) a SU handoff call is blocked on the lth attempt. Given that the handoff blocking probability P_h and the probability P_r of handoff demand are uniquely especified, one can obtain the dropout probability P_d.

The probability P_r that a SU call in service requires at least one handoff before completion is the same as the probability that the service time of this call is longer that the residual idle period of PU on a channel. Thanks to the memoryless property, the residual idle period of PU traffic on a channel is the same as the generic idle period of PU traffic. Hence, at any moment that a SU call is in service, the time X it takes to complete a SU call and the time Y it takes until a PU return to take a channel, are independent exponential random variables with rates β and λ respectively, each competing to occur before the other. It follows that the probability P_r can be expressed as

$$
\begin{aligned}
P_r &= \text{Prob}(Y > X) \\
&= \int_0^\infty [1 - F_Y(t)]\, f_X(t)\, dt \\
&= \int_0 e^{-\beta t}\, f_X(t)\, dt \\
&= \int_0^\infty \lambda e^{-(\beta+\lambda)t}\, dt \\
&= \frac{\lambda}{\beta + \lambda}
\end{aligned} \tag{4.26}
$$

Therefore, we obtain the droupout probability as

$$
\begin{aligned}
P_d(i) &= \sum_{l=1}^{\infty} \left[\big(1 - P_h(i)\big)^{l-1} P_r^{l-1} \right] P_r P_h(i) \\
&= \frac{P_h(i)}{1 - P_h(i)} \sum_{l=1}^{\infty} \left[\big(1 - P_h(i)\big)^{l} P_r^{l} \right] \\
&= \frac{P_r P_h(i)}{1 - \big(1 - P_h(i)\big) P_r}
\end{aligned} \tag{4.27}
$$

Similarly, we consider $P_d(i)$ as weights over the channels state i, so we obtain the total droupout probability as a weighted average:

$$
\mathcal{P}_d = \sum_{i=0}^{N-1} \frac{P_r P_h(i)}{1 - \big(1 - P_h(i)\big) P_r} \pi_i \tag{4.28}
$$

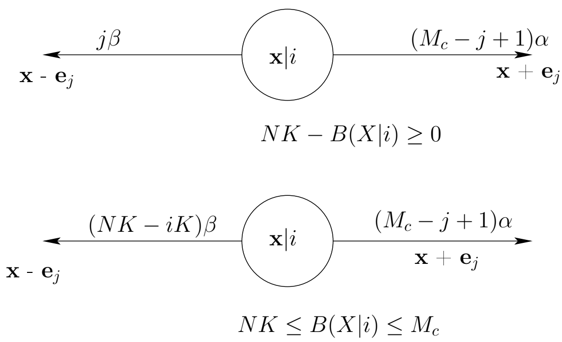

Figure 4.3: *The state diagram of a one dimension CTMC with buffering*

4.3.3 Sub-Channel Assignment Scheme with Buffering

The system is characterized in this case by a secondary admission strategy employing a random sub-channel allocation scheme with buffering. This means that all SU requests, new and handoff, are supplied to a buffer and then drained according to the available sub-channels associated with the idle channels each. When a SU call gets a sub-channel it keeps it until either it is satisfactory completed or queued waiting for a sub-channel. When a SU call is completed, the sub-channel becomes available to serve another SU call.

Thus, when a SU new setup request arrives and finds the system in state $\mathbf{x}$, an admission decision is made based on the following rule:

$$\text{setup decision in } \mathbf{x} = \begin{cases} \text{accepted} & \text{if} \quad (NK - B(\mathbf{x})) \geq 0 \quad (4.29a) \\ \text{queued} & \text{if} \quad (NK - B(\mathbf{x})) < 0 \quad (4.29b) \end{cases}$$

Since we are studying secondary admission control in the presence of system model uncertainty, we consider the system state $\mathbf{x}|i$ conditional to the number i of channels occupied by PUs. We denote possible number of SUs in conditional sytem state $\mathbf{x}|i$ by $j = 0,\ 1, 2, \cdots, M_c$. Also, the conditional number of resources is given by $B(\mathbf{x}|i) = iK + j$.

Thus, the state diagram of the reduced one dimension CTMC shown in Figure 4.3. We obtain the steady state probabilities $\{\pi(\mathbf{x}|i)\}$ by using the following steady state balance equations:

$$(M_c - j + 1)\ \alpha\ \pi(\mathbf{y}|i) = j\ \beta\ \pi(\mathbf{x}|i) \qquad \mathbf{y} = \mathbf{x} - \mathbf{e}_j \tag{4.30}$$

and

$$(M_c - j + 1)\ \alpha\ \pi(\mathbf{y}|i) = (NK - iK)\ \beta\ \pi(\mathbf{x}|i) \qquad \mathbf{y} = \mathbf{x} - \mathbf{e}_j \tag{4.31}$$

for each $i = 0,\ 1, \cdots, N-1$

The expressions for steady state probabilities are hence given by

$$\begin{aligned} \pi(\mathbf{x}|i) &= \frac{M_c!}{(M_c - j)!\ j!} \left(\frac{\alpha}{\beta}\right)^j \pi(i,0) \\ &= \binom{M_c}{j} \left(\frac{\alpha}{\beta}\right)^j \pi(i,0) \qquad B(\mathbf{x}|i) \le NK \end{aligned} \tag{4.32}$$

$$\begin{aligned} \pi(\mathbf{x}|i) &= \frac{M_c!}{(M_c - j)!\ (NK - iK)^{(j-K(N-i))}\ (NK - iK)!} \left(\frac{\alpha}{\beta}\right)^j \pi(i,0) \\ &= \binom{M_c}{j} \frac{j!}{(NK - iK)^{(j-K(N-i))}\ (NK - iK)!} \left(\frac{\alpha}{\beta}\right)^j \pi(i,0) \end{aligned} \tag{4.33}$$

if $NK \leq B(\mathbf{x}|i) \leq M_c$

By normalization we have

$$\pi(i,0) = \left[\sum_{j=0}^{NK-iK-1} \binom{M_c}{j} \left(\frac{\alpha}{\beta}\right)^j \right. \tag{4.34}$$
$$\left. + \sum_{j=NK-iK}^{M_c} \binom{M_c}{j} \frac{j!}{(NK-iK)^{(j-K(N-i))}\,(NK-iK)!} \left(\frac{\alpha}{\beta}\right)^j \right]^{-1}$$

To obtain the average number $L(i)$ of SU calls in the system, we select the weights in the respective states $\mathbf{x}|i$ to be

$$w(\mathbf{x}|i) = j \tag{4.35}$$

so that L can be expressed in the form of weighted average:

$$L(i) = \sum_{\mathbf{x}} w(\mathbf{x}|i)\ \pi(\mathbf{x}|i) \tag{4.36}$$
$$= \sum_{j=0}^{M_c} j\ \pi(\mathbf{x}|i)$$

To obtain the average arrival rate $\hat{\alpha}$ of SUs calls, we choose the weights of the system state $\mathbf{x}|i$ as

$$w(\mathbf{x}|i) = \alpha(M_c - j) \tag{4.37}$$

so that

$$\begin{aligned}\hat{\alpha}(i) &= \sum_{\mathbf{x}} w(\mathbf{x}|i)\ \pi(\mathbf{x}|i) \\ &= \alpha(M_c - j)\ \pi(\mathbf{x}|i)\end{aligned} \tag{4.38}$$

By Little's formula, the mean response time D of the system is given for $i = 0,\ 1, \cdots, N-1$, by

$$\begin{aligned}D(i) &= \frac{L(i)}{\hat{\alpha}(i)} \\ &= \frac{L(i)}{\alpha(M_c - L(i))}\end{aligned} \tag{4.39}$$

where $D(i)$ are considered as weights attached to the channel state i so that the total response time can be written as

$$\mathcal{D} = \sum_{i=0}^{N} \frac{L(i)}{\alpha(M_c - L(i))} \pi_i \tag{4.40}$$

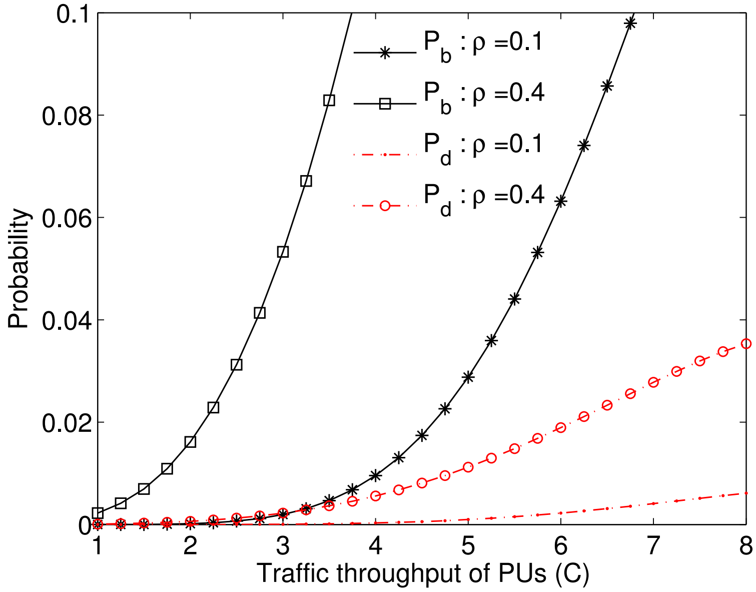

Figure 4.4: *SU blocking probability and SU dropout probability with* $K = 2$, $N = 10$ $M_p = 20$, $M_c = 30$

4.4 Numerical Results

Figure 4.4 shows the blocking probability $\mathcal{P}_b$ and dropout probability $\mathcal{P}_d$ of SUs as functions of traffic capacity C of PUs (Erlangs). The number of sub-channels per channel is fixed to $K = 2$ while the traffic intensity ρ of SUs takes different values in the range $[0.1, 0.4]$. A wirelss network with a population size of PUs and SUs fixed to $M_p = 20$ and $M_c = 30$, respectively, is considered. PUs share $N = 10$ channels in total. SUs share $N \times K = 20$ opportunistic channels in total. The traffic capacity C is used to measure the traffic performance of PUs from which we derive the channels state

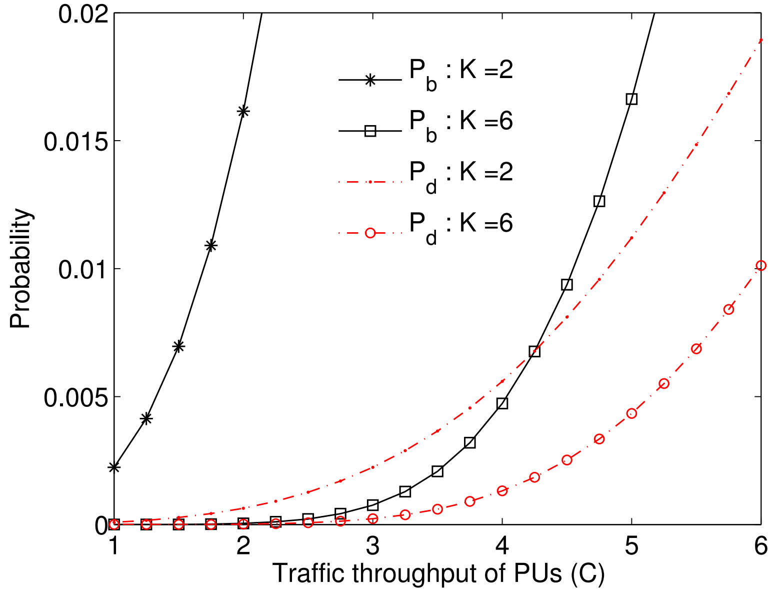

Figure 4.5: *SU blocking probability and SU dropout probability with* $\rho = 0.4$, $N = 10$, $M_p = 20$, $M_c = 30$

information. This knowledge is then used to operate the system of SUs.

The effects of PUs having higher priority over SUs is reflected in different levels of C. It is observed that, at light SU traffic load ($\rho = 0.1$), $\mathcal{P}_b$ and $\mathcal{P}_d$ can be maintained at relatively small levels, several orders of magnitude below, before the measured capacity C is becomes larger than 6 Erlangs.

Also, it is observed that, when ρ is small (0.1), the loss (increase) in $\mathcal{P}_d$ is far too much (more than 25 times) below $\mathcal{P}_b$. Hence, at small ρ the droupout effects are rendered void. On the other hand, when ρ increases up to 0.4, $\mathcal{P}_b$ and $\mathcal{P}_d$ increase by orders of magnitude.

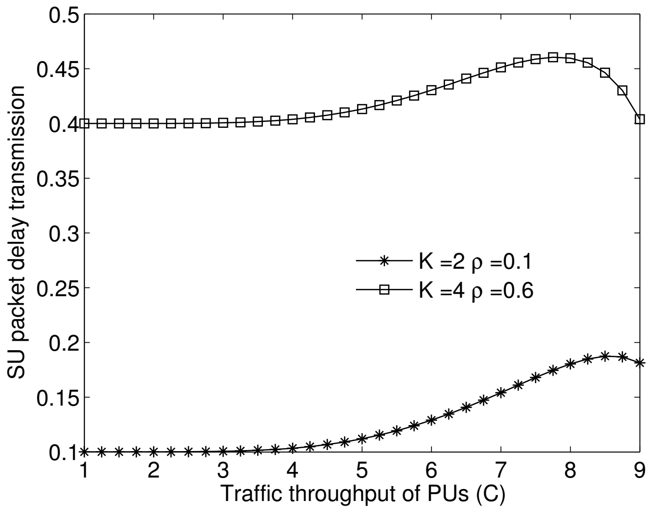

Figure 4.6: *SU packet delay transmission with* $N = 10$, $M_p = 20$, $M_c = 30$

The improvements in the blocking probability $\mathcal{P}_b$ and the dropout probability $\mathcal{P}_d$ are shown in Figure 4.5. The traffic intensity of SUs is fixed to $\rho = 0.4$ while the number of sub-channels K supported per channel takes various interger values in range $[2,8]$. When the number of sub-channels is increased up to 8, that is each PU channel can be divided up to 8 parallel sub-channels, $\mathcal{P}_b$ is decreased by orders of magnitude with only small to moderate decrease in $\mathcal{P}_d$. Hence, by increasing the number of sub-channels, the blocking effects become void.

In Figure 4.6, the delay performance of SUs is shown as a function of the traffic capacity C of PUs for different values of ρ and K. It is observed that a SU can have a better response time when the number of sub channels is relatively low. This is because increasing the number of sub-channels does not improve too much on droupout effects.

4.5 Summary

Admission control of SUs to achieve a certain degree of QoS has been studied. A multiple dimension Markov chain whose size increases quickly with the number of channels of PUs is assumed to characterize the known system. In the presence of uncertainty associated with the system size reduction, we consider a one dimension Markov chain approximation to characterize the uncertain system. Appropriate analytical models are developed and used to derive the performance characteristics of the secondary admission strategies. Numerical results are reported and show a robust capacity in terms of blocking probability, dropping probability and delay performance as functions of spectrum utilization by PUs.

Chapter 5

Conclusion

5.1 Introduction

The exploding increase of wireless communications combined with the existing inefficient usage of the licensed spectrum gives a strong impetus to the development and standardization of cognitive radio networking and communications. Much of the interest in RSA comes from the need to match increasing wireless demands of today's societies and the recognition that current gridlock approach to spectrum use results in inefficient use of the radio spectrum.

5.2 Summary of the Thesis

In this dissertation, the alternative approach to RSA, called DSA, is first presented as a means for alleviating spectrum gridlocks by using spectrum sharing. Based on this, CR can be developed to support DSA. The DSA concept means that the users of a wireless system are divided into a multi-tiered hierarchy with the primary users (PUs) entitled to protection and cognitive radio capable secondary users (SUs).

DSA views the spectrum as an open resource to be shared among PUs and SUs in such way that performance benefits of such a sharing arrangement can be fully realized while at the same time avoiding excessive interference caused to PUs. The performance sharing gains are obtained by means of a medium access control protocol with knowledge of the statistical properties or available local information of the channels already occupied by PUs as well as knowledge of the interference tolerance within which the interference to PUs is kept to a given level.

5.3 Concluding Remarks

An important contribution of this thesis is to match the discrepancy from the apparent spectrum overcrowding and the recognition that outdated spectrum access techniques usually result in inefficient use of the spectrum. To do this, we use the DSA concept as a method that views the spectrum not only as an open resource, but as a resource more effectively and efficiently used. Spectrum overcrowding comes from outdated spectrum policies and wireless technologies. The effect is that the societal and economic benefits that come with a new wireless technology are limited. However, methods supporting diverse forms of spectrum sharing such as DSA have the potential to unleash new technologies as long as policies are reformed to match the technology.

Further, emphasis has been laid on the potential role of DSA in cognitive radio communications in light the channel models provided by DSA. Three classes of communication models that describe several situations arising in the cognitive communication process have been studied considering the nature of information available about the channels over which transmissions of SUs take place. These are: with dynamic, distributed channels information; with dynamic, parallel channels information; and under a dynamic sub-channels allocation scheme. Related to this, emphasis is laid on the MAC protocol capabilities to determine the efficiency of the secondary sharing of spectrum.

The performance results clearly indicate the feasibility of the DSA concept as a candidate for future wireless infrastructures that has the potential to support economic growth and evolving societal needs. This approach offers the choice of performance that is restricted to the capabilities of spectrum sharing arrangements, and less on the spectrum gridlocks. In specific it offers robust performance with reference to the uncertainty associated with localized sensing of distributed, dynamic channels and with timely sensing of parallel, dynamic channels. The extension to dynamic, parallel sub-channels enables resource allocation to be carried out in sub-channels. The analytical results on the performance of sub-channel allocation indicate a robust traffic capacity in terms of blocking probability, drop-out probability and delay performance as function of PUs traffic loads.

5.4 Future Work

Several important issues have been left open in this thesis and they need to be thoroughly investigated in the future. The following topics are of particular interest:

Policy enforcement: DSA is a wireless technology whose implementation greatly depends on spectrum policy deliberations at the national and the international level. Policies are enforced primarily by preventing beforehand the deployment of devices that are capable of violating regulations, rather than detecting and punishing violators afterwards. To do this, regulators test and approve products before they can be widely distributed. However, technologies behind cognitive radio complicate the problem of ensuring a priori that a device complies with regulator test. The more a device alters its behavior in accordance with what it senses from its surrounding radio environment the harder it is to exhaustively test such behaviors. Hence the viability of DSA hinger on designer ability to make testing simple for the regulator that must grant permission before a new device become operational.

Statistical modeling: Spectrum occupancy data are pertinent to DSA experimentation and implementation. Traditional measurements and analyses intended to quantify the performance of one particular system usually cannot be applied directly to the design of DSA systems. Instead, spectrum measurements and data analyses need to be directed to developing statistical models with sufficient richness to describe the occupancy in real spectrum bands, which can be conveniently parameterized, and readily amenable to mathematical analysis.

Testbed platform: Monte Carlo computer simulation is a flexible performance prediction tool widely used to assess communications infrastructures. Its flexibility stems from the fact that it consists of a computer program that behaves like the system under study. The network operations can be conveniently described by simulation programs. The stochastic nature of demands can be modeled by using pseudo-random generators. Therefore, a Monte Carlo simulation program can serve as a flexible testbed platform for conduction system experimentation without constructing software/hardware prototypes. However, several factors must be considered in oder to apply a successful simulation technology to DSA experimentation. To begin with, the nature of secondary applications are incompletely known. Further, the secondary network is cognitive radio based, that is, the more a cognitive radio alters its behavior in accordance with what it senses from its surrounding radio environment the harder it is to exhaustively model its behaviors.

Bibliography

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017," White Paper, Cisco, Feb. 2013.

[2] "Fifteenth annual report and analysis of competitive market conditions with respect to mobile wireless, including commercial mobile services," Docket No. 10-133, FCC, Jun. 2011.

[3] J. Robinson, "Spectrum management policy in the united states: An historical account," Working Paper, FCC, Apr. 1985.

[4] M. A. McHenry, D. McCloskey, and G. Lane-Roberts, "Spectrum occupancy measurements location 4 of 6: Republican national convention new york city, new york," White Paper, Shared Spectrum Company, Aug. 2005.

[5] K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, and C.E. Wheatley, "On the capacity of a cellular cdma system," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, pp. 303–12, May 1991.

[6] G.J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Per-*

sonal Communications (Wireless Pers Commun), vol. 6, no. 3, pp. 311–335, Apr. 1998.

[7] R. W. Chang, “Synthesis of band-limited orthogonal signals for multichannel data transmission,” Bell Lab Tech. J., Dec. 1996.

[8] “Federal communications commission, authorization of spread spectrum systems under parts 15 and 90 of the FCC rules and regulations,” Docket 81-413, FCC, 1985.

[9] J. Mitola, “An integrated agent architecture for software defined radio,” Ph.D. dissertation, School of Teleinformatics, Royal Inst. Technol., Stockholm, Sweden, May 2000.

[10] Steven H. Low, Fernando Paganini, and John C. Doyle, “Internet congestion control,” *IEEE Control Systems Magazine*, vol. 22, no. 1, pp. 28–43, Feb. 2002.

[11] S. Jafar and S. Srinivasa, “Cognitive medium access: Constraining interference based on experimental models,” *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 529–37, Apr. 2007.

[12] I. Macaluso, T.K. Forde, L. DaSilva, and L. Doyle, “Impact of cognitive radio: recognition and informed exploitation of grey spectrum opportunities,” *IEEE Vehicular Technology Magazine (IEEE Veh. Technol. Mag.)*, vol. 7, no. 2, pp. 85–90, Jun. 2012.

[13] C. Stevenson, G. Chouinard, Zhongding Lei, Wendong Hu, S. Shellhammer, and W. Caldwell, “IEEE 802.22: The first cognitive radio wireless regional area network standard,” *IEEE Communications Magazine (IEEE Commun. Mag.)*, vol. 47, no. 1, pp. 130–8, Jan. 2009.

[14] Qing Zhao, Lang Tong, Ananthram Swami, and Yunxia Chen, "Cross-layer design of opportunistic spectrum access in the presence of sensing error," in *2006 IEEE Conference on Information Sciences and Systems, CISS 2006*, Princeton, NJ, USA, Mar. 2007, pp. 778–782.

[15] Y. Chen, Q. Zhao, and A. Swami, "Distributed cognitive MAC for energy-constrained opportunistic spectrum access," in *2006 IEEE Military Communication Conference(MILICOM)*, Washington, DC, USA, 2006.

[16] Yunxia Chen, Qing Zhao, and Ananthram Swami, "Distributed spectrum sensing and access in cognitive radio networks with energy constraint," *IEEE Transactions on Signal Processing(IEEE Trans Signal Process)*, vol. 57, no. 2, pp. 783–797, 2009.

[17] Danijela Cabric and Robert W. Brodersen, "Physical layer design issues unique to cognitive radio systems," in *IEEE New Frontiers in Dynamic Spectrum (DySPAN'10)*, Berlin, Apr. 2005, pp. 759–763.

[18] Yunxia Chen, Qing Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access," in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2006, pp. 696–700.

[19] Yunxia Chen, Qing Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.

[20] Amir Ghasemi and Elvino S. Sousa, "Optimization of spectrum sensing for opportunistic spectrum access in cognitive radio networks," in *2007 4th Annual IEEE Consumer Communications and Networking*

Conference (CCNC 2007), Las Vegas, NV, United states, Jan. 2007, pp. 1022–1026.

[21] Ying-Chang Liang, Yonghong Zeng, E. Peh, and Anh Tuan Hoang, "Sensing-throughput tradeoff for cognitive radio networks," in *2007 IEEE International Conference on Communications*, Glasgow, UK, Jun. 2007, pp. 5330–5.

[22] Ying-Chang Liang, Yonghong Zeng, E. Peh, and Anh Tuan Hoang, "Sensing-throughput tradeoff in cognitive radio networks: how frequently should spectrum sensing be carried out?" in *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, Athens, Greece, 2007, pp. 3266–70.

[23] Qing Zhao, Lang Tong, and A. Swami, "A cross-layer approach to cognitive MAC for spectrum agility," in *2005 39th Asilomar Conference on Signals, Systems and Computer*, Pacific Grove, CA, USA, Nov. 2005, pp. 200–4.

[24] Tang Liang, Yunfei Chen, Evor L. Hines, and Mohamed-Slim Alouini, "Effect of primary user traffic on sensing-throughput tradeoff for cognitive radios," *IEEE Transactions on Wireless Communications(IEEE Trans. Wirel. Commun. (USA))*, vol. 10, no. 4, pp. 1063–1068, Apr. 2011.

[25] S. Geirhofer, Lang Tong, and B.M. Sadler, "A measurement-based model for dynamic spectrum access in WLAN channels," in *MILCOM 2006*, Washington, DC, USA, Oct. 2006.

[26] S. Geirhofer, Lang Tong, and B.M. Sadle, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 66–72, 2007.

[27] Qianchuan Zhao, S. Geirhofer, Lang Tong, and B.M. Sadler, "Opportunistic spectrum access via periodic channel sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 785–796, Jan. 2008.

[28] Qianchuan Zhao, S. Geirhofer, Lang Tong, and Sadler B.M., "Optimal dynamic spectrum access via periodic channel sensing," in *2007 8th IEEE Wireless Communications and Networking Conference*, Kowloon, China, Mar. 2007, p. 5.

[29] Senhua Huang, Xin Liu, and Zhi Ding, "Opportunistic spectrum access in cognitive radio networks," in *2008 Proceedings IEEE INFOCOM*, Phoenix, AZ, USA, Apr. 2008, pp. 2101–9.

[30] Xiao Qinghai, Gao Qunyi, Xiao Limin, Zhou Shidong, and Wang Jing, "An optimal opportunistic spectrum access approach," in *2009 IEEE International Conference on Communications Workshops (ICC 2009*, Dresden, Germany, Jun. 2009.

[31] Li Xin, Zhao Qianchuan, Guan Xiaohong, and Tong Lang, "Optimal cognitive access of markovian channels under tight collision constraints," in *2010 IEEE International Conference on Communications (ICC 2010)*, Cape Town, South africa, May 2010.

[32] Senhua Huang, Xin Liu, and Zhi Ding, "Optimal transmission strategies for dynamic spectrum access in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 12, pp. 1636 – 1648, Dec. 2009.

[33] Qinghai Xiao, Yunzhou Li, Zhong Xiaofeng, Xu Xibin, and Wang Jing, "A unified approach to optimal opportunistic spectrum access under collision probability constraint in cognitive radio systems," *Eurasip Journal on Advances in Signal Processing(Eurasip. J. Adv. Sign. Process.)*, 2010.

[34] Jeonghoon Mo, Hoi-Sheung Wilson So, and M.C. Jean Walrand, "Comparison of multichannel mac protocols," *IEEE Transactions on Mobile Computing (IEEE Trans. Mob. Comput.*, vol. 7, no. 1, pp. 50–65, Jan. 2008.

[35] P. Kyasanur, , So Jungmin, C. Chereddi, and N.H. Vaidya, "Multichannel mesh networks: challenges and protocols," *IEEE Wireless Communications (IEEE Wirel. Commun.*, vol. 13, no. 2, pp. 30–6, Apr. 2006.

[36] Bany Salameh, , A. Haythem, and Krunz Marwan, "Channel access protocols for multihop opportunistic networks: Challenges and recent developments," *IEEE Network*, vol. 23, no. 4, pp. 14–19, Apr. 2009.

[37] S. Geirhofer, Lang Tong, and B.M. Sadle, "Cognitive medium access: Constraining interference based on experimental models," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 95–101, Jan. 2008.

[38] D.V Djonin, Qing Zhao, and V. Krishnamurthy, "Optimality and complexity of opportunistic spectrum access: a truncated markov decision process formulation," in *2007 IEEE International Conference on Communications*, Glasgow, UK, Jun. 5787-92, p. 2008.

[39] Xin Li, Qianchuan Zhao, Xiaohong Guan, and Lang Tong, "Optimal cognitive access of markovian channels under tight collision constraints," vol. 29, no. 4, pp. 746–756, Apr. 2011.

[40] E. David and Lon-Rong Hu, "Traffic engineering of the radio interface for cellular mobile networks," *IEEE Proceedings of the IEEE*, vol. 82, no. 9, pp. 1383–1397, 1994.

[41] S. Stephen Rappaport and Lon-Rong Hu, "Microcellular communication systems with hierarchical macrocell overlays: Traffic performance

models and analysis," *IEEE Proceedings of the IEEE*, vol. 82, no. 9, pp. 1383–1397, 1994.

[42] Xiaorong Zhu, Lianfeng Shen, and Tak-Shing Peter Yum, "Analysis of cognitive radio spectrum access with optimal channel reservation," *IEEE Commun. Lett.*, vol. 11, no. 4, pp. 304–306, 2007.

[43] W. Ahmed, H. Suraweera, and M. Faulkner, "Comments on analysis of cognitive radio spectrum access with optimal channel reservation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4488–4491, Apr. 2009.

[44] Eric W. M. Wong and Chuan Heng Foh, "Analysis of cognitive radio spectrum access with finite user population," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 294–296, May 2009.

[45] Yan Zhang, "Dynamic spectrum access in cognitive radio wireless networks," in *IEEE ICC 2008*, Beijing, China, May 2008, pp. 4927–4932.

[46] Yong Yao, S. R. Ngoga, D. Erman, and A. Popescu, "Performance of cognitive radio spectrum access with intra- and inter-handoff," in *IEEE ICC 2012*, Ottawa, Canada, May 2010.

[47] S. R. Ngoga, Yong Yao, and A. Popescu, "Beacon-enabled cognitive access for dynamic spectrum access," in *Proceedings of the 2012 8th Euro-NF Conference on Next Generation Internet (NGI)*, Karlskrona, Sweden, Jun. 2012, pp. 49–56.

[48] S. R. Ngoga, Yong Yao, and A. Popescu, "A simple time-based threshold strategy for dynamic spectrum access," in *Proceedings 2012 9th International Conference on Communications (COMM 2012)*, Bucharest, Romania, Jun. 2012, pp. 225–8.

[49] S. R. Ngoga, Yong Yao, and A. Popescu, "On multi-channel opportunistic access in heterogeneous cognitive radio networks," in *The Third International Conference on Advances in Cognitive Radio (COCORA 2013)*, Venice, Italy, Apr. 2013, pp. 29–34.

[50] Qing Zhao and B.M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[51] J. M. Peha, "Sharing spectrum through spectrum policy reform and cognitive radio," *Proc. IEEE (USA)*, vol. 97, no. 4, pp. 708–709, Apr. 2009.

[52] Qing Zhao and A. Swami, "A decision-theoretic framework for opportunistic spectrum access," *IEEE Wireless Commun. Mag.*, vol. 14, no. 4, pp. 14–20, Aug. 2007.

[53] Matthias Wellens and Petri M., "Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model," *(Springer) Mobile Networks and Applications*, vol. 15, no. 3, pp. 461–474, Jun. 2010.

[54] J. Mitola, "Cognitive radio architecture evolution," *Proc. IEEE (USA)*, vol. 97, no. 4, pp. 626–41, Apr. 2009.

[55] A. P. Jardosh, K. N. Ramachandran, K. C. Almeroth, and E. M. Belding-Royer, "Understanding congestion in IEEE 802.11b wireless networks," in *Proc. ACM SIGCOMM Workshops Conf. Comput. Commun.*, Philadelphia, PA, Aug. 2005, pp. 11–16.

[56] Kishor S. Trivedi, *Probability and Statistics with Reliability, Queuing and Computer Science Applications.* Wiley-Interscience, 2001.

[57] N. Balakrishnan and V. B. Nevzorov, *A Primer on Statistical Distributions.* Wiley-Interscience, 2001.

[58] Samuel Kotz and Saralees Nadarajah, *Extreme Value Distributions: Theory and Applications.* World Scientific, 2001.

[59] M J Marcus, "Dynamic spectrum access as a mechanism for transition to interference tolerant systems," in *IEEE 2010 IEEE Symposium on New Frontiers in Dynamic Spectrum, DySPAN 2010 (DySPAN'10)*, Singapore, Apr. 2010.

[60] R Tandra, A. Sahai, and V. Veeravalli, "Unified space-time metrics to evaluate spectrum sensing?" *IEEE Wireless Commun. Mag.*, vol. 49, no. 3, pp. 54–61, May 2011.

[61] Kwang-Cheng Chen, "Medium access control of wireless lans for mobile computing," *IEEE Network*, vol. 8, no. 5, pp. 50–63, Oct. 1994.

[62] C. Ghosh, S. Pagadarai, and A. Wyglinski, "A framework for statistical wireless spectrum occupancy modeling," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 38–44, Jan. 2010.

[63] J. Mitola and Jr Maguire G.Q., "Cognitive radio: making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[64] P. Pollin, Hoi-Sheung So, A. Bahai, R. Prasad, and R. Hekmat, "Quality of service assessment of opportunistic spectrum access: a medium access control approach," *IEEE Wireless Commun. Mag.*, vol. 15, no. 5, pp. 20–29, Oct. 2008.

[65] James Gross and Mathias Bohge, "Dynamic mechanisms in OFDM wireless systems: A survey on mathematical and system engineering contributions," Technical University Berlin, TKN Technical Report TKN-06-001, May 2006.

[66] S. Shakkottai and T. S. Rappaport, “Cross-layer design for wireless networkss,” *IEEE Wireless Commun. Mag.*, vol. 41, no. 10, pp. 74–80, Oct. 2003.

[67] A. Gyasi-Agyei, “Multiuser diversity based opportunistic scheduling for wireless data networks,” *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 670–2, 2005.

[68] R. Jain and S.A. Routhier, “Packet trains-measurements and a new model for computer network traffic,” *IEEE J. Sel. Areas Commun.*, vol. 4, no. 6, pp. 986–95, 1986.

[69] Yakov Ben-Haim, *Info-gap decision theory for engineering design. Or: why Good is preferable to Best.* CRC Press Boca Raton, 2005.

[70] D. D. Kouvatsos, “Entropy maximisation and queueing network models,” *Ann. Oper. Res. (Switzerland)*, vol. 48, no. 1-4, pp. 63–126, Jan. 1994.

[71] Dongxu Shen, Li, and O. K. Victor, “Performance analysis for a stabilized multi-channel slotted aloha algorithm,” in *PIMRC2003 - 14th IEEE 2003 International Symposium on Personal, Indoor and Mobile Radio Communications, Proceedings*, Beijing, China, May 2003.

[72] Zhao Liu and El Zarki, “Medium access control of wireless lans for mobile computing,” *(Netherlands) Wireless Networks*, vol. 1, no. 1, pp. 1–16, 1995.

[73] Jyh-Horng Wen and Jee-Wey Wang, “A recursive solution to an occupancy problem resulting from tdm radio communication application,” *Appl. Math. Comput. (USA)*, vol. 101, no. 1, pp. 1–3, 1999.

[74] P.F Marshall, “Extending the reach of cognitive radio,” *Proceedings of the IEEE(IEEE Commun. Mag. (USA)*, vol. 97, no. 4, pp. 612–25, Apr. 2009.

[75] M.C Vuran, I.F Akyildiz, Lee Won-Yeol, and S. Mohanty, “Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey,” *Computer Networks(Comput. Netw. (Netherlands))*, vol. 50, no. 13, pp. 2127–59, 2006.

[76] I.F. Akyildiz, Lee Won-Yeol, M.C. Vuran, and S. Mohanty, “A survey on spectrum management in cognitive radio networks,” *IEEE Communications Magazine(IEEE Commun. Mag. (USA))*, vol. 46, no. 4, pp. 40–8, Apr. 2008.

[77] Liang Ying-Chang, Zeng Yonghong, E.C.Y. Peh, and Hoang Anh Tuan, “Sensing-throughput tradeoff for cognitive radio networks,” *IEEE Transactions on Wireless Communications(IEEE Trans. Wirel. Commun. (USA))*, vol. 7, no. 4, pp. 1326–37, Apr. 2008.

[78] P. Viswanath, D.N.C. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Transactions on Information Theory(IEEE Trans. Inf. Theory (USA))*, vol. 48, no. 6, pp. 1277–94, Jun. 2002.

[79] J. Martinez-Bauset, V. Pla, M.J., Domenech-Benlloch, and D. Pacheco-Paramo, “Admission control and interference management in dynamic spectrum access networks,” *EURASIP Journal on Wireless Communications and Networking(EURASIP J. Wirel. Commun. Netw. (USA))*, vol. 48, no. 6, pp. 1277–94, Jun. 2002.

ABSTRACT

The exploding increase of wireless communications combined with the existing inefficient usage of the licensed spectrum gives a strong impetus to the development and standardization of cognitive radio networking and communications. In this dissertation, a framework for Dynamic Spectrum Access (DSA) is first presented, which is the enabling technology for increasing the spectral efficiency of wireless communications. Based on that, Cognitive Radio (CR) can be developed as an enabling technology for supporting the DSA, which means that the wireless users are provided with enhanced capability for sensing the operating radio environment and for exploiting the network side information obtained from this sensing.

The DSA concept means that the users of a wireless system are divided into a multi-tiered hierarchy with the primary users (PUs) entitled to protection and with cognitive radio capable secondary users (SUs). The improved spectrum efficiency is obtained by means of a medium access control protocol with knowledge about the statistical properties or available local information of the channels already occupied by PUs as well as knowledge about the interference tolerance within which the interference to PUs is kept to a given level. Related to this, emphasis is laid on the protocol capability to determine the efficiency of the secondary sharing of spectrum. Based on the type of available local information, the capacity of opportunistic communication is investigated for three models. These are: with dynamic, distributed channels information; with dynamic, parallel channels information; and under a dynamic sub-channels allocation scheme.

The results indicate that this capacity is robust with reference to the uncertainty associated with localized sensing of distributed dynamic channels and with timely sensing of parallel dynamic channels. The extension to dynamic parallel sub-channels enables resource allocation to be carried out in sub-channels. The analytical results on the performance of sub-channel allocation indicate a robust traffic capacity in terms of blocking probability, drop-out probability and delay performance as function of PUs traffic loads.

ISSN 1653-2090
ISBN: 978-91-7295-267-6

2013:12