Yuehong Gao

Performance Analysis of a Cognitive Radio Network Using Network Calculus

Thesis for the degree of Philosophiae Doctor

Trondheim, December 2012

Norwegian University of Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Telematics

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Telematics

ISBN 978-82-471-4046-8 (printed ver.)
ISBN 978-82-471-4047-5 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2012:358

Printed by NTNU-trykk

Abstract

Cognitive radio has become a promising technology to increase spectrum utilization through spectrum sharing between licensed users (primary users) and unlicensed users (secondary users). Performance evaluation and analysis is of key importance to get more knowledge of this newly emerged technology.

Queueing theory and Markov chain model have been applied to conduct the analysis, where some results are obtained. The existing research sheds light on the performance of cognitive radio networks. However, there are also some limitations. Poisson arrival and exponentially distributed service time are mostly assumed in existing analysis. With these assumptions, existing queueing theory results, particularly $M/G/1$ priority queue results, can be directly applied, and the Markov chain model can be established. However, these assumptions are too restrictive for modern wireless communication networks, where the traffic can be of different types and the channel capacity can vary over time. In addition, particular focus is made on average values (such as average delay) with little investigation on probabilistic distribution bounds.

Therefore, new methodology is needed to make a breakthrough and to bring new insights regarding performance of cognitive radio networks. Network calculus, a newly developed theory, provides a possible solution. It was firstly proposed by R. L. Cruz in 1991, and has involuted into two branches now, i.e., deterministic network calculus and stochastic network calculus. There are two basic concepts in network calculus: (stochastic) arrival curve and (stochastic) service curve, which are used to describe the arrival process of input traffic and the service process of server, respectively. Probabilistic performance guarantees can be analyzed using stochastic network calculus. In addition, the independence between arrival process and service process can be exploited to obtain tighter probabilistic bounds, which is called as independent case analysis.

This work is devoted to applying network calculus analysis, particularly stochastic network calculus, to performance evaluation of a cognitive radio

network. Generally, this work contains two scenarios: a single-channel scenario and a multiple-channel scenario.

In the single-channel scenario, several key factors are considered and discussed. Different traffic models, including Periodical source and (Compound) Poisson source, are considered. The sensing error process, which can be further divided into mis-detection process and false alarm process, is modeled, and the corresponding stochastic arrival process is analyzed. Both constant channel and Gilbert-Elliott ON-OFF fading channel are investigated, and the stochastic service curves are obtained. In addition, the influence of different re-transmission schemes is also studied. These schemes include no-re-transmission, re-transmission until success and maximum-N-time re-transmission. Probabilistic delay distribution bound and probabilistic backlog distribution bound are obtained and discussed. Furthermore, delay-constrained capacity is defined and capacity regions for both primary users and secondary users are also studied. Independent case analysis is applied in some cases, and results are compared and discussed. In the multiple-channel scenario, the main concern is put on the guaranteed service of multiple parallel channels and delay-constrained capacity by assuming perfect sensing and constant channel.

In validating the theoretical results, system configurations are specified mainly based on Long Term Evolution (LTE) networks, and numerical calculations executed in Matlab are made to visually depict the results. Simulation platform by C++ language is also constructed to obtain simulation results. Related results show the influence of different schemes or parameters on system performance and capacity region. Comparison between the numerical and simulation results is made, which further verifies the theoretical deductions.

Preface

This thesis is submitted in partial fulfillment of the requirement for the degree of philosophies doctor (PhD) at the Norwegian University of Science and Technology (NTNU). The PhD study was formally conducted at the Department of Telematics (ITEM) at NTNU, and I have been hosted and funded by the Centre of Quantifiable Quality of Service in Communication Systems, Centre of Excellence (Q2S), at NTNU. Professor Yuming Jiang has been the supervisor of this work, and Professor Geir E. Øien has been the co-supervisor.

I would like to thank several people who gave me a lot of support during my stay at Q2S. First and foremost is my supervisor Professor Yuming Jiang. I feel quite fortunate to have Yuming Jiang as my supervisor. His outstanding research skills and elaborate research attitude influence me a lot. When I encountered difficulties in my work, he was always there to give me guidance and confidence to stride forward. Without his discussion, advice and patience, this work would not be finished. I also would like to thank my co-supervisor Professor Geir E. Øien, who suggested the interesting topic when I started my phd work.

All former and current people working at Q2S are thanked for forming a peaceful and joyful atmosphere. I also thank my friends for their encouragement and accompany. I could not list all of their names, but I remember all the happy times we spent together. Special thanks are given to Anniken Skotvoll for her kind and patient help, which makes me feel warm in the heart and reduces my homesickness as a foreigner.

My sincere appreciation goes to my current colleagues, especially Dr. Dacheng Yang and Dr. Xin Zhang, from School of Information and Telecommunication Engineering (SICE), Beijing University of Posts and Telecommunications (BUPT). Their cooperation and encouragement greatly support me during my phd study time at NTNU and also during my working time at BUPT.

Last but not the least, I am thankful to my family, especially my husband Ning Yin. He took care of our parents when I was abroad, and he always encouraged me to stick on when I encountered problems and felt frustrated. His support is an important power for me to persist and complete this work.

Yuehong Gao
September 2012

Abbreviations

3G	3rd Generation
ACK	Acknowledge
AMC	Adaptive Modulation and Coding
CCDF	Complimentary Cumulative Distribution Function
CDMA	Code Division Multiple Access
CE	Competitive Equilibrium
CP	Cyclic Prefix
CS	Correct Sensing
CSGC	Color Sensitive Graph Coloring
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CTMC	Continuous Time Markov Chain
ETSI	European Telecommunications Standards Institute
FA	False Alarm
FCC	Federal Communications Commission
FDMA	Frequency Division Multiple Access
FIFO	First-In-First-Out
FSMC	Finite State Markov Channel
GA-SAA	Genetic Algorithm Spectrum Allocation Algorithm
GBR	Guaranteed Bit Rate
GE	Gilbert-Elliott
ISI	Inter-Symbol-Interference
ISM	Industrial, Scientific and Medical
ITU	International Telecommunication Union
LO	Local Oscillator
LTE	Long Term Evolution
MAC	Media Access Control
Max-N-RT	Maximum-N-time Re-Transmission
MD	Mis-Detection
MF	Matched Filter
MGF	Moment Generating Function

OFDM	Orthogonal Frequency Division Multiplexing
PDSS	Price-based Distributed Spectrum Sharing
PF	Proportional Fair
PSO-SAA	Particle Swarm Optimization Spectrum Allocation Algorithm
PUs	Primary Users
QCI	QoS Class Identifier
QGA-SAA	Quantum Genetic Algorithm Spectrum Allocation Algorithm
QoS	Quality of Service
RB	Resource Block
RT	Re-Transmission
RT-S	Re-Transmission until Success
SAC	Stochastic Arrival Curve
SE	Sensing Error
SNIR	Signal to Noise and Interference Ratio
SNR	Signal to Noise Ratio
SSC	Stochastic Service Curve
SUs	Secondary Users
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
UMTS	Universal Mobile Telecommunications System
vbc	virtual backlog centric
VoIP	Voice over IP
WO-RT	Without Re-Transmission

Notations

A(t)	Cumulative amount of arrival traffic up to time t
$A^*(t)$	Cumulative amount of departure traffic up to time t
$B(t)$	System backlog at time t
C_i	Transmission rate of channel i
$D(t)$	System delay at time t
F	Set of non-negative wide-sense increasing functions
$\bar{F}$	Set of non-negative wide-sense decreasing functions
fl^P	Traffic flow formed by primary users
fl^S	Traffic flow formed by secondary users
$h(\alpha + x, \beta)$	maximum horizontal distance between functions $\alpha(t) + x$ and $\beta(t)$
$I(t)$	Cumulative amount of impaired service up to time t
L	Packet length
p^{MD}	The probability that a mis-detection happens in a slot
p^{FA}	The probability that a false alarm happens in a slot
p^{SE}	The probability that a sensing error happens in a slot
q_{ij}	Transition probability from state i to state j of a GE channel
$S(t)$	Cumulative amount of provided service up to time t
T	Slot length
λ	Arrival rate of Poisson traffic
ϵ	Maximum tolerable delay violation probability
η_i	Active factor of the primary network on channel i

List of Figures

List of Tables

Table of Contents

Part I

Introduction

Chapter 1

Motivation and Focus of This Thesis

1.1 Motivation

As the requirements for mobile communications increase greatly, wireless communication systems have been developed quickly. However, the available spectrum resources, which are suitable for wireless transmissions, are scarce. Therefore, how to utilize the limited spectrum in a more efficient way becomes a vital task, and attracts significant attention from both industry and academy. The proposal of cognitive radio provides a remarkable solution, and leads to fruitful results after more than ten years development. Many research efforts have been put on specific problems in cognitive radio networks, such as spectrum sensing algorithms and spectrum allocation mechanisms. The evaluation and analysis from the network performance viewpoint is one of the important issues.

The inherent characteristics of a cognitive radio network pose challenges for performance evaluation. First, there are two types of users, named primary users and secondary users, sharing the spectrum according to different priorities. Second, spectrum sensing is relied on by the secondary users to find the vacant spectrum before their transmissions. Imperfect sensing can greatly impact the system performance, which will introduce collisions between a primary user and a secondary user, or will lead to waste of transmission opportunities. Third, retransmission is usually employed in order to cope with loss. In addition, channel fading and offered traffic will also affect the performance. To sum up, all these factors make the evaluation process complicated.

In the literature, the classic queueing theory has been used to conduct performance analysis of cognitive radio networks. In [1], an $M/D/1$ priority queueing system model is used to derive the average waiting times and average queueing lengths in a cognitive radio network with perfect spectrum sensing. $M/M/1$ queueing model is employed to analyze the average queueing time of secondary users in [2]. The authors of [3] relied on the $M/G/1$ preemptive priority queue to obtain analytical forms of average delay and throughput for both PUs and SUs. In these works, the impact of spectrum sensing errors on the system performance is not well studied, and they only provide results in terms of average values with little investigation on probabilistic delay bounds. Furthermore, the $M/G/1$ model assumes that the system is work-conserving. However, due to sensing errors, the system may not be work-conserving, which implies that the $M/G/1$ priority results can not be directly applied when sensing error is taken into consideration. In addition, $M/G/1$ results can not be applied to the multi-channel case.

The Markov chain model has also been relied on to conduct performance analysis. Considering the cognitive radio scenario, the state space of the Markov chain can be defined in two ways: (1) based on the channel occupancy state (that is, whether a channel is free, occupied by PU, occupied by SU or collision), and (2) the number of PUs and the number of SUs in the system. In most literatures, the second is used for indices of the Markov state space, such as in [4–6], because, with it, the dimensionality and complexity of the Markov model (especially in the multi-channel case) can be more easily reduced as highlighted by the authors of [5]. The Markov chain based analysis helps to derive the blocking probability for SUs, average number of users in the system as well as throughput. However, to the best of our knowledge, no delay-related results are available from this analysis.

In all the related works, the authors assume Poisson arrival and most of them also assume exponentially distributed service time, so that existing queueing theory results, particularly $M/G/1$ priority queue results, can be directly applied, and the Markov chain model can be established. However, these assumptions are too restrictive for modern wireless communication networks, where the traffic can be of different types and the channel capacity can vary over time.

Therefore, it is of great importance to make all-rounded performance analysis of a cognitive radio network by considering different traffic models, channel models and other factors. Network calculus, an analysis theory based on min-plus/max-plus algebra, provides a novel approach to fill in the blank area. By applying network calculus, system performance for more traffic models can be evaluated, the influence of channel fading and re-transmission schemes can be studied, and capacity region for primary users and secondary users can be obtained. All these results will then provide more insights on the performance of cognitive radio networks. At the same time, application of network calculus will be extended from computer networks to wireless communication networks.

1.2 Focus of This Thesis

In this work, a newly developed approach – network calculus, specifically stochastic network calculus – is employed to perform stochastic service guarantee analysis of cognitive radio networks. Several key factors that will influence the final performance are considered, including:

- *Traffic models.* Poisson traffic will be used as a classical model. Besides, periodical source model and compound Poisson traffic will also be studied. The mathematical model, named (stochastic) arrival curve in network calculus, will be used to describe the characteristics of traffic.

- *Wireless channel models.* Fading is perhaps the most significant difference between wireless channels and wired channels. Thus, channel fading and its impact should be considered and studied. In this work, the wireless channel will be modeled as a Gilbert-Elliott (GE) channel with two states: *ON* and *OFF*. In state *ON*, data can be transmitted and received correctly, while in state OFF, no data can be received by the receiver due to deep fading. The mathematical model, named stochastic service curve for such a channel will be obtained. In addition, constant channel is also employed and studied in this work.

- *Sensing error process.* Spectrum sensing can be thought as the basis of a cognitive radio network. The secondary users obtain the information about spectrum occupancy state through spectrum sensing mechanisms, and then they will make use of the unused spectrum so that the spectrum utilization can be improved. However, spectrum sensing results are not always consistent with the real conditions due to many difficulties during the sensing process, which are called sensing errors. In this work, the sensing errors will be classified into two types, i.e. Mis-Detection (MD) and False Alarm (FA). Stochastic arrival curves of these processes will be derived, and their influences on performance will also be studied and analyzed.

- *Re-transmission schemes.* The transmitted packet may not be received successfully by the receiver because of deep fading or misdetection. Therefore, Re-Transmission (RT) is needed in order to guarantee a certain packet loss probability. Different schemes will be studied in this work, including without re-transmission, re-transmit until success and max-N-time re-transmission.

The performance will be evaluated and compared mainly by the probabilistic delay/backlog distribution bound and delay-constrained capacity. Both a single-channel scenario and a multiple-channel scenario are modeled and studied. Furthermore, independent case analysis approach in stochastic network calculus is also investigated and applied to obtain better bounds and capacity region.

Chapter 2

Background

In this part, a short introduction to the related background will be presented. First, cognitive radio network is described including the network architecture and spectrum sensing techniques. Second, network calculus related contents are introduced, including its mathematic basis, traffic model, server model, useful properties and independent case analysis. Then, the cognitive radio network considered in this work is modeled, and the concerned output parameters are presented.

2.1 Cognitive Radio Network

Nowadays, the development of wireless communication techniques cannot meet the fast increasing of communication requirements. One of the most essential bottlenecks is the scarce spectrum resource, which is suitable for wireless transmissions. The fixed spectrum allocation policy makes the problem even severer. In most countries, the use of spectrum bands is regulated by governments through a spectrum management process known as spectrum allocation. A number of forums and standards bodies are also involved in the spectrum management process, such as International Telecommunication Union (ITU) and European Telecommunications Standards Institute (ETSI). Spectrum bands are assigned in three types [7]:

- No one may transmit: frequencies reserved for radio astronomy to avoid interference at radio telescopes;
- Anyone may transmit, as long as they respect certain transmission power and other limits: open spectrum bands such as the unlicensed Industrial, Scientific and Medical (ISM) bands. The "listen before talk" contention based protocol is mostly used in this case;
- Only the licensed user of that band may transmit: the licensing body may give the same frequency to several users as a form of frequency reuse if they do not interfere because their coverage map areas never overlap.

For those high-demand sections of the electromagnetic spectrum, auctions may be used to decide who can use them. Generally, the aforementioned spectrum allocation is static or fixed for a certain time length. It has been found that some bands are not efficiently utilized in space domain or in time domain. According to [8], the utilization of the fixed spectrum assignment is approximately 15-85% based on temporal and geographical variations. On the contrary of under-utilization, the requirements for spectrum resource increase urgently. In other words, dynamic spectrum allocation mechanism is required in order to make better use of the spectrum.

The proposal of cognitive radio provides a novel solution to improve spectrum utilization efficiency. The concept of cognitive radio was firstly proposed by Joseph Mitola III in a seminar in 1998, and then published in an article by Mitola and Maguire in 1999 [9]. After that, cognitive radio has gained a lot of attention. A cognitive radio is the key enabling technology for dynamic spectrum access, and it has various definitions given by different regulatory bodies, such as Federal Communications Commission (FCC) and International Telecommunication Union.

- (Definition by FCC) A Cognitive Radio is a radio that can change its transmitter parameters based on interaction with the environment in which it operates.

- (Definition by ITU) A radio or system that senses and is aware of its operational environment and can dynamically and autonomously adjust its radio operating parameters accordingly.

No matter how the specific definitions are made, the essential concept is the ability to know the environment and the ability to adjust to use vacant resources.

In a cognitive radio network, there are two coexistent networks, i.e. primary network and secondary network. The users belonging to the primary network, called primary users, are licensed users and have a license to access a certain spectrum band. Primary users do not need any modification or additional functions for co-existence with the secondary network. However, the users belonging to the secondary network, called secondary users, are non-licensed users. Therefore, additional functionalities are required to share the spectrum band with primary users. The required tasks form a cognitive cycle as summarized in Figure 2.1.

The cycle starts from the spectrum sensing process. In this part, the secondary users sense the available spectrum bands (named spectrum holes), capture their information, and then detect the vacant spectrum resources. Based on the sensing results, secondary users can decide which resource they want to utilize. Then, spectrum sharing is conducted to prevent collisions between multiple secondary users trying to access the same spectrum. In addition, secondary users should switch to other vacant bands when the primary users need the specific portion of spectrum, because secondary users are *visitors* to the spectrum. This operation is the so-called spectrum mobility.

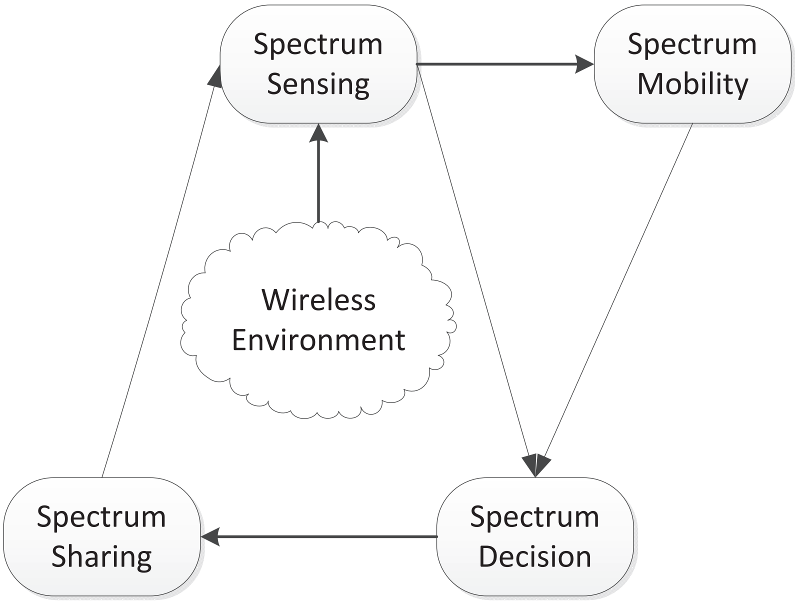


Figure 2.1: Functionalities in a Cognitive Cycle

2.1.1 Coexistence Architectures

In a cognitive radio system, the primary network and the secondary network coexist with each other in one architecture. The basic design principle is that the primary users are as unaffected as possible. In order to fulfill such coexistence, there are three possible solutions: overlay, underlay and interweave.

In the *overlay* architecture, concurrent transmissions between primary and secondary users are allowed. Secondary users have the ability to sense primary users' message, and then use advanced coding schemes (e.g. dirty paper coding [10]) for interference cancelation, so that primary users' transmissions remain unaffected.

The *underlay* architecture also allows simultaneous primary and secondary transmissions. The secondary users spread their signals over a wide bandwidth, which is large enough to ensure that the amount of interference caused to primary transmissions is under the tolerable thresholds. Such interference constraint restricts the usage of underlay architecture to short range communications.

The *interweave* architecture is proposed based on the opportunistic communication. Secondary users have the intelligent to periodically monitor the spectrum, detect the activities of primary users in time and frequency domain, and then opportunistically interweave secondary transmissions through the sensed spectrum holes. In this scenario, accurate spectrum sensing is critical to the performance, especially when Signal to Noise Ratio (SNR) is low.

In this work, the interweave architecture is considered, in which secondary users share the spectrum with primary users by Time Division Multiple Access (TDMA) mode along time axis and by Frequency Division Multiple Access (FDMA) mode in frequency domain.

2.1.2 Spectrum Sensing Techniques

As mentioned in the interweave architecture above, spectrum sensing is of great importance to a cognitive radio network. In reality, however, it is not easy for a secondary user to detect primary users' activity in the absence of interaction between primary users and itself. One of the key challenging issues is the so-called *hidden terminal* problem, which refers to terminals that are out of the range of other terminals. For example, a terminal B at the edge of an access point's range, which is known as A, can see the access point, but it is unlikely that B can see another terminal C on the opposite end of the access point's range, as illustrated by Figure 2.2. These terminals are known as hidden terminals. In order to prevent this problem, it is required that the spectrum sensing sensitivity should outperform primary user's receiver by a large margin. In addition, it is also required that the implementation of the spectrum sensing function has a high degree of flexibility, because it should work in various environments with different types of primary network, bandwidth, frequency, propagation property, interferences and other special characteristics.

A lot of work has been done and many spectrum sensing algorithms have been proposed (e.g. [11–16]). Generally, these algorithms can be classified into three categories: transmitter detection, receiver detection and interference temperature management detection.

In transmitter detection approaches, the detection of primary users is performed based on the received signal at secondary users. The most well-known approaches in this category include Matched Filter (MF) detection [13], Energy detection [17], and Cyclostationary-Feature detection [18]. A matched filter is obtained by correlating a known signal with an unknown signal to detect the presence of the known signal. The MF detection requires less time to achieve high processing gain on the basis of a priori knowledge of the primary user signal, such as the modulation type and order, the pulse shape, and the packet format. This further means, if such information is not accurate, then the matched filter performs poorly. For

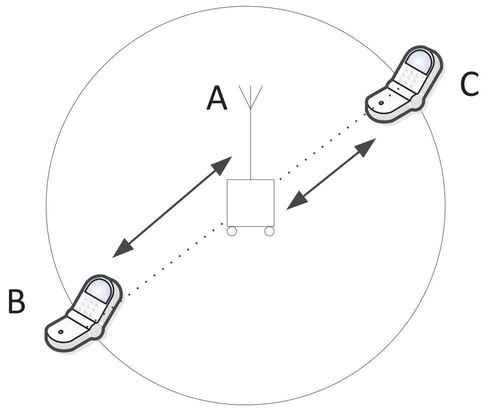


Figure 2.2: An Example of Hidden Terminal

energy detection, its implementation is easy. It can determine the presence of the signal but cannot differentiate signal types, and it does not work for spread spectrum signals. The Cyclostationary-Feature detection has better performance even in low SNR regions, but has higher computational complexity and requires significantly long observation time. With the transmitter detection algorithms, the secondary users cannot avoid the interference due to the lack of primary receivers' information. Moreover, the transmitter detection cannot prevent the hidden terminal problem.

In receiver detection, secondary users detect the Local Oscillator (LO) leakage power for the detection of primary users instead of the transmitted signals. Same methods, such as matched filter, can be used to detect the LO leakage power. The receiver detection can solve the receiver uncertainty problem; however, implementation of a reliable detector is challenging since the LO leakage signal is typically weak.

The interference temperature management detection manages interference at the receiver according to the interference temperature limit, which is represented by the amount of interference that the receiver could tolerate. As far as secondary users do not exceed this limit by their transmissions, they can use this spectrum band. It faces several implementation problems. For example, there is no practical way for a secondary user to measure or estimate the interference temperature at the primary receiver. In addition, the interference temperature limit of a primary user is location dependent, which is not easy to determine.

From the viewpoint of detection behavior, spectrum sensing can be divided into non-cooperative and cooperative detection. In the non-cooperative detection, secondary users conduct detection independently by themselves.

While in the cooperative detection, secondary users share their detection information to improve the detection performance.

In this work, the only aspect that matters is the detection performance without considering which specific detection approach is used. The performance is represented by average sensing error probability in a time slot, which is further divided into mis-detection probability and false alarm probability.

2.2 Network Calculus

Network calculus is a newly developed queueing theory for service guarantee analysis, and it is used to deal with flow problems in networks. The concept of network calculus was firstly introduced in 1990s [19, 20], in which arrival curve and service curve concepts were proposed. During the 20-year's development, network calculus has evolved into two branches: deterministic network calculus and stochastic network calculus. At the same time, some nice properties have been proved, such as output characteristics and concatenation property, on the mathematical basis of min-plus algebra. Book [21] and Book [22] give detailed introduction to deterministic network calculus and stochastic network calculus, respectively.

In this section, a short introduction on min-plus algebra is given. Then, the traffic model and service model in both deterministic network calculus and stochastic network calculus are presented, followed by a summary of several widely-used properties. Lastly, an important approach, i.e. independent case analysis, is described in Section 2.2.6.

2.2.1 Min-Plus Algebra

Min-plus algebra is the mathematical foundation of network calculus. This section will give a brief introduction to min-plus algebra. The definition of min-plus algebra and comparison with traditional algebra are illustrated firstly. Then min-plus convolution and de-convolution as well as their properties are discussed. Thirdly, the vertical and horizontal deviations are expressed, which are two important quantities in network calculus.

In conventional algebra, the two most common operations are addition and multiplication, and the algebraic structure is $(\Re, +, \times)$, that is the set of reals endowed with the two operations of addition and multiplication. Let us change the operations in the following way: addition becomes computation of the minimum and multiplication becomes addition. Besides, we also include $+\infty$ in the set of elements on which min-operations are carried out.

Then we can obtain the min-plus algebraic structure as $(\Re \cup \{+\infty\}, \wedge, +)$, where $\wedge$ denotes the infimum (or, when it exists, the minimum).

Min-plus algebra has some operation properties, which are similar with traditional algebra. Some of the properties are listed below.

- Closure of $\wedge$ and $+$: For all $a, b \in \{\Re \cup \{+\infty\}\}$, $a \wedge b$ and $a + b$ also belong to $\{\Re \cup \{+\infty\}\}$.
- Associativity of $\wedge$ and $+$: For all $a, b, c \in \{\Re \cup \{+\infty\}\}$, $(a \wedge b) \wedge c = a \wedge (b \wedge c)$ and $(a + b) + c = a + (b + c)$.
- Commutativity of $\wedge$: For all $a, b \in \{\Re \cup \{+\infty\}\}$, $a \wedge b = b \wedge a$.
- Idempotency of $\wedge$: For all $a \in \{\Re \cup \{+\infty\}\}$, we have $a \wedge a = a$.
- Existence of a zero element for $\wedge$: There is some $e = +\infty \in \{\Re \cup \{+\infty\}\}$ such that for all $a \in \{\Re \cup \{+\infty\}\}$, we have $a \wedge e = a$.
- The zero element for $\wedge$ is absorbing for $+$: For all $a \in \{\Re \cup \{+\infty\}\}$, we have $a + e = e = e + a$.
- Existence of a neutral element for $+$: There is some $u = 0 \in \{\Re \cup \{+\infty\}\}$ such that for all $a \in \{\Re \cup \{+\infty\}\}$, we have $a + u = u = u + a$.
- Distributivity of $+$ with respect to $\wedge$: For all $a, b, c \in \{\Re \cup \{+\infty\}\}$, we have $(a + b) \wedge c = a \wedge c + b \wedge c$.

For functions in min-plus algebra, there are two important ones that are often used, i.e. min-plus convolution and min-plus de-convolution, defined as follows.

- Min-plus convolution $\otimes$: For any functions a and b, $a \otimes b(x) = \inf_{0 \leq y \leq x}[a(y) + b(x - y)]$.
- Min-plus de-convolution $\oslash$: For any functions a and b, $a \oslash b(x) = \sup_{y \geq 0}[a(x + y) - b(y)]$.

2.2.2 Notations

Before going to the details of stochastic network calculus, the important notations used all over this thesis are described and defined in this section.

$A(t)$ denotes the cumulative amount of traffic generated by an arrival process during period $(s, t]$. $S(t)$ represents the cumulative amount of service that can be provided by a service process. The arrival process is served

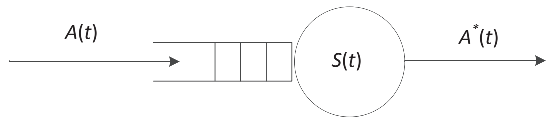


Figure 2.3: System Elements and Notations

by the service process as depicted in Figure 2.3, where the output process is denoted by $A^*(t)$. By definition, the backlog in the system at time t is

$$B(t) = A(t) - A^*(t).$$

For the delay in the system at time t, it is

$$D(t) = inf\{\tau : A(t) \leq A^*(t + \tau)\}.$$

2.2.3 Traffic Model

Arrival curve is defined to describe the characteristics of a traffic flow based on the concept of cumulative arrival process, denoted as $A(t)$, which is the total amount of traffic generated by the flow during time period $[0, t)$. The deterministic arrival curve gives an upper bound on the generated traffic defined as:

Definition 1. ***(Deterministic Arrival Curve.)*** *Given a wide-sense increasing function $\alpha(t)$ defined for $t \geq 0$, we say that a flow A is constrained by α if and only if for all $s \leq t$:*

$$A(t) - A(s) \leq \alpha(t - s),$$

where $A(u)$ denotes the total amount of traffic from flow A during time period $[0, s)$. We say that flow A has α as a deterministic arrival curve.

A typical traffic model that is constrained by a deterministic arrival curve is the Periodic Traffic defined as following.

Example 1. ***(Periodic Source)*** *A periodic source produces an amount of workload, denoted by δ, at times $\{U\tau + n\tau, n = 0, 1, 2, ...\}$, where τ is the* *period time length and U is uniformly distributed on the interval* $[0, 1]$. *It has a deterministic arrival curve as:*

$$\alpha(t) = \delta + \frac{\delta}{\tau} t.$$

For some stochastic arrival processes, deterministic arrival curve is not applicable, because such stochastic processes are not deterministically upper bounded. Stochastic arrival curve is then defined while it has variations, among them the following one is mostly used in this work. This model, originally proposed in [23] and generalized in [24], exploits the *virtual-backlog-property* of deterministic arrival curve [22], and is now also known in the literature as *sample path envelope* [25].

Definition 2. ***(Stochastic Arrival Curve).*** *A flow $A(t)$ is said to have a virtual-backlog-centric (v.b.c) stochastic arrival curve (sac) $\alpha \in F^1$ with bounding function $f \in \bar{F}^2$, denoted by $A(t) \sim_{sac} \langle f, \alpha \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\left\{\sup_{0 \leq s \leq t} \{A(s,t) - \alpha(t-s)\} > x\right\} \leq f(x).$$

For two typical models that are employed in this work, i.e. Poisson and Compound Poisson traffic model, the v.b.c stochastic arrival curves are already obtained as follows.

Example 2. ***(Poisson Traffic.)*** *Suppose all packets of a flow have the same size L and they arrive according to a Poisson process with mean arrival rate λ. Then the flow has a v.b.c stochastic arrival curve $A(t) \sim_{sac} \langle f, \alpha \rangle$ for any $r > \lambda L$ with bounding function [22]:*

$$f(x) = 1 - (1-a)\sum_{i=0}^{k}\left[\frac{[a(i-k)]^i}{i!}e^{-a(i-k)}\right]$$
$$\alpha(t) = rt,$$

where $a = \frac{\lambda L}{r}$ and $k = \lceil \frac{x}{L} \rceil$.

[1] F: the set of non-negative wide-sensing increasing functions
[2] $\bar{F}$: the set of non-negative wide-sensing decreasing functions

Example 3. ***(Compound Poisson Traffic.)*** *Suppose packets of a flow arrive according to a Poisson process with mean arrival rate λ. If the packet lengthes $L_1, L_2, ...$ are independent and exponentially distributed with parameter σ. Then the flow has a v.b.c stochastic arrival curve $A(t) \sim_{sac} \langle f, \alpha \rangle$ for any $\theta_1 \geq 0$ and $\theta > 0$ [26]:*

$$f(x) = e^{-\theta\theta_1} e^{-\theta x} \tag{2.1}$$

$$\alpha(t) = \frac{\lambda t}{\sigma - \theta} + \theta_1 t. \tag{2.2}$$

2.2.4 Server Model

Similar to arrival curve, service curve depicts the properties of a server. Deterministic service curve gives a deterministic bound on the amount of service that can be guaranteed by a server, while a stochastic service curve depicts the probabilistic property of the server with several variations. Here, the definitions for deterministic service curve and stochastic service curve are presented.

Definition 3. ***(Deterministic Service Curve.)*** *Consider a system S and a flow through S with input function $A(t)$ and output function $A^*(t)$. We say that S offers to the flow a deterministic service curve β if β is wide sense increasing, $\beta(0) = 0$ and*

$$A^* \geq A \otimes \beta(t) \triangleq \inf_{0 \leq s \leq t} \{A(s) + \beta(t - s)\},$$

where $\otimes$ denotes the min-plus convolution.

Definition 4. ***(Stochastic Service Curve[3].)*** *A system S is said to provide a stochastic service curve (ssc) $\beta \in F$ with bounding function $g \in \bar{F}$, denoted by $S \sim_{ssc} \langle g, \beta \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\{A \otimes \beta(t) - A^*(t) > x\} \leq g(x). \tag{2.3}$$

[3]The definition given by Eq.(2.3) is called *Weak* stochastic service curve in [22]. For the sake of consistence in this thesis, the name *Stochastic Service Curve* is used here.

Note that: *The deterministic branch can be considered as a special case of the stochastic branch, where the bounding functions are ZEROs for any x. Therefore, the stochastic branch is referred to by default in the remaining parts.*

2.2.5 Properties

This section introduces three basic results of stochastic network calculus, which play important roles in the analysis of this work. The *Backlog Bound* and *Delay Bound* theorems show bounds on the probabilistic distributions of backlog and delay based on stochastic arrival curve and stochastic service curve. The *Leftover Service* theorem tells how to deal with aggregated flows. For conciseness, the theorems are summarized below without detailed deduction steps, which can be found in [22].

Theorem 1. ***(Backlog Bound.)*** *Consider a server S with input A. Suppose that the input has a v.b.c stochastic arrival curve denoted as $A \sim_{sac} \langle f, \alpha \rangle$, and the server provides to the input a stochastic service curve written as $S \sim_{ssc} \langle g, \beta \rangle$, then for any $t \geq 0$ and $x \geq 0$, the backlog $B(t) \equiv A(t) - A^*(t)$ is bounded by*

$$P\{B(t) \geq x\} \leq f \otimes g(x - \alpha \oslash \beta(0)), \tag{2.4}$$

where $\alpha \oslash \beta(x)$ is de-convolution of functions α and β.

Theorem 2. ***(Delay Bound.)*** *Consider a server S with input A. Suppose that the input has a v.b.c stochastic arrival curve denoted as $A \sim_{sac} \langle f, \alpha \rangle$, and the server provides to the input a weak stochastic service curve written as $S \sim_{ssc} \langle g, \beta \rangle$, then for any $t \geq 0$ and $x \geq 0$, the delay of traffic arriving at t, denoted as $D(t) = \inf\{\tau \geq 0 : A(t) \leq A^*(t+\tau)\}$, is bounded by*

$$P\{D(t) \geq h(\alpha + x, \beta)\} \leq f \otimes g(x), \tag{2.5}$$

where $h(\alpha + x, \beta) = \sup_{t\geq 0}\{\inf\{\tau \geq 0 : \alpha(t) + x \leq \beta(t+\tau)\}\}$ is the maximum horizontal distance between $\alpha(t) + x$ and $\beta(t)$.

Theorem 3. ***(Leftover Service.)*** *Consider a system S with input A which is the aggregation of two constituent flows A_1 and A_2. Suppose A_2 has a v.b.c stochastic arrival curve denoted as $A_2 \sim_{sac} \langle f_2, \alpha_2 \rangle$, and the system provides to the aggregated input A a stochastic service curve written as $S \sim_{ssc} \langle g, \beta \rangle$. Then, if $\beta - \alpha_2 \in F$, A_1 receives a stochastic service curve of $S_1 \sim_{ssc} \langle g \otimes f_2, \beta - \alpha_2 \rangle$.*

2.2.6 Independent Case Analysis

As it is noticeable by the definitions of arrival/service curve, they are not unique. Take the deterministic arrival curve as an example: if $\alpha(t)$ is a deterministic arrival curve, then $\alpha(t) + C$ with $C \geq 0$ and $k\alpha$ with $k \geq 1$ can also fulfill the definition. In other words, they are also deterministic arrival curves. Due to this fact, the performance bounds obtained from Theorem 1 and Theorem 2 are also not unique. Then, a key challenge in network calculus analysis is to obtain tight bounds. Unfortunately, it has been shown that the performance bounds in Theorem 1 and Theorem 2 are still not tight in some cases even optimized arrival/service curves are relied on, especially when aggregated flows are involved.

Some efforts have been made to improve the performance bounds, among which an important method is the one named *independent case analysis* [22]. It tries to explore the independence between arrivals based on the concepts of strict server and impairment process.

An impairment process I is a process that its amount during any period will never exceed the amount of service that can be provided by the server during this period. The impairment process does not need/have a buffer: it represents the amount of service that is "impaired".

Definition 5. ***(Strict Service Curve.)*** *A system is said to be a stochastic strict server providing stochastic strict service curve $\beta(t)$ with bounding function $g(x) \in \bar{F}$, denoted by $S \sim_{s-ssc} \langle g, \beta \rangle$, if during any period $(s, t]$, the amount of service $S(s, t)$ provided by the system satisfies $P\{S(s, t) < \beta(t - s) - x\} \leq g(x)$ for any $x \geq 0$.*

The following theorem summarizes results for the independent case, with particular focus on delay.

Theorem 4. ***(Independent Case Analysis).***

(i) Consider a stochastic strict server S providing strict service curve $\hat{\beta}(t)$ with impairment process I. If the impairment process has a stochastic arrival curve of $I \sim_{sac} \langle g, \gamma \rangle$, then the server is a stochastic strict server providing stochastic strict service curve of $S \sim_{s-ssc} \langle g, \beta \rangle$, where $\beta(t) = \hat{\beta}(t) - \gamma(t)$.

(ii) Consider a flow A served by the stochastic strict server described in part (i). Suppose the input has a v.b.c stochastic arrival curve of $A \sim_{sac} \langle f, \alpha \rangle$. If process A and I are independent, the delay $D(t)$ is guaranteed such that, for any $x \geq 0$,

$$P\{D(t) > h(\alpha + x, \beta)\} \leq 1 - \bar{f} * \bar{g}(x),$$

where $\bar{f}(x) = 1 - [f(x)]_1$, $\bar{g}(x) = 1 - [g(x)]_1$, $[\cdot]_1 = \max\{0, \min\{1, \cdot\}\}$ and $\bar{f} * \bar{g}(x) \equiv \int_0^x \bar{f}(x-y) d\bar{g}(y)$.

2.3 Performance Analysis of a Cognitive Radio Network

In this section, modeling of the considered cognitive radio network is made, where several key aspects are taken into account, including system model, channel model, sensing error model, and re-transmission schemes. The research results mainly focus on probabilistic backlog distribution, probabilistic delay distribution and delay-constrained capacity.

2.3.1 Modeling of the Considered Cognitive Radio Network

In a cognitive radio network, secondary users try to make use of the spectrum when the primary users are away from the spectrum. Secondary users should vacant the spectrum when primary users need to use the spectrum resource. Due to this, a cognitive radio network is naturally modeled as a priority system with two classes of inputs.

The spectrum access decision of secondary users is made based on spectrum sensing results as depicted in Figure 2.1. However, spectrum detection is not always reliable, errors may happen sometimes. Spectrum sensing errors can be classified into two types, i.e. mis-detection and false alarm, which have different impacts on users. A mis-detection happens when a primary user is actually occupying the spectrum but the sensing result tells a secondary user that the spectrum is vacant. Then collision between a primary user and a secondary user will occur, which leads to both transmission failures. In a false alarm, the spectrum is available for a secondary user while the sensing result indicates the spectrum is occupied. Therefore, secondary users will miss the transmission opportunity. The impacts of mis-detection and false alarm should be properly modeled, so that the system performance can be better evaluated. In this work, sensing error processes are modeled by the concept of *wasted service*.

Re-transmission is a normally employed scheme to compensate for transmission failures, and its strategy has particular impacts on delay and loss probability. Generally, more re-transmission attempts will lead to more reliable transmission and longer delay. In this work, three typical schemes are considered and discussed: without re-transmission (WO-RT), re-transmission until success (RT-S) and maximum-N-time re-transmission (Max-N-RT). For WO-RT scheme, a packet will be cleared from the queue when it is transmitted out, no matter it is received by the receiver or not. For the RT-S scheme, a packet will be re-transmitted until the receiver sends back the acknowledge (ACK) signal. In the Max-N-RT scheme, a packet can be re-transmitted for N times at most, and then it will be removed from the queue. Based on the working principle, WO-RT scheme has better delay guarantee, RT-S provides more reliable transmission, where the Max-N-RT scheme is a tradeoff between delay and loss probability.

For the spectrum sharing policy, it is assumed that there exists a virtual central control point to coordinate the transmissions from all primary users in a First-In-First-Out (FIFO) manner, and another point to coordinate all secondary users by FIFO. Therefore, no collision will happen between two primary users (or between two secondary users). This assumption may be too ideal in practical networks, however, it helps to obtain tractable performance bounds.

In wireless networks, channel fading is also an important aspect to be modeled. In this work, a constant rate channel without fading is considered at the beginning. Then, the two-state Gilbert-Elliott channel is employed to study the impacts of fading, which can also be considered as a constant channel with an impairment process. A GE channel has two states, i.e. state *ON* and state *OFF*, and the channel transfers between these two states according to a Markov chain as shown in Figure 2.4.

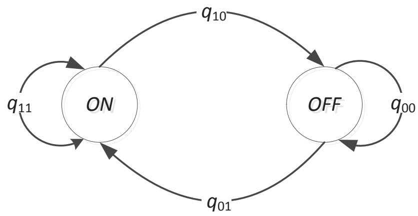


Figure 2.4: GE Channel State Transition

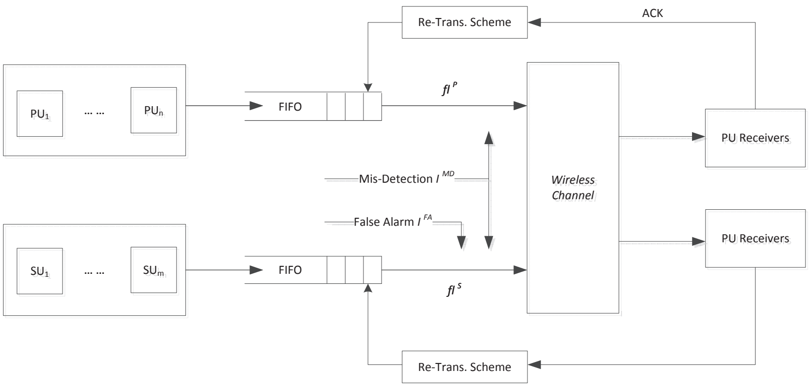


Figure 2.5: Considered Network Model

In state ON (state 1), data can be received correctly, while in state OFF (state 0), the channel is in deep fading and no data can be successfully received at the receiver. The transition rate from state $i(i \in \{0, 1\})$ to state $j(j \in \{0, 1\})$ is represented by q_{ij}.

Figure 2.5 summarizes the considered cognitive radio network. Two types of users are organized into two independent FIFO buffers. Mis-detection and false alarm have their impacts on two flows. Packets are transmitted through the wireless channel, and then receivers send back their ACK[4] signals, which are used by re-transmission scheme to decide how to schedule the next packet waiting in the queue.

2.3.2 Output Metrics

In this work, performance of a cognitive radio network is evaluated by network calculus. Main output is given in the form of Complimentary

[4]ACK delay is ignored in the analysis

Cumulative Distribution Function (CCDF) bound, which shows the upper bound on the probability that a real-valued random variable X exceeds a threshold x. Network calculus provides theoretical basis to obtain the CCDF bounds for delay and backlog as given in Theorem 2 and Theorem 1, i.e. $Pr\{D > x\} \leq f_D(x)$ and $Pr\{B > x\} \leq f_B(x)$.

The CCDF bounding function $f_D(x)$ is closely related with several aspects, and their impacts on the network performance are studied, such as characteristics of input traffic, channel condition, channel fading speed, spectrum sensing error probability, and re-transmission scheme.

Delay-constrained capacity is also an important evaluation aspect in this work, which is defined as the maximum arrival rate of input traffic when required delay and its violation probability are still met. Let d denote the delay threshold and ϵ denote the maximum violation probability. Then, the delay-constrained capacity C can be expressed as

$$C = max\,\{R|Pr\{D(R) > d\} \leq \epsilon\}$$

where R represents the arrival rate and its value will influence the delay distribution $D(R)$.

Furthermore, in order to validate the theoretical analysis, system-level simulation platform is established to obtain the corresponding simulation results for comparison with numerical results mentioned above.

Chapter 3

Publications and Contributions

This thesis consists of six publications, including one book chapter and five academic papers, which are attached in Chapter A to Chapter F. Each of these publications has its particular concern and discusses different key aspects in the considered cognitive radio network. The thesis author played an active role in research and writing these papers under the supervision of Prof. Yuming Jiang. Xin Zhang also contributed to the work presented in Publication C and D. Tao Lin, Jinxing Yang and Wenting Jiang respectively contributed to the running of numerical calculation and simulation in Publications C, Publication D and Publication F. In the following, a brief summary of these publications is presented with short descriptions on the authors' contributions.

3.1 List of Publications Included in This Thesis

3.1.1 Publication A

- Yuehong Gao and Yuming Jiang; Advanced Cognitive Radio Network: Chapter 4 Spectrum Allocation; Scientific Research Publishing; ISBN: 978-1-935068-74-7; Sept 2011.

In this book chapter, we study spectrum allocation in cognitive radio networks, which has the aim of making efficient spectrum utilization. An introduction to the various policies in spectrum sharing is studied first. Then, the concept of spectrum pooling that represents the idea of merging spectral ranges from possibly different spectrum owners is introduced. After these, the focus is put on performance aspects of spectrum allocation. The application of game theory to spectrum allocation is also introduced, because it provides a well-defined model to describe conflict and cooperation among intelligent rational decision makers, which has a natural match to spectrum allocation in cognitive radio networks. Following this, spectrum utilization efficiency is discussed in more detail, including various measures in evaluating spectrum utilization efficiency. Before summarizing the chapter with highlighting directions for further research in spectrum allocation, spectrum allocation algorithms and performance analysis of them are reviewed.

3.1.2 Publication B

- Yuehong Gao and Yuming Jiang; Performance Analysis for a Cognitive Radio with Imperfect Spectrum Sensing; IEEE INFOCOM 2010

Workshop on Cognitive Wireless Communications and Networking, March 2010.

In this paper, a cognitive radio network with a single constant wireless channel is studied, where imperfect spectrum sensing and re-transmission schemes are considered. The concept of wasted service process is used for modeling the mis-detection process, false alarm process and service process under three different re-transmission schemes, i.e. without re-transmission (WO-RT), re-transmission until success (RT-S) and maximum-N-time re-transmission (Max-N-RT). Then, stochastic service curve for primary flow and secondary flow is derived on the basis of interference process as shown in Theorem 2 in the paper. Two types of traffic, (σ, ρ)-constrained traffic and Poisson traffic, are used as two examples to obtain numerical results. The backlog and delay distribution bounds are compared and discussed under different configurations.

3.1.3 Publication C

- Yuehong Gao, Yuming Jiang, Tao Lin and Xin Zhang; Performance Bounds for a Cognitive Radio Network with Network Calculus Analysis; 2010 International Conference on Network Infrastructure and Digital Content, September 2010.

In this paper, the considered cognitive radio network is simplified, where perfect spectrum sensing and constant rate channel are assumed. Two approaches, i.e. min-plus convolution and independent case analysis, are applied for the analysis. Min-plus convolution means that the distribution bound is obtained after a min-plus convolution between bounding functions. The independent case analysis tries to explore the independence between arrival process and service process, so that the distribution bound can be improved. One pre-condition of applying independent case analysis is that the functions should be differentiable. Thus, two stochastic arrival curves for Poisson traffic are used. Theoretical conduction is made for each approach, and the mathematical expressions are summarized in Theorem 3 in the paper. Delay distribution bound and backlog distribution bound of primary flow and secondary flow are obtained both by numerical calculation and system level simulation. It is shown that these two methods have the same results for the primary flow, while independent case analysis will significantly improve the bounds for the secondary flow.

3.1.4 Publication D

- Yuehong Gao, Jinxing Yang, Xin Zhang and Yuming Jiang; Capacity Limits for a Cognitive Radio Network under Fading Channel; Springer Lecture Notes in Computer Science: IFIP Networking 2011 workshops, pp. 42 - 51, May 2011.

In this paper, spectrum sensing errors and channel fading are considered. Spectrum sensing errors are further divided into mis-detections and false alarms, and their stochastic arrival curves are obtained. The wireless channel is modeled as a two-state Gilbert-Elliott channel, and its stochastic service curve is derived. After that, delay distribution bound is obtained by stochastic network calculus. A Long Term Evolution (LTE) system using Orthogonal Frequency Division Multiplexing (OFDM) technology is relied on to conduct numerical calculations. Two types of QoS requirement are studied, i.e. Voice over IP (VoIP) and Transmission Control Protocol (TCP). VoIP has strict requirement on delay budget but it can tolerant some loss; while TCP requires low loss probability and has larger delay budget. Then, delay-constrained capacity is defined based on delay budget and loss probability. The capacity limit of primary flow and the capacity limit of secondary flow under different traffic load are presented and discussed.

3.1.5 Publication E

- Yuehong Gao and Yuming Jiang; Analysis on the capacity of a cognitive radio network under delay constraints; IEICE Transactions on Communications; Vol. E95-B, No. 04, 2012.

This paper further extends the work in Publication D with more detailed deductions and discussions. Delay-constrained capacity is defined for primary flow and secondary flow respectively, which is decided by the maximum traffic rate. Furthermore, specific expressions of delay-constrained capacity are given for Poisson traffic and Periodic traffic. Both numerical results and simulation results are compared and discussed. Capacity limits for primary flow are listed, and it is found that numerical results approach to the simulation results, which validates the analysis. The capacity of secondary flow relies on the traffic load from primary flow, therefore, a capacity region is depicted with x-axis as the PU's traffic load. The impacts of mis-detections are studied under different MD probabilities. It is obvious that higher MD probability will reduce the capacity. How the channel fading will influence the capacity is investigated from the viewpoint of fading speed. It is found that the capacity limits are reduced when the fading is slow. In addition, the gap between theoretical results and simulation results increases under slow fading, especially when the primary network is fed with heavy load.

3.1.6 Publication F

- Yuehong Gao, Wenting Jiang and Yuming Jiang; Guaranteed Service and Delay-Constrained Capacity of a Multi-Channel Cognitive Secondary Network; 7th International Conference on Cognitive Radio Oriented Wireless Networks (CrownCom); June 2012.

In this paper, the considered network has multiple parallel channels instead of only one channel, and all of these channels are shared by primary users and secondary users. By assuming that a certain amount of resource is exclusively reserved and used on each channel by the primary network, we derive the traffic transportation capacity that is guaranteed to the secondary network. Then, the traffic delay distribution bound and the guaranteed capacity of the secondary network in serving traffic with probabilistic delay requirement are got. Both numerical and simulation results are obtained, where the secondary network traffic follows a model taken from an LTE system. The results are compared and discussed by delay distribution, average delay and delay-constrained capacity. The good match between numerical and simulation results validates the analysis.

3.2 Summary of Publications and Contributions

All of the publications focus on the performance analysis of cognitive radio networks with different emphases. Publication A introduces fundamental principles of a cognitive radio network with particular description on spectrum allocation mechanism. Publications B to E investigate the single-channel scenario, where Publication B studies the impacts of retransmission scheme and spectrum sensing errors, Publication C compares independent case analysis with min-plus convolution, Publication D includes channel fading, and Publication E gives deep discussions based on the results in Publications B to D. Multiple-channel scenario is considered in Publication F. Delay and backlog distribution bounds are the main outcomes in Publications B and C, while delay-constrained capacity is evaluated in Publications D to F.

The main contributions of this thesis are now summarized as follows:

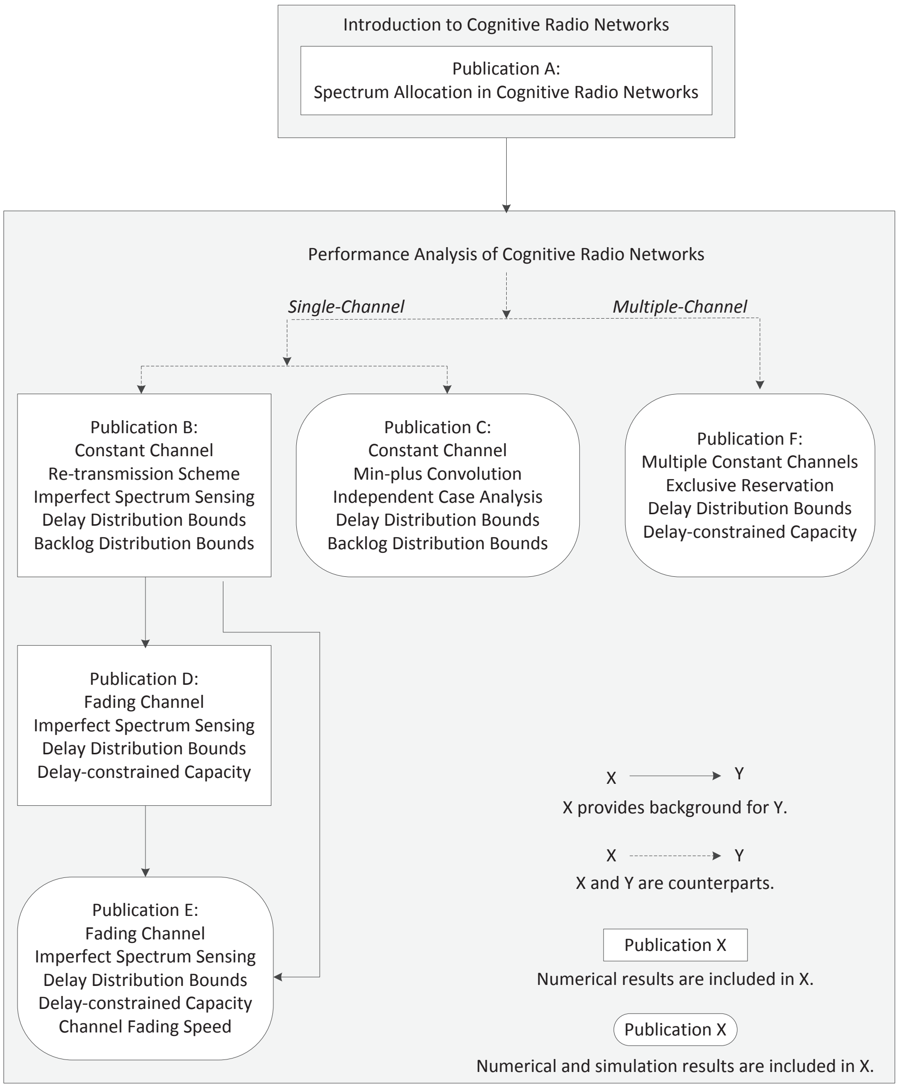


Figure 3.1: Connections of Included Publications

- Network calculus is relied on for performance analysis of a cognitive radio network, which extends the application of network calculus into wireless networks. At the same time, more results about the probabilistic distributions of delay and backlog are discussed, instead of average values in most literatures.

- Characteristics of spectrum sensing error process are studied and divided into mis-detection and false alarm. The arrival processes of mis-detection and false alarm are modeled, where the stochastic arrival curves are derived. Their impacts on the system performance are obtained both by numerical calculation and simulation.

- Three typical re-transmission schemes, i.e. WO-RT, RT-S and Max-N-RT, are modeled, and corresponding stochastic arrival curves are derived based on the concept of wasted service. The performances under different re-transmission schemes are studied by numerical calculations.

- In the single channel scenario, the Gilbert-Elliott fading channel is considered, and its stochastic service curve is derived. The impact of fading speed on the network's performance is studied by numerical and simulation results.

- The service guarantee provided to each type of users is derived based on the Leftover Service property. For the primary flow, mis-detection and channel fading are two aspects that influence its service guarantee. While for the secondary flow, the arrival of primary flow is the most significant factor to be considered. Specific expressions of stochastic service curves are obtained and summarized under different configurations.

- The research results not only focus on the probabilistic backlog and delay distribution bounds, but also investigate the delay-constrained capacity, which is defined as the maximum arrival rate when a certain probabilistic delay requirement (d, ϵ) can still be met. The obtained capacity or capacity region provides reference for the network to conduct admission control.

- Independent case analysis is studied and relied on for performance analysis, where the special requirement on bounding function is also discussed. Its results are compared with the results got by min-plus convolution. The presented results validate that independent case analysis can improve the probabilistic distribution bound significantly, and its results coincide with simulation results in particular cases.

- A cognitive radio network with multiple channels is modeled and studied by assuming exclusive resource reservation for primary users. The upper bound and lower bound on the service guarantee for secondary users are obtained. Mathematical expressions for the delay-constrained capacity is also derived. Numerical results and simulation results are obtained, and their similarity validates the theoretical analysis.

- Numerical calculation and simulation platforms are established for a cognitive radio network using LTE parameter setting.

3.3 List of Publications not Included

During the PhD study time, the thesis author also contributed to the following papers. Some of them (i-iv) focus on the radio resource management and performance evaluation issues of the 3rd Generation (3G) systems. Paper (v) proposes a new cooperative spectrum sensing scheme. Paper (vi) extends the application of network calculus to analyze the energy consumption problem for wireless terminals. These papers are not included in this thesis, but the work therein has provided in-depth knowledge to conduct the research work and to write this thesis.

(i) Yuehong Gao, Li Chen, Xin Zhang and Yuming Jiang; Performance Evaluation of Mobile WiMAX with Dynamic Overhead; Proceedings of 2008 IEEE 68th Vehicular Technology Conference (VTC 2008-Fall), Calgary Canada, 2008.

(ii) Yuehong Gao, Xin Zhang, Dacheng Yang and Yuming Jiang; Unified Simulation Evaluation for Mobile Broadband Technologies; IEEE Communications Magazine, March 2009, Page(s): 142–149.

(iii) Yuehong Gao, Xin Zhang and Yuming Jiang; An Evaluation Model for Call Admission Control Scheme in CDMA System; Proceedings of 2009 International Forum on Information Technology and Applications; Chengdu China, May 2009.

(iv) Yuehong Gao, Xin Zhang, Yuming Jiang and Jeong-woo Cho; System Spectral Efficiency and Stability of 3G Networks: A Comparative Study; Proceedings of IEEE International Conference on Communications (ICC), Dresden Germany, June 2009.

(v) Qun Pan, Xin Zhang, Yuehong Gao, Dacheng Yang and Yuming Jiang; Efficient Quantification Using Local Information for Cooperative Spectrum Sensing Cognitive Radio; Proceedings of the International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), Cannes France, June 2010.

(vi) Yuehong Gao and Yuming Jiang; Analysis on the Battery Lifespan of a Mobile Terminal under Probabilistic Delay Constraint; Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet (e-Energy 2012); Article No. 11, Madrid Spain, May 2012.

Chapter 4

Research Issues for Future Work

In this work, network calculus is applied to conduct performance analysis for a cognitive radio network. Although several key aspects are considered and the corresponding results are obtained, there are still some interesting issues left for future study.

- In the later part of this thesis work, which consists of Publication D and Publication E, the fading channel is modeled by a two-state Gilbert-Elliott channel. This model is simple to be applied since it has only two states, *ON* with constant rate R_1 ($R_1 > 0$) or *OFF* with constant rate $R_0 = 0$. However, it is too rough to catch the characteristics of a real fading channel, where the channel supports more that two rate levels. Therefore, advanced channel model, such as Finite State Markov Channel (FSMC) [27], can be studied in the future, and the possible difficulty may be how to obtain the stochastic service curve for such a multiple-rate fading channel. The authors of [28] present an approach to map a Rayleigh channel to a GE channel. The work in [29] studies statistical properties of the capacity of mobile fading channels in various wireless communication systems, and the results therein may be useful.

- In this work, it is assumed that there exists a virtual central control point to coordinate transmissions from the same type of users, and all packets will be served in a FIFO manner without considering any fairness or QoS requirement. Hence, more scheduling mechanisms can be studied in the future. In addition, this issue may be connected with channel model, since scheduling algorithm in a wireless network is usually linked with channel fading properties. In [30], some results have recently been obtained along this direction.

- In the theoretical analysis using network calculus, no restrict is assumed for input traffic models as far as their stochastic arrival curves can be found. However, Poisson traffic and Periodic traffic are mainly used in this work in order to simplify the numerical calculation. Other types of traffic can be evaluated in future work. Some typical models have been summarized in [22, 31].

- Further study on independent case analysis is needed, especially when the convolution operation involves several bounding functions. In addition, more scenarios can be considered to further validate possible improvements led by applying independent case analysis.

- Multiple parallel channels are considered in this work under the assumption of periodical and exclusive reservation for primary users along each channel. This assumption is made in order to obtain the guaranteed service provided to secondary users, and it greatly restricts the generality of the analysis. The biggest challenge on the way to remove this assumption is how to derive the stochastic service guarantee of parallel servers, and this blank area is waiting to be filled.

Bibliography

[1] I. Suliman and J. Lehtomaki. Queueing analysis of opportunistic access in cognitive radios. In *Proc. 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management*, pages 153–157, Aalborg, Denmark, May 2009.

[2] C. Ghosh, S. Pagadarai, D. P. Agrawal, and A. M. Wyglinski. Queuing theory representation and modeling of spectrum occupancy employing radio frequency measurements. In *Proc. IEEE Vehicular Technology Conference (VTC 2009-Fall)*, pages 1–5, Anchorage, USA, September 20–23, 2009.

[3] C. Zhang, X. Wang, and J. Li. Cooperative cognitive radio with priority queueing analysis. In *Proc. IEEE International Conference on Communications (ICC '09)*, pages 1–5, Dresden, Germany, June 14–18, 2009.

[4] C. An, H. Ji, and P. Si. Dynamic spectrum access with qos provisioning in cognitive radio networks. In *Proc. IEEE Global Communications Conference*, pages 1–5, Miami, USA, December 2010.

[5] X. Gelabert, O. Sallent, J. Perez-Romero, and R. Agusti. Spectrum sharing in cognitive radio networks with imperfect sensing: A discrete-time markov model. *Computer Networks*, 54:2519–2536, April 2010.

[6] Y. R. Kondareddy, N. Andrews, and P. Agrawal. On the capacity of secondary users in a cognitive radio network. In *IEEE Sarnoff Symposium*, pages 1–5, Princeton, USA, March 2009.

[7] Online. *http://en.wikipedia.org/wiki/Spectrum_allocation*.

[8] Federal Communications Commission Spectrum Policy Task Force. Report of the spectrum efficiency working group. November 2002.

[9] J. III Mitola and G. Q. Jr. Maguire. Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6:13–18, August 1999.

[10] M. Costa. Writing on dirty paper. *IEEE Transactions on Informatio Theory*, 29:439–441, May 1983.

[11] M. Subhedar and G. Birajdar. Spectrum sensing techniques in cognitive radio networks: A survey. *International Journal of Next-Generation Networks (IJNGN)*, 3(2):37–51, June 2011.

[12] D. Cabric, S. M. Mishra, and R. W. Brodersen. Implementation issues in spectrum sensing for cognitive radios. In *Proc. 38th Asilomar Conference on Signals, Systems and Computers*, pages 772–776, November 2004.

[13] D. Cabric, A. Tkachenko, and R. W. Brodersen. Spectrum sensing measurements of pilot, energy, and collaborative detection. In *Proc. IEEE Military Commnication Conference*, pages 1–7, October 2006.

[14] G. Ganesan and Y. Li. Cooperative spectrum sensing in cognitive radio: Part I: Two user networks. *IEEE Transactions on Wireless Communications*, 6(6):2204–2213, November 2006.

[15] G. Ganesan and Y. Li. Cooperative spectrum sensing in cognitive radio: Part II: Multiuser networks. *IEEE Transactions on Wireless Communications*, 6(6):2214–2222, November 2006.

[16] A. Ghasemi and E. S. Sousa. Collaborative spectrum sensing for opportunistic access in fading environment. In *Proc. IEEE Dynamic Spectrum Access Networks (DySPAN)*, pages 131–136, November 2005.

[17] M.-S Alouini F. F. Digham and M. K. Simon. On the energy detection of unknown signals over fading channels. In *Proc. IEEE International Conference on Communication (ICC)*, pages 3575–3579, May 2003.

[18] B. Deepa. *Design and Analysis of Cyclostationarity Based Spectrum Sensing for Cognitive Radio.* M. E. Thesis, Indian Institute of Science, May 2009.

[19] R. L. Cruz. A calculus for network delay, part I: Network elements in isolation. *IEEE Transactions on Information Theory*, 37(1):114–131, January 1991.

[20] R. L. Cruz. A calculus for network delay, part II: Network analysis. *IEEE Transactions on Information Theory*, 37(1):132–141, January 1991.

[21] J-Y. Le Boudec and Patrick Thiran. *Network Calculus – A Theory of Deterministic Queuing Systems for the Internet*. Springer-Verlag, May 2004.

[22] Y. Jiang and Y. Liu. *Stochastic Network Calculus*. Springer, 2008.

[23] Q. Yin, Y. Jiang, S. Jiang, and P. Y. Kong. Analysis on generalized stochastically bounded bursty traffic for communication networks. In *Proc. 27th Annual IEEE Conference on Local Computer Networks (L-CN02)*, 2002.

[24] Y. Jiang, Q. Yin, Y. Liu, and S. Jiang. Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks. *Computer Networks*, 53(12):2011–2021, 2009.

[25] M. Fidler. A survey of deterministic and stochastic service curve models in the network calculus. *IEEE Communications Surveys and Tutorials*, 12(1):59–86, 2010.

[26] Yuming Jiang. A note on applying stochastic network calculus. *Online: http://q2s.ntnu.no/j̃iang/publications/note-on-applying-snetcal-v20100501.pdf*, May 2010.

[27] H.S. Wang and N. Moayeri. Finite-state markov channel - a usuful model for radio communication channels. *IEEE Transactions on Vehicular Technology*, pages 163–171, 995.

[28] L. Liu, P. Parag, J. Tang, W-Y. Chen, and J.-F. Chamberland. Resource allocation and quality of service evaluation for wireless communication systems using fluid models. *IEEE Transactions on Information Theory*, 2007.

[29] G. Rafiq. *PhD Thesis: Statistical Analysis of the Capacity of Mobile Radio Channels*. Doctoral Dissertation by the University of Agder, 2011.

[30] K. Zheng, F. Liu, L. Lei, C. Li, and Yuming Jiang. Stochastic performance analysis of a wireless finite-state markov channel. *Submitted for publication*, 2011.

[31] F. Kelly. Notes on effective bandwidths. *Royal Statistical Society Lecture Notes Series*, 4:141–168, 1996.

Part II

Included Publications

Appendix A

Publication A

Yuehong Gao and Yuming Jiang; Advanced Cognitive Radio Network: Chapter 4 Spectrum Allocation; Scientific Research Publishing; ISBN: 978-1-935068-74-7; Sept 2011.

A.1 Introduction

With the development of wireless communication technologies, the requirements for high speed data services also increase dramatically. This makes the scarcity of wireless spectrums, which are suitable for wireless communications, become more and more significant. Therefore, how to enhance the utilization of limited frequency bands has attracted a lot of research attention. The proposal of cognitive radio technology provides a possible solution.

Cognitive radio networks can better utilize the scarce spectrum via heterogenous system architecture and flexible spectrum access techniques. In a cognitive radio network, secondary users (SUs) can sense and access the available spectrum holes. However, due to the stochastic appearances and departures of the primary users (PUs), the positions (time and/or bandwidth) of available spectrum holes will change. In addition, the wireless channel capacity also varies and different users/sessions may have distinguish Quality-of-Service (QoS) requirements. All these factors lead to the vast attention paid on spectrum allocation in cognitive radio networks, which has the aim of making efficient spectrum utilization.

In this chapter, we start with an introduction to the various policies in spectrum sharing. Then, we introduce the concept of spectrum pooling, which represents the idea of merging spectral ranges from possibly different spectrum owners, and the idea of spectrum sharing with spectrum pooling. After these, we make our focus on performance aspects of spectrum allocation. Game theory provides a well-defined model to describe conflict and cooperation among intelligent rational decision makers, which has a natural match to spectrum allocation in cognitive radio networks. Game theory basics and its applications to spectrum allocation will be introduced. Following this, we discuss spectrum utilization efficiency in more detail, including various measures in evaluating spectrum utilization efficiency. Before summarizing the chapter with highlighting directions for further research in spectrum allocation, we review spectrum allocation algorithms and performance analysis of them.

A.2 Spectrum Sharing

In a cognitive radio network, multiple secondary users may compete to access the available spectrums, and therefore, access coordination of transmission attempts between secondary users is needed in order to prevent collisions. Spectrum sharing is proposed as a solution to this issue, and it has similar functionalities as the Media Access Control (MAC) layer protocol in traditional wireless networks. However, due to the unique characteristics of cognitive radio networks, such as dynamic changes of available spectrum resources and co-existence with primary users, the spectrum sharing policy in cognitive radio networks faces many new challenges. A lot of research effort has been made in this topic, and the existing proposed spectrum sharing algorithms can be classified in the following four ways from different perspectives [1].

- *Centralized and Distributed Spectrum Sharing.* This classification is made based on the network architecture. In centralized spectrum sharing, there exists a central point, which collects the information about available spectrums and controls the spectrum allocation and access. In this case, different requirements from different users can be better considered and coordinated by the central point. In distributed spectrum sharing, spectrum allocation is performed by each node distributively.

- *Cooperative and Non-cooperative Spectrum Sharing.* From the aspect of allocation behavior, spectrum sharing policies can be divided into cooperative and non-cooperative policies. In the cooperative mode, users cooperate with each other so that the impact of interference can be evaluated and controlled. At the same time, frequent message exchanges are needed, which will occupy some system resources. Compared with the cooperative mode, message exchanges are not required in the non-cooperative mode, where users operate only based on their own profits. However, the non-cooperative mode may lead to reduced utilization efficiency due to lacking cooperation.

- *Overlay and Underlay Spectrum Sharing.* In this classification, the adopted co-existence method between primary and secondary users is the basis to differentiate spectrum sharing algorithms. In overlay spectrum sharing, secondary users can access some part of the spectrum when primary users do not use that part of spectrum, which can avoid introducing interference to the primary users. On the contrast, the spread spectrum technique is used in the underlay method, where the whole spectrum can be used by secondary users with the disadvantage of introducing noise to the primary users.

- *Intra-network and Inter-network Spectrum Sharing.* Based on the range of spectrum sharing, we can define intra-network and inter-network policies. When the spectrum sharing policy only occurs among the secondary users within the same network, it is called an intra-network policy. In contrast, inter-network spectrum sharing can

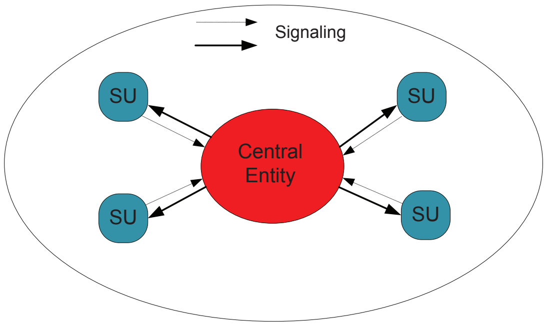


Figure A.1: Centralized Spectrum Sharing

happen between the users of different networks deployed in overlapping areas and spectrums.

In this section, spectrum sharing policies will be discussed in detail from these four perspectives.

A.2.1 Centralized and Distributed Spectrum Sharing

The key difference between a centralized and a distributed scheme is if there exists a central entity that performs spectrum allocation in the deployed area. In the centralized scheme, while the measurement and sensing procedures are performed decentralizedly by secondary users, the central entity collects the feedback information from each secondary user, based on which the spectrum allocation decision is constructed, as illustrated in Fig.A.1. The advantage is that optimal scheduling can be implemented by the central entity since it knows the characteristics of each node through some signaling.

In [2], the concept of Spectrum Server is introduced, which coordinates the transmissions of a group of links sharing a common spectrum by finding an optimal schedule that maximizes the average sum rate subject to a minimum average rate constraint for each link. In [3], a centralized method for managing and coordinating spectrum access, which is called Dynamic Spectrum Access Protocol, is proposed. This method enables lease-based dynamic spectrum access through a coordinating central entity and allows efficient resource sharing and utilization in wireless environments. In [4], the competition for both spectrum and customers is considered under the regulation of a spectrum policy server, which mediates spectrum sharing among communicating devices, and also monitors the relevant spectrum.

On the contrary, in the distributed scheme, secondary users participate in spectrum allocation by themselves equally. This approach reduces the complexity of system deployment while the message exchanges between users may occupy considerable amount of system resource. Much research effort has been put on distributed spectrum sharing, such as in [5–8]. In [5], the authors present the design and realization of a game theory based solution, which is called price-based distributed spectrum sharing (PDSS) to achieve optimal spectrum utilization. In [6], the multi-channel scenario is considered and a Distributed Multichannel Spectrum Sharing algorithm is proposed to approach the optimal Nash Equilibrium. The distributed media access strategy with partial spectrum information is discussed from a game theoretical view in [7]. A busy-burst signaling based dynamic spectrum assignment algorithm is proposed in [9], which works in a distributed and short term manner. In [8], the authors focus on the Quality of Service (QoS) issue by proposing a distributed algorithm which only requires local feedback information.

Comparisons between the centralized and distributed policies have been made in [10–12]. In [10], a collaborated distributed approach is proposed, and it is shown that the distributed approach can provide similar QoS compared with a centralized algorithm using global information. A detailed case study is conducted in [11], where several typical scenarios are considered and discussed. Ref. [12] presents a survey of centralized and distributed spectrum management approaches and discusses technical challenges and remaining open issues.

A.2.2 Cooperative and Non-cooperative Spectrum Sharing

This classification is made according to the relationship between secondary user, to be specific, the criteria is whether they share their spectrum sensing information with the others or not.

In non-cooperative spectrum sharing, secondary users make spectrum access decisions by themselves without exchanging information with other users. In [13], by considering different QoS demands and spectrum characterization parameters, the authors propose a demand matching algorithm based on game theory for a non-cooperative cognitive radio network, which enables secondary users to make spectrum access decision distributively and to access multiple appropriate spectrum bands. In [14], a spectrum sharing problem in an unlicensed band is considered, where results, proved to be tight and quantify the best achievable performance in a non-cooperative scenario, are presented.

Compared with non-cooperative algorithms, cooperative (or coordinat-

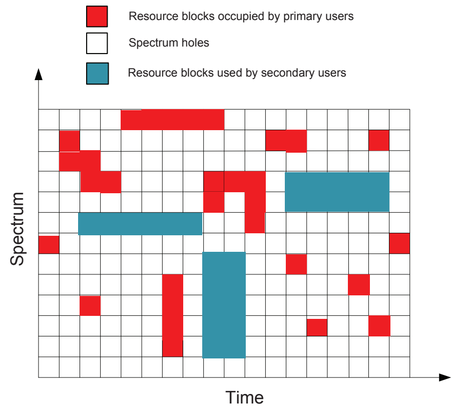


Figure A.2: Overlay Spectrum Sharing

ed) spectrum sharing has gained more attention, because via cooperation, the secondary users can share more information about the environment and better performance can then be provided. One common method is to deploy agents on the devices, which can cooperate to improve spectrum utilization. It is shown that multi-agent based cooperative spectrum sharing can achieve up to 80% of the whole utility [15]. Different design aims can be set for spectrum sharing, such as to reduce the transmission power or to reduce the interference, and corresponding techniques are required in order to fulfill such aims [16]. Although coordination can improve the performance, it may result in large amount of coordination delay and overhead [17].

A.2.3 Overlay and Underlay Spectrum Sharing

In overlay spectrum sharing, secondary users access the spectrum holes opportunistically, and they transmit on some frequency bands only when they do not find any primary transmission on those bands, as shown in Fig.A.2. In this way, overlapping transmission with primary users can be avoided, and hence the interference introduced to the primary users is minimized. However, information about primary users' location and transmission is required, which may be obtained through spectrum sensing. In underlay spectrum sharing, secondary users spread their signals across the available bandwidth, which allows simultaneous transmission of both primary and secondary users, as illustrated in Fig.A.3. With underlay spectrum sharing, the interference caused to primary users can be treated as wide-band background noise. As far as the noise level is under a tolerable threshold, the primary transmissions can be guaranteed.

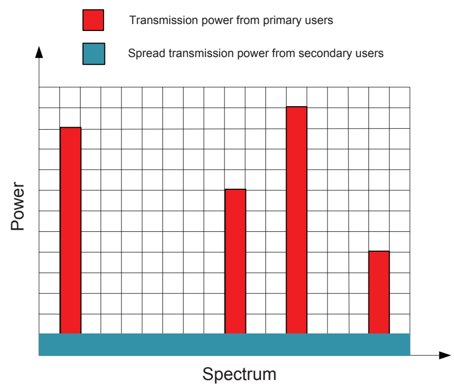


Figure A.3: Underlay Spectrum Sharing

Much research has been conducted on these two types of algorithms from different aspects. The power allocation problem is studied in a joint spectrum overlay and underlay cognitive radio network in [18, 19]. Joint power and rate adaptation is considered in [20–22]. A game theoretical rate allocation scheme for an overlay cognitive radio network is proposed in [23]. In [24], outage probability is used as the evaluation matrix to compare the performance of different spectrum sharing policies, including spreading-based underlay and interference avoidance based overlay. Multi-antenna technique has also been considered for both overlay and underlay scenarios [25]. The comparisons and improvements between different overlay and underlay schemes have been covered by [24, 26–28].

A.2.4 Intra-network and Inter-network Spectrum Sharing

When the practical system deployment is considered, sometimes one area is covered by a single network, while it also happens that several networks deployed by different operators or multi-mode systems owned by the same operator coexist in the area. For the former scenario, intra-network spectrum sharing is adopted to allocate resources; and for the latter scenario, inter-network schemes may be used.

A centralized intra-network spectrum sharing is proposed in [29], where the algorithms have been studied thoroughly and one optimal algorithm in terms of fairness and throughput has been found. A hybrid scenario with DS-CDMA and MC-CDMA system, deployed in the same cell, i.e., intra-cell inter-vendor case, is considered in [30], and it is shown that the spectrum utilization can be improved by spectrum sharing compared to the

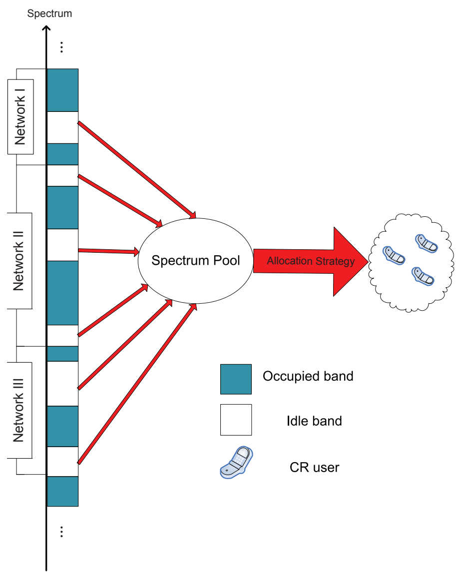


Figure A.4: Workflow of Spectrum Pooling

system without sharing. Inter-network spectrum sharing for UMTS, 802.22 system and IMT-Advanced systems has been investigated in [31–33].

A.3 Spectrum Pooling Concepts

A.3.1 Spectrum Pooling: State of the Art

Spectrum pooling, firstly introduced in [34] by Mitola, enables public access to those already licensed spectrum bands. It aims to enhance the spectrum utilization efficiency by allowing new wireless networks to work overlapped with an existing licensed system without requiring any changes to the existing system. The basic concept of spectrum pooling is to merge the unoccupied spectrum gaps owned by different owners into one common pool, which can be allocated to cognitive radio users for temporary use, as Fig.A.4 illustrates. Through this way, the under-utilized licensed spectrums can be further allocated to some new users when they are not occupied by the licensed users, so that the spectrum utilization can be improved. The basic requirement for the new networks is that they need to be highly flexible in order to efficiently fill the available spectrum gaps.

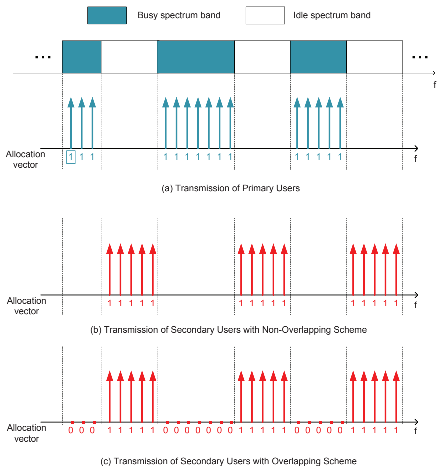


Figure A.5: Overlapping Architecture for OFDM-Based Spectrum Pooling

A.3.2 System Architecture

In the mixed network with spectrum pooling, the primary (or licensed) network can be Time-Division/Frequency-Division/Code-Division Multiple Access (TDMA/FDMA/CDMA) based. However, due to the reason that it is not easy to know the spreading code used in CDMA networks, CDMA-based primary networks are not widely investigated. As for the secondary network, high flexibility is required in order to efficiently utilize the spectrum gaps. Orthogonal Frequency Division Multiplexing (OFDM) based physical transmission scheme is a popular candidate because it is possible to modulate data to spectrum gaps with different bandwidths [35].

In OFDM systems, high-rate serial data stream is transformed into multiple parallel sub-streams with reduced rate. The purpose of this operation is to increase the duration of each symbol, so that the system is more robust to delay spread. By introducing the Cyclic Prefix (CP), the Inter-Symbol-Interference (ISI) can be eliminated. In addition, each sub-stream is modulated and transmitted on a separate orthogonal subcarrier. By changing the scale of serial to parallel transformation, the required number of sub-carries (equivalent to the required bandwidth) can be adjusted. Also, it is possible for primary and secondary networks to coexist in the same spectrum, because the transmission energy on some certain spectrum bands from the OFDM-based secondary network can be suppressed by setting the inputs of the corresponding sub-carriers as zero, and the interference to primary network will be controlled. Fig.A.5.(b) shows this overlapping architecture.

Fig.A.5.(c) shows another example of OFDM-based system spectrum pooling, where the sub-carriers of secondary network are modulated to those unoccupied spectrums. In other words, no spectrum overlapping occurs, so that the coexistence interference between primary and secondary network can be avoided. Furthermore, it is preferred that the spectrum gaps are integer multiple of sub-carrier spacing in secondary network, which leads to higher spectrum utilization.

A.3.3 Spectrum Sharing with Spectrum Pooling

In a spectrum pooling strategy, there are two key factors. They are the time interval of updating the spectrum gaps stored in the pool and how to allocate the available resources. The selection of updating interval differs based on different scenarios. For example, when the traffic density of the primary network is heaver, shorter update period should be used in order not to influence the primary transmissions. If the primary transmissions are sparse, then the updation can be performed infrequently. It has been studied in [36] that the update time of spectrum pool varies according to spectrum characteristics and user's quality of service demands. It is also shown that tradeoff is needed between many factors in spectrum pooling, such as access time delay, system efficiency, costs of software and hardware resources, and complexity in establishing multi-dimensional cost function.

In practical spectrum pooling scenarios, multiple networks may be deployed by different operators. This indicates that it is difficult to have a central control point, and thus spectrum sharing is preferred to work in a distributed manner. In [37, 38], how to optimally allocate spectrum among the networks with spectrum pooling in a distributed and cooperative way is investigated. In them, the dynamic inter-network spectrum sharing among multiple networks is formulated as a restless bandits model-based optimization system, where the multiple networks cooperate with each other to share the spectrum dynamically, and therefore no central control point is needed. It is shown that the proposed scheme in [37, 38] maximizes the utilities of the networks sharing the spectrum pool, where a utility is defined based on the price and spectrum efficiency.

Spectrum pooling can also be used in relay systems. In this case, both the source-relay transmission and relay-destination transmission need to access the spectrum resource stored in the corresponding spectrum pool. In [39], the authors propose an optimal solution to maximize the capacity of an OFDM-based relay system by appropriately choosing the assigned subcarriers in the spectrum pools. It is shown in [39] that the capacity of the system is maximized when the subcarriers' Signal-Noise-Ratios (SNRs) in the spectrum pool corresponding to the source-relay and relay-destination are similarly ordered.

Spectrum pooling is mostly considered in the scenarios where the primary network is deployed in licensed bands. However, spectrum pooling can also be used in unlicensed bands to compensate for spectrum scarcity caused by fixed licensing rules. In the unlicensed bands, Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) is widely employed. In [40], the authors studied the performance of a modified CSMA/CA protocol as the spectrum sharing algorithm for a distributed-overlay cognitive radio network, where three different backoff algorithms are evaluated.

A.4 Game-theory and its Application in Spectrum Allocation

Game theory provides a well-defined model describing conflict and cooperation among intelligent rational decision makers. It matches in nature to the spectrum allocation problem in cognitive radio networks. This has led to its extensive usage in spectrum allocation in cognitive radio networks. In this section, some fundamental knowledge on game theory and its application in spectrum allocation are introduced.

A.4.1 Game Theory Basics

The first known discussion on game theory emerged about three hundred years ago by James Waldegrave, which is about a two-person card gambling game (called Le Her). After that, many researchers, such as James Madison, Antoine Augustin Cournot and John Nash, have contributed a lot to the flourishing of game theory. From 1950s, game theory began to develop systematically as a branch of applied mathematics. It has been employed in many fields, such as social science, biology, political problems as well as computer and information science. Nowadays, as it is remarked by Aumann in [41] that, game theory is a sort of umbrella or unified field theory for the rational side of social science, where social is interpreted broadly, including human as well as non-human players (computers, animals, plants).

Generally, a game has three elements. They are the set of players, the strategy space for each player and the payoff function. The strategy space defines the actions a player can select in every distinguishable state. The payoff (or utility) function associates the payoff for each player with

Table A.1: Payoff Matrix for a 2-Player 2-Strategy Normal Game

	P1 and S-A1	P1 and S-B1
P2 and S-A2	(4,3)	(-1,-1)
P2 and S-B2	(0,0)	(3,4)

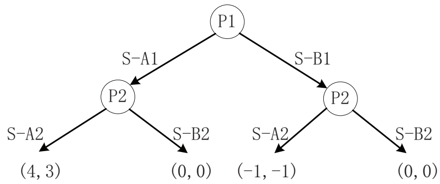


Figure A.6: Payoff Tree for a 2-Player 2-Strategy Extensive Form Game

every possible combination of actions and thus indicates the motivations of players.

When each player in a game acts simultaneously or, at least, without knowing the actions of the others, the game is called a normal (or strategy form) game. In this scenario, a payoff matrix is usually used to represent the game for ease of understanding. Table. A.1 shows an example payoff matrix with 2-player and each player has 2 strategies. There are two numbers in each matrix element, where the first number represents the payoff for Player 1 (P1) and the second number for Player 2 (P2). Take the element (4,3) for an example: 4 is the payoff for Player 1 and 3 is the payoff for Player 2 when Player 1 takes strategy A1 and Player 2 takes strategy A2.

When the players have some information about the choices of other players, the extensive form should be used to formalize the game, where the action orders are considered. Here, games are often represented by trees, instead of matrix, as shown in Fig.A.4.1. In this case, element (4,3) means the payoff for Player 1 and Player 2 are 4 and 3, respectively, when Player 1 firstly takes Strategy A1 and then Player 2 uses Strategy A2.

Games can also be classified into different types from other aspects, for example, cooperative and non-cooperative game, simultaneous and sequential game, perfect information and imperfect information game, and so on. Each type has its own characteristics, and matches with some problems encountered in communication systems. Game theory has been widely applied to modeling and analysis in communication systems, including the spectrum allocation issues in cognitive radio networks.

A.4.2 Nash Equilibrium

In game theory, Nash Equilibrium, named after John Forbes Nash, is an important analysis model to measure the outcome of a non-cooperative game and has been comprehensively studied. It provides a solution to a game involving two or more players, where each player is assumed to know the equilibrium strategies of the other players, and no player can gain anything by changing only his or her own strategy unilaterally. To be short, each player makes the best decision that he or she can, taking into account the decisions of the others in Nash Equilibrium.

In spectrum allocation mechanisms of cognitive radio networks, distributed non-cooperative spectrum allocation is preferred when it is difficult to deploy central control entities to reach centralized optimal solution. By further assuming that each user is aware of the strategies of others through some mechanisms, the application conditions of Nash Equilibrium are fulfilled, and game theory analysis is ready to be used.

In [7], non-cooperative distributed media access under two strategy-aware scenarios is discussed. One scenario is that the base station broadcasts public spectrum information. Another scenario is that each cognitive user obtains private spectrum information via individual spectrum sensing. In the system considered in [7], three types of devices are involved, which are primary mobile station, cognitive radio mobile station and spectrum agent. They are mapped to players in game theory. For the cognitive radio mobile stations, channel access probabilities are defined as the set of strategy profiles, based on which the payoff function can be obtained. Then the equilibrium strategy profile can be found.

For a game, Nash equilibrium may lead to multiple equilibria. Therefore, how to judge and select the optimal solution from all equilibria is a crucial issue. Pareto optimality, among many proposed criteria, is probably the most important and widely adopted one. An outcome is Pareto optimal if there doesn't exist any other outcome that can make at least one player strictly better off without making any other player worse. In other words, a Pareto optimality cannot be improved without hurting any player. In many cases, a Nash equilibrium may not be a Pareto optimality, which means the payoffs can be further increased. In many works, the Nash equilibrium is found firstly and then it is proved to be a Pareto optimality, such as in [42].

A.4.3 Bargaining Game

As for the cooperative spectrum allocation schemes, a barging game can be used to make the analysis. In a barging game, each player has a minimum expected payoff, and the outcome of each player's payoff should not be smaller than the expected value, otherwise the player will not cooperate. Among the possible types of solutions, the Nash Barging Solution plays an important role, which provides a unique Pareto optimal operation point under certain conditions [43].

For an N-person bargaining problem, let S be a closed and convex subset of $\Re^N$ representing the set of feasible payoff allocations that the players can obtain if they all work together. In addition, let $u_{min} = (u^1_{min}, u^2_{min}, ..., u^N_{min})$ be the minimum payoff vector that each player expects, otherwise, the player will not cooperate. Then, $\bar{u} = \phi(S, u_{min})$, is said to be a Nash bargaining solution in S for u_{min} if the following axioms are satisfied:
1. Individual rationality: $\bar{u}_i \geq u^i_{min}, \forall i$.
2. Feasibility: $\bar{u} \in S$.
3. Pareto optimality: For every $\hat{u} \in S$, if $u_i \geq \bar{u}_i, \forall i$, then $\hat{u}_i = \bar{u}_i, \forall i$.
4. Independence of irrelevant alternatives: If $\bar{u} \in S' \subset S, \bar{u} = \phi(S, u_{min})$, then $\bar{u} = \phi(S', u_{min})$.
5. Independence of linear transformations: For any linear scale transformation ψ, $\psi(\phi(S, u_{min})) = \phi(\psi(S), \psi(u_{min}))$.
6. Symmetry: If S is invariant under all exchanges of agents, $\phi_j(S, u_{min}) = \phi_{j'}(S, u_{min}), \forall j, j'$.

In a Nash barging solution, the intuitive concept is to fulfill the minimum requirements of each player firstly, and then allocate the remaining resources to players proportionally to their conditions. This allocation idea matches well with the design requirement of fair radio resource allocation. Indeed, the Nash barging solution has been applied to wireless communication systems broadly, such as in [44, 45].

A.4.4 Auction-based Game

In cognitive radio networks, the primary users may be willing to lease or sell the unused spectrums to cognitive radio users in order to gain some extra payoffs. Meanwhile, the primary users do not want to expose their private information, but on the other hand, the secondary users try to acquire such information to coordinate and communication with each other, which leads to a confliction. In addition, the game models described previously assume that the sets of strategies are known by each player. This assumption can hardly hold in real cognitive networks. The auction game model, which is a well-developed and important applied branch of game theory, helps to formulate and analyze the scenarios discussed here.

In auction games, the players are divided into two groups: auctioneers and bidders. The auctioneers determine how to allocate the resources and how much to charge for the resources based on the bids proposed by bidders. According to the winner selection criteria, auction games mainly have the following types, which have been widely adopted in many areas, e.g. in human society as well as in academic research.

- First-price sealed-bid auction. The bidders submit their bids in a sealed envelop to the auctioneers simultaneously, the one with the highest bid wins the auction, and he/she pays the exact amount of the highest bid.
- Second-price sealed-bid auction. The bidders submit their bids in a sealed envelop to the auctioneers simultaneously, the one with the highest bid wins the auction, but he/she pays the exact amount of second highest bid.
- Open ascending-bid auction. It is also called English auction. The auctioneer increases the price step by step, until there is only one bidder who wants to pay the current price. This bidder is the auction winner.
- Open descending-bid auction. This is also called Dutch auction, and it operates in an opposite way as English auction. The auctioneer starts from a sufficiently high price and decreases the price gradually, until one bidder would like to pay the current price, and he/she wins the bid.

In cognitive radio networks, primary users are the auctioneers, and secondary users try to bid for available spectrums. In [46], the first-price and second-price sealed-bid auction mechanisms in dynamic spectrum allocation of cognitive radio networks are studied, where the authors show that these two methods yield similar outcomes in terms of throughput and efficiency. The out-band sensing time is modeled as the price, which the secondary users have to pay in order to get the transmission opportunities. The secondary users consider channel condition, traffic and other payoff-relevant information before they make a bid decision. An improved auction is also proposed in [46] by considering the impact of packet deadline checking.

A.5 Concepts of Spectrum Utilize Efficiency

Spectrum utilize efficiency, also called spectral efficiency or bandwidth efficiency, refers to the information rate that can be successfully transmitted over a certain bandwidth by using some specific transmission techniques. It is a vital parameter to measure the effectiveness of transmissions over the air interface, which has become the bottleneck of wireless communications. High spectrum utilize efficiency, as one of the most important design objectives of modern wireless systems, helps to alleviate the spectrum scarcity challenge to some extent. In this section, various definitions of spectrum utilize efficiency are introduced first, and then its role in system design and optimization is discussed.

A.5.1 Definition

The definition for spectrum utilize efficiency changes along with the development of communication technology, because for some newly emerged system, the original definition may not be suitable or the related parameters are difficult to measure in real systems. For a simple point-to-point physical link, spectrum utilize efficiency is defined as follows [47].

Definition 1. *Link spectrum utilize efficiency is the ratio of effective information rate transferred through the link to the occupied spectrum bandwidth, where the effective information excludes the error-correcting codes. [47]*

Usually, the information rate is calculated in the unit of bits/second, and spectrum bandwidth in the unit of hertz. Therefore, the unit for spectrum efficiency is bits/second/hertz, or b/s/hz for short. Note that other units, such as symbols/second/hertz, can also be used, which depends on particular cases.

When considering the spectrum utilization of a whole system, the above definition is not applicable any more. Take a typical cellular network as an example, as shown in Fig.A.7, in which there are multiple cells, and the frequencies are reused with reuse factor of 3. In this scenario, more factors are taken into consideration when defining system spectrum utilize efficiency [48].

Definition 2. *System spectrum utilize efficiency is the ratio of maximum average throughput for a cell (or sector, or site) to the total bandwidth in a defined geographic area, which results in the unit of bits/s/hz/cell or bits/s/hz/m [48].*

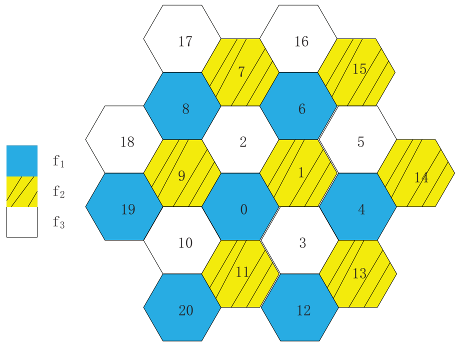


Figure A.7: Typical Cellular Network Topology with Frequency Reuse

In Fig.A.7, the system spectrum efficiency is calculated as:

$$U = \frac{\sum_{i=1}^{i=N} Th_i/N}{\sum_{j=1}^{j=3} f_i}, \tag{A.1}$$

where N denotes the number of cells, Th_i represents the throughput of cell i, f_j is the bandwidth of j^{th} spectrum bandwidth, and 3 is the frequency reuse factor, which is also equal to the number of spectrum bands.

In cognitive radio networks, the cognitive users make use of the spectrum holes, which may not be continuous or fixed. Therefore, it is difficult to define and calculate the amount of spectrum bandwidth. Due to this, several definitions have been proposed for cognitive radio networks. In [49], the spectrum efficiency calculation is based on outage probability P_{out} as follows, where an outage event is considered to occur when the channel capacity falls below some data rate threshold,

$$\eta = \frac{1 - P_{out}}{P_a}. \tag{A.2}$$

Here, $1 - P_{out}$ is the quantity of spectrum holes that can be utilized by cognitive uses and have satisfactory transmission quality, and P_a denotes the total amount of spectrum holes available for cognitive users. In other words, the spectrum efficiency defined here is a measure to show the percentage of available and usable spectrum holes.

In addition, a definition based on link spectrum efficiency has also been adopted in some literatures, such as in [50, 51]. Let p_b denote the probability that the channel is not occupied by the primary users, P_t^P and P_t^S is the transmitted power level from the primary and secondary users, respectively, h_P represents the channel gain between a pair of primary transmitter and receiver, and h_S the channel gain between a pair of secondary transmitter and receiver. Then, the spectrum efficiency on a certain band from Shannon's formula can be expressed as [50]:

$$\eta = (1 - p_b)\log_2\left(1 + \frac{h_P^2 P_t^P}{N_0 B}\right) + p_b \log_2\left(1 + \frac{h_s^2 P_t^S}{N_0 B}\right), \tag{A.3}$$

where B is the frequency bandwidth and N_0 is the power spectrum density of white Gaussian noise.

A more practical model is used in [51], where the impact of adaptive modulation and coding (AMC) scheme is considered in a cognitive radio system over a wireless fading interference channel. The range of Signal to Noise and Interference Ratio (SNIR) of the transmitter is divided into $N+1$ non-overlapping consecutive segments, expressed as $[\theta_0, \theta_1, ..., \theta_i, ..., \theta_{N+1}]$, where θ_0 is set to 0, and θ_N is infinity. When the measured SNIR falls in the interval $[\theta_{i-1}, \theta_i], (i = 1, ..., N)$, the i^{th} rate in the rate set $R_1, R_2, ..., R_N$ will be selected. Then the authors define the average spectral efficiency as:

$$\eta_m^{avg} = \sum_{i=1}^{i=N} R_i \times Pr\left\{\theta_{m,i} \leq \gamma_i < \theta_{m,i+1}\right\} \tag{A.4}$$

where $m \in \{p, s\}$ is the subscript to represent for *p*rimary or *s*econdary user, γ_i denotes the measured SINR, and $Pr\left\{\theta_{m,i} \leq \gamma_i < \theta_{m,i+1}\right\}$ is the probability that the i^{th} rate is used for transmission. Actually, this definition is equivalent to average transmission rate, which can be used to evaluate systems with the same bandwidth. When the impact of different bandwidths should be considered, we can divide η_m^{avg} by the bandwidth.

A.5.2 Optimization

In the previous subsection, different definitions for spectrum utilize efficiency are introduced. Since the principal advantage of cognitive radio is to enhance the spectrum efficiency, it has been considered as the ultimate goal of many optimization problems. Here, system optimization based on spectrum efficiency will be discussed.

Essentially, how to design an efficient system turns into how to effectively control the interference, and further how to reasonably manage and allocate the power resources, constrained by the interference level.

The power control problem of secondary users is discussed in [50], where the spectrum efficiency defined in equation (A.3) is used. When both the primary and secondary users operate in a greedy manner, the constraint on the secondary users' transmission power is that the introduced interference should not exceed a threshold, which can be expressed as:

$$h_s^2 P_t^S + N_0 B \leq k T_{max} B, \tag{A.5}$$

where k is a constant, called the Boltzmann's constant [52], and T_{max} is the allowed upper limit of interference temperature. In order to maximize the spectrum efficiency in (A.3) under the constraint in (A.5), the highest transmission power of the secondary user should be:

$$P_t^S = \frac{k T_{max} B - N_0 B}{h_s^2}. \tag{A.6}$$

Other three spectrum sharing schemes are also discussed in [50] with corresponding power control expressions for secondary users, which can optimize the spectrum efficiency.

The selection of adaptive modulation and coding schemes is considered in [51]. Besides power control, the consideration is based on the scaled SNIR feedback of primary and secondary transmission links. The optimization objective is to maximize the spectrum efficiency of secondary users, i.e., to maximize k_2^{avg}, under two constraints. Firstly, the average spectrum efficiency for primary users, denoted by k_1^{avg}, must be guaranteed above a minimum level $\bar{K_1}$. Secondly, the available transmission power for secondary users (denoted by P_2) has an upper bound $\bar{P_2}$. This problem can be summarized as follows:

$$\begin{aligned} \text{Objective:} \quad & \max k_2^{avg} \\ \text{Subject to:} \quad & k_1^{avg} \geq \bar{K_1} \quad \text{and } P_2 \leq \bar{P_2} \end{aligned}$$

It has been proved that increasing P_2 will increase k_2^{avg} and decrease k_1^{avg} at the same time. The solution for the above optimization problem is obtained by increasing P_2 from zero up to a point where both the two constraints are satisfied with equality.

A.6 Spectrum Allocation Algorithms and Performance Analysis

Much research effort has been put on spectrum allocation mechanisms in cognitive radio networks, and various types of algorithms have been proposed and studied. In this section, several typical algorithms are selected as examples to illustrate the spectrum allocation problem. The performance analysis results are also included.

A.6.1 Pricing-based Spectrum Allocation

Generally, cognitive radio technology helps to improve the spectrum efficiency. However, the collusion among selfish users may deteriorate the efficiency seriously. Therefore, prices should be charged for selfishness and collusion. In [53], an efficient pricing-based distributive collusion-resistant dynamic spectrum allocation approach is proposed in order to optimize the overall spectrum efficiency, where the spectrum allocation problem is modeled as a multistage dynamic game. The authors consider a cognitive radio network with J primary users and K secondary users, where all users are assumed to be selfish and are allowed to cheat. All available spectrums from primary users are collected in a spectrum pool. Secondary users can lease the spectrums and they must pay for successful leases.

Then, the utility for each primary and secondary user is expressed as:

$$U_{P_i}\left(\phi_{\mathbf{A}_i}, \alpha_i^{\mathbf{A}_i}\right) = \sum_{j=1}^{n_i}\left(\phi_{a_i^j} - c_i^j\right)\alpha_i^{a_i^j}, \tag{A.7}$$

$$U_{S_i}\left(\phi_{\mathbf{A}}, \beta_i^{\mathbf{A}}\right) = \sum_{j=1}^{N}\left(v_i^j - \phi_j\right)\beta_i^j, \tag{A.8}$$

with the following notations:

- P_i: primary user i
- S_i: secondary user i
- $\mathbf{A}_i = \{a_i^j\}_{j\in\{1,2,\ldots,n_i\}}$, where a_i^j represents the channel index in the spectrum pool, and n_i is the total number of channels belonging to primary user P_i
- $\phi_{\mathbf{A}_i} = \{\phi_{a_i^j}\}_{j\in\{1,2,\ldots,n_i\}}$, where $\phi_{a_i^j}$ is the payment that primary users P_i obtains from secondary users by leasing channel a_i^j in the spectrum pool
- $\alpha_i^{\mathbf{A}_i} = \{\alpha_i^{a_i^j}\}_{j\in\{1,2,\ldots,n_i\}}$, where $\alpha_i^{a_i^j} \in \{0,1\}$, which indicates whether channel j of primary user P_i has been rent by a secondary user or not
- $\mathbf{C}_i = \{c_i^j\}_{j\in\{1,2,\ldots,n_i\}}$, where c_i^j represents the acquisition cost of the j^{th} channel belonging to primary user P_i
- $\mathbf{V}_i = \{v_i^j\}_{j\in\{1,2,\ldots,n_i\}}$, where v_i^j represents the payoff if secondary user successfully leases the j^{th} channel from primary user P_i

- $\phi_{\mathbf{A}} = \{\phi_j\}_{j \in \{1,2,\ldots,N\}}$, where N denotes the total number of channels in the spectrum pool
- $\beta_i^{\mathbf{A}} = \{\beta_i^j\}_{j \in \{1,2,\ldots,n_i\}}$, where $\beta_i^j \in \{0,1\}$ indicates if secondary user S_i has successfully leased the j^{th} channel

It is straightforward that the interest of primary users conflicts with that of secondary users, because primary users want to earn as much as possible by leasing spectrums, while secondary users would like to obtain transmission opportunities by paying as least as possible. When multiple channels are considered and the selfish users are not willing to share their own information with the others, the spectrum allocation problem becomes a multi-stage non-cooperative pricing game, which can be solved by the auction theory. The auction games have been briefly described in Section A.4.4. However, the auction problem here is different, because it includes two scenarios, i.e., not only the secondary users but also the primary users need to compete with each other to attract secondary users to rent their spectrums. Hence, optimizations for primary and secondary users need to be implemented separately, which makes the problem considered here even more complex. The authors rely on the competitive equilibrium to solve this problem. Particularly, the objective functions for primary and secondary users are expressed as:

$$\widetilde{O}(P_i) = \max_{\phi_{\mathbf{A}_i}, t, \alpha_{i,t}^{\mathbf{A}_i}} E_{c_i^j, v_i^j} \left[\sum_{t=1}^{\infty} \gamma^t \cdot U_{P_i,t}(\phi_{\mathbf{A}_i}, t, \alpha_{i,t}^{\mathbf{A}_i}) \right] \tag{A.9}$$

$$\widetilde{O}(S_i) = \max_{\phi_{\mathbf{A}}, t, \beta_{i,t}^{\mathbf{A}}} E_{c_i^j, v_i^j} \left[\sum_{t=1}^{\infty} \gamma^t \cdot U_{S_i,t}\left(\phi_{\mathbf{A}}, t, \beta_{i,t}^{\mathbf{A}}\right) \right] \tag{A.10}$$

where the subscript t denotes the index for time stages.

Based on the multi-stage pricing game model, the impact of user collusion is further discussed and a collusion-resistent spectrum allocation with multi-secondary and multi-primary users is proposed in [53], where the optimal reserve prices are obtained. Nash barging solution is relied on to derive the performance lower bounds. Simulation results show that the proposed algorithm can achieve high spectrum efficiency by only using limited overhead under various situations of user collusion. Fig.A.8 compares the total utilities of several algorithms with different configurations, such as the Competitive Equilibrium (CE) without user collusion, Nash Bargaining Solution with all.inclusive collusion and the proposed pricing-based collusion-resistant algorithm. It is shown that when there is no user collusion, the dynamic pricing scheme without reserve prices is able to achieve similar performance compared to the theoretical CE outcomes. However,

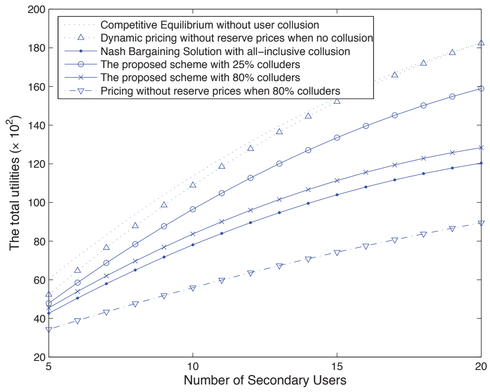


Figure A.8: Comparison of total utilities

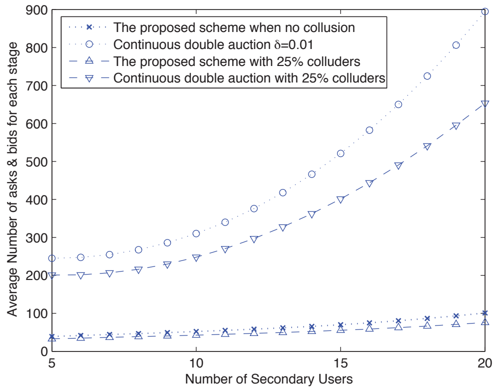


Figure A.9: Comparison of total utilities

with the presence of user collusion, the proposed scheme with reserves prices achieves much higher total utilities than those of the scheme without reserve prices. Fig.A.9 plots the simulation results about overhead, which shows that the proposed approach substantially decreases the pricing communication overhead under either the situations with user collusion or without user collusion.

A.6.2 Primary-prioritized Markov Spectrum Allocation

In [54], the interactions between primary users and secondary users are modeled as a continuous-time Markov chains (CTMC), and a primary-prioritized Markov approach is proposed for dynamic spectrum access. The optimal access probabilities for secondary users under different fairness criteria are derived.

A basic network with one primary user (denoted by P) and two secondary users (denoted by A and B) is considered firstly, where the service request is modeled as Poisson traffic with rate λ_P, λ_A and λ_B per second, and the service duration is negative-exponentially distributed with mean rate μ_P, μ_A and μ_B per second. The CTMC without/with queueing for the basic network is studied, and the results are extended to the scenarios with more secondary users. Then the authors propose a primary-prioritized dynamic spectrum access under different optimality criteria, including maximal-throughput, max-min fairness and proportional fair, and the optimal access probabilities are obtained. Based on these probabilities, secondary users decide how to access the spectrum. The operation flow for the proposed mechanism is summarized in Fig.A.10.

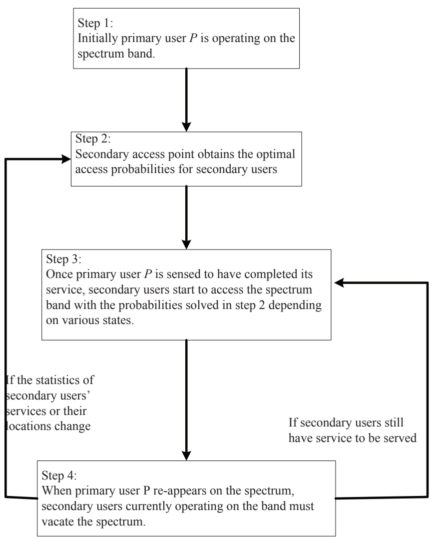


Figure A.10: Operation flow of the primary-prioritized Markov Spectrum Allocation

Simulation results under different configurations are shown and discussed. By comparisons, the authors conclude that the proposed primary-prioritized spectrum access with proportional fair criterion can achieve up

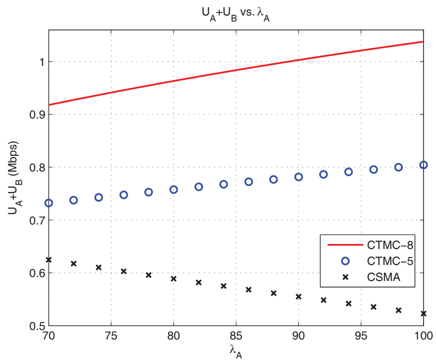


Figure A.11: Overall Throughput Comparison

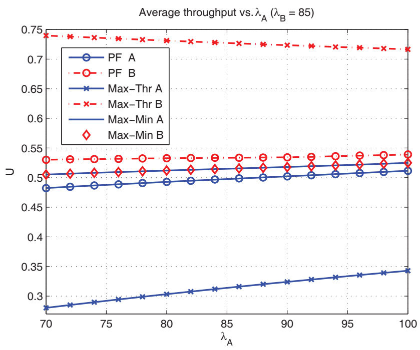


Figure A.12: Comparison of Fairness

to 95% performance gain over the normally used CSMA-based random access approach as indicated by the results of $CTMC-8$ in Fig.A.11, where $CTMC-8$ means the 8-state Continuous Time Markov Chain. In addition, related simulation results also show that the proposed algorithm achieves the optimal tradeoff between efficient spectrum utilization and fairness. Fig.A.12 compares the throughput for two secondary users under three different optimization goals (PF: Proportional Fair, Maximum Throughput: Max-thr, Max-min fairness: Max-min) for two secondary users (A and B). It can be noticed that, the results for *Max-min A* and *Max-min B* are the same, which leads to the fairest allocation, however, it also results in the lowest efficiency. The difference between *Max-thr A* and *Max-thr B* is the largest, indicating the poorest fairness. The results for *PF A* and *PF B* give a tradeoff between throughput and fairness.

A.6.3 Evolutionary Algorithms Based Spectrum Allocation

Evolutionary algorithms, which are stochastic search methods that mimic natural evolution and/or the social behavior of species, has also been applied to spectrum allocation optimization issues. In [55], the authors use three algorithms to search for the optimal channel allocation schemes under different objectives.

A network with N secondary users, who compete with each other among M non-overlapping channels, is focused on. To describe the spectrum allocation model, the following notations are defined.

- $L = \{l_{n,m} | l_{n,m} \in \{0,1\}\}_{N \times M}$: denotes the channel availability, where $l_{n,m}$ is a binary indicator. When $l_{n,m} = 1$, the channel m is available to user n, and $l_{n,m} = 0$ otherwise.
- $B = \{b_{n,m}\}_{N \times M}$: is the channel reward matrix, where $b_{n,m}$ represents the reward that user n obtains by using channel m.
- $C = \{c_{n,k,m} | c_{n,k,m} \in \{0,1\}\}_{N \times N \times M}$: represents the interference constraint, where $c_{n,k,m} = 1$ means user n and user k will interfere with each other if they user channel m at the same time, and $c_{n,k,m} = 0$ otherwise. In addition, $c_{n,n,m} = 1 - l_{n,m}$.
- $A = a_{n,m} | a_{n,m} \in \{0,1\}\}_{N \times M}$: is the channel assignment matrix, where $a_{n,m}$ indicates that channel m is allocated to secondary user n. Here, A is a conflict-free channel assignment matrix if the interference constraints defined by C is satisfied, i.e., $a_{n,m} \cdot a_{k,m} = 0$, if $c_{n,k,m} = 1, \forall 1 \leq n, k \leq N, 1 \leq m \leq M$.
- $R = \{r_n\}_{N \times 1}$: represents the reward vector, where $r_n = \sum_{m=1}^{m=M} a_{n,m} \cdot b_{n,m}$ is the reward user n obtains, given the conflict-free channel assignment matrix A and the reward matrix B.

For a given L and C, there may be several conflict-free channel assignment matrixes, and let $\Lambda_{L,C}$ denote the set of these matrixes. Then, the optimization objective is to find the conflict-free channel assignment matrix, that can maximize the network utilization $U(R)$, i.e.,

$$A^* = \arg \max_{A \in \Lambda_{L,C}} U(R). \tag{A.11}$$

Three definitions for network utilization $U(R)$ are considered, including

- Max-Sum-Reward (MSR): $U(R) = \sum_{n=1}^{N} r_n$
- Max-Min-Reward (MMR): $U(R) = min_{1 \leq n \leq N} r_n$

Table A.2: Comparison of Average Award

Generation or iteration	Algorithm	Average reward ($N = 5, M = 5$)			Average reward ($N = 20, M = 20$)		
		MSR	MMR	MPF	MSR	MMR	MPF
10	GA-SAA	150.6038	25.0771	53.8800	1170.9525	2.3642	62.4950
	QGA-SAA	151.0952	27.8634	65.1482	1206.1508	9.5613	37.0068
	PSO-SAA	151.0952	27.8634	64.8348	1204.7035	8.3244	13.7659
50	GA-SAA	151.0952	25.6343	64.0926	1229.1568	7.2553	96.2020
	QGA-SAA	151.0952	27.8634	66.1267	1237.3000	39.5744	87.5857
	PSO-SAA	151.0952	27.8634	67.8842	1238.1758	28.0520	82.4346
300	GA-SAA	151.0952	27.3061	64.5525	1238.9552	12.3750	116.7429
	QGA-SAA	151.0952	27.8634	66.1267	1238.9561	56.2500	118.0215
	PSO-SAA	151.0952	27.8634	67.9928	1240.1890	50.9594	120.5298
	CSGC	138.3981	21.1016	56.0257	1206.0437	2.7769	60.1252

- Max-Proportional-Fair (MPF): $u(R) = (\prod_{n=1}^{N}(r_n + 10^{-6}))^{\frac{1}{N}}$

Accordingly, three evolutionary algorithms, namely Genetic Algorithm Spectrum Allocation Algorithm (GA-SAA), Quantum Genetic Algorithm SAA (QGA-SAA) and Particle Swarm Optimization SAA (PSO-SAA), are used to solve the above optimization problem. And color sensitive graph coloring (CSGC) algorithm is used for performance comparison. Results show that the proposed methods greatly outperform CSGC under all experiments. For example, Tabel.A.2 summarizes the average reward for each algorithm under different objective functions defined above. It can be noticed that the average rewards obtained by GASAA, QGA-SAA and PSO-SAA after 50 generations are better than CSGC, which validates the effectiveness of the proposed evolutionary algorithms-based spectrum allocation methods.

A.7 Summary

In this section, the spectrum allocation problem in cognitive radio networks has been discussed from several aspects, including spectrum sharing policy, concept of spectrum pooling and spectrum utilize efficiency, game theory and its applications, and three specific spectrum allocation algorithms. As for the topics, there are a number of issues that are not discussed or need further study, such as:

In some spectrum allocation algorithms, it is assumed that the secondary users can share the spectrum sensing information with each other, and then spectrum access can be performed based on the overall information. However, how to obtain and how to share the spectrum sensing information are challenging, since users are vigilant to disclose their information and the available resource for information sharing is also limited.

Much research focuses on the scenario with one primary user and multiple secondary users. However, coexistence of multiple primary users in one system is quite common in reality. Then, more factors should be considered in the spectrum allocation problem, such as the spectrum allocation between primary users.

The modeling of wireless fading channel is seldomly included in existing works on spectrum allocation, which is a vital characteristic of wireless communications. The channel quality plays an important role in resource allocation as indicated by tremendous research effort on wireless scheduling. As for the spectrum allocation among secondary users, the channel quality should also be considered in order to better utilize the resources.

Another important characteristic of wireless communication is mobility. The available spectrum resource changes when a user moves from one place to another. Then how to guarantee continuous allocation and transmission is challenging.

Bibliography

[1] I.F. Akyildiz, Won-Yeol Lee, M.C. Vuran, and S. Mohanty. A survey on spectrum management in cognitive radio networks. *IEEE Communications Magazine*, 46(4):40 – 48, April 2008.

[2] C. Raman, R.D. Yates, and N.B. Mandayam. Scheduling variable rate links via a spectrum server. In *Proceeding of 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005)*, pages 110 – 118, 2005.

[3] V. Brik, E. Rozner, S. Banerjee, and P. Bahl. DSAP: a protocol for coordinated spectrum access. In *Proceeding of 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005)*, pages 611 – 614, 2005.

[4] O. Ileri, D. Samardzija, and N.B. Mandayam. Demand responsive pricing and competitive spectrum allocation via a spectrum server. In *1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005)*, pages 194 – 202, 2005.

[5] Binyang Xu, Feng Yang, Jigang Qiu, and Di Lu. Distributed spectrum sharing algorithm design and realization. In *IEEE 71st Vehicular Technology Conference (VTC 2010-Spring)*, 2010.

[6] Zengqin Xu and Qinyu Zhang. Distributed multichannel spectrum sharing algorithm for cognitive radio networks. In *Proceeding of International Conference on Communication Software and Networks (ICCSN '09)*, pages 781 – 785, 2009.

[7] Yu-Yu Lin and Kwang-Cheng Chen. Distributed spectrum sharing in cognitive radio networks - game theoretical view. In *7th IEEE Consumer Communications and Networking Conference (CCNC)*, 2010.

[8] Y. Xing and R. Chandramouli. Qos constrained secondary spectrum sharing. In *Proceeding of 1st IEEE International Symposium*

on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005), pages 658 – 661, 2005.

[9] G. Auer, H. Haas, and P. Omiyi. Interference aware medium access for dynamic spectrum sharing. In *Proceeding of 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2007)*, pages 399 – 402, 2007.

[10] Chunyi Peng, Haitao Zheng, and Ben Y. Zhao. Utilization and fairness in spectrum assignment for opportunistic spectrum access. *Mobile Networks and Applications*, 11(4):555 – 576, 2006.

[11] T. Ulversoy, T. Maseng, and J. Karstad. On spectrum sharing in autonomous and coordinated dynamic spectrum access systems: A case study. In *Proceeding of 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE 2009)*, pages 649 – 654, 2009.

[12] G Salami, O Durowoju, A Attar, O Holland, R Tafazolli, and H Aghvami. A comparison between the centralized and distributed approaches for spectrum management. *IEEE Communications Surveys & Tutorials*, pages 1 – 17, 2010.

[13] Jishun Liu, Lianfeng Shen, Tiecheng Song, and Xiaoxia Wang. Demand-matching spectrum sharing game for non-cooperative cognitive radio network. In *International Conference on Wireless Communications & Signal Processing (WCSP 2009)*, 2009.

[14] R. Etkin, A. Parekh, and D. Tse. Spectrum sharing for unlicensed bands. *IEEE Journal on Selected Areas in Communications*, 25(3):517 – 528, April 2007.

[15] U. Mir, L. Merghem-Boulahia, and D. Gaïti. A cooperative multiagent based spectrum sharing. In *Proceeding of Sixth Advanced International Conference on Telecommunications (AICT)*, pages 124 – 130, 2010.

[16] Fulu Li. A networking perspective of cooperative spectrum sharing in wireless networks: Analysis and experiments. In *Proceeding of Wireless Telecommunications Symposium (WTS 2008)*, pages 220 – 229, 2008.

[17] D. Niyato and E. Hossain. Market-equilibrium, competitive, and cooperative pricing for spectrum sharing in cognitive radio networks: Analysis and comparison. *IEEE Transactions on Wireless Communications*, 7(11):4273 – 4283, November 2008.

[18] Haipeng Yao, Zheng Zhou, Luyong Zhang, He Liu, Liang Tang, and Kyung Sup Kwak. An efficient power allocation scheme in joint spectrum overlay and underlay cognitive radio networks. In *Proceeding of 9th International Symposium on Communications and Information Technology (ISCIT 2009)*, pages 102 – 105, 2009.

[19] Haipeng Yao, Zheng Zhou, He Liu, and Luyong Zhang. Optimal power allocation in joint spectrum underlay and overlay cognitive radio networks. In *4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM'09.)*, 2009.

[20] Dong In Kim, Long Le, and E. Hossain. Resource allocation for cognitive radios in dynamic spectrum access environment. In *3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM'08)*, 2008.

[21] Dong Kim, Long Le, and E. Hossain. Joint rate and power allocation for cognitive radios in dynamic spectrum access environment. *IEEE Transactions on Wireless Communications*, 7(12):5517 – 5527, December 2008.

[22] Long Bao Le and E. Hossain. Resource allocation for spectrum underlay in cognitive radio networks. *IEEE Transactions on Wireless Communications*, 7(12):5306 – 5315, December 2008.

[23] L.S. Pillutla and V. Krishnamurthy. Game theoretic rate adaptation for spectrum-overlay cognitive radio networks. In *IEEE Global Telecommunications Conference (IEEE GLOBECOM 2008)*, 2008.

[24] Rekha Menon, R.M. Buehrer, and J.H. Reed. Outage probability based comparison of underlay and overlay spectrum sharing techniques. In *Proceeding of 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005)*, pages 101 – 109, 2005.

[25] Yingxiao Wu and Zhen Yang. Spectrum sharing under spectrum overlay and underlay with multi-antenna. In *Proceeding of 9th International Conference on Networks (ICN)*, pages 237 – 242, 2010.

[26] A. Mishra. A multi-channel mac for opportunistic spectrum sharing in cognitive networks. In *IEEE Military Communications Conference (MILCOM 2006)*, 2006.

[27] M. Khoshkholgh, K. Navaie, and H. Yanikomeroglu. Access strategies for spectrum sharing in fading environment: Overlay, underlay and

mixed. *IEEE Transactions on Mobile Computing*, 9(12):1780 – 1793, December 2010.

[28] R. Menon, R. Buehrer, and J. Reed. On the impact of dynamic spectrum sharing techniques on legacy radio systems. *IEEE Transactions on Wireless Communications*, 7(11):4198 – 4207, November 2008.

[29] H. Mujeeb, F. Aslam, and S. Aslam. Device-centric spectrum sharing for cognitive radios. In *Proceeding of International Conference on Networking and Information Technology (ICNIT)*, pages 410 – 414, 2010.

[30] A. Pezeshk and S.A. Zekavat. Ds-cdma vs. mc-cdma, a performance survey in inter-vendor spectrum sharing environment. In *Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers*, pages 459 – 464, 2003.

[31] A. Yarmohammad, M. Abaii, S. Thilakawardana, and R. Tafazolli. Inter-operator dynamic spectrum selection in umts. In *IEEE 69th Vehicular Technology Conference (VTC Spring 2009)*, 2009.

[32] R. Chen, Weiying Li, and Jiandong Li. Auction-based inter bs spectrum sharing for ieee 802.22 wran systems. In *Proceeding of 4th IEEE Conference on Industrial Electronics and Applications (ICIEA 2009)*, pages 1787 – 1791, 2009.

[33] A. Mihovska, F. Meucci, N.R. Prasad, F.J. Velez, and O. Cabral. Multi-operator resource sharing scenario in the context of imt-advanced systems. In *Proceeding of 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART 2009)*, pages 12 – 16, 2009.

[34] III Mitola, J. Cognitive radio for flexible mobile multimedia communications. In *IEEE International Workshop on Mobile Multimedia Communications (MoMuC '99)*, pages 3 – 10, 1999.

[35] T.A. Weiss and F.K. Jondral. Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency. *IEEE Communications Magazine*, 42(3):8 – 14, March 2004.

[36] Hongjie Liu, Shufang Li, Zhongxu Wang, Weijun Hong, and Min Yi. Strategy of dynamic spectrum access based-on spectrum pool. In *4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, 2008.

[37] Pengbo Si, Enchang Sun, Ruizhe Yang, and Yanhua Zhang. Cooperative and distributed spectrum sharing in dynamic spectrum pooling networks. In *19th Annual Wireless and Optical Communications Conference (WOCC)*, 2010.

[38] Pengbo Si, Hong Ji, F.R. Yu, and V.C.M. Leung. Optimal cooperative internetwork spectrum sharing for cognitive radio systems with spectrum pooling. *IEEE Transactions on Vehicular Technology*, 59(4):1760 – 1768, May 2010.

[39] A. Pandharipande and Chin Keong Ho. Stochastic spectrum pool reassignment for cognitive relay systems. In *Proceeding of IEEE Wireless Communications and Networking Conference (WCNC '08)*, pages 588 – 592, 2008.

[40] A.V. Adamis and P. Constantinou. Performance study of csma/ca over spectrum pooling environment for cognitive radios. In *Proceeding of 3rd IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMOB 2007)*, pages 31 – 31, 2007.

[41] Robert J. Aumann. Game theory. *The New Palgrave: A Dictionary of Economics*, 2:460–82, 1987.

[42] Jun Sun, E. Modiano, and Lizhong Zheng. Wireless channel allocation using an auction algorithm. *IEEE Journal on Selected Areas in Communications*, 24(5):1085 – 1096, May 2006.

[43] Ekram Hossain, Dusit Niyato, and Zhu Han. *Dynamic Spectrum Access and Management in Cognitive Radio Networks*. Cambrige University Press, 1 edition, 2009.

[44] Z. Han, Z. Ji, and K.J.R. Liu. Fair multiuser channel allocation for ofdma networks using nash bargaining solutions and coalitions. *IEEE Transactions on Communications*, 53(8):1366 – 1376, August 2005.

[45] Chih-Yu Wang, Kuo-Tung Hong, and Hung-Yu Wei. Nash bargaining solution for cooperative shared-spectrum wlan networks. In *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007)*, 2007.

[46] Bin Chen, A.T. Hoang, and Ying-Chang Liang. Cognitive radio channel allocation using auction mechanisms. In *Proceeding of IEEE Vehicular Technology Conference (VTC Spring 2008)*, pages 1564 – 1568, 2008.

[47] Sergio Benedetto and Ezio Biglieri. *Principles of Digital Transmission: With Wireless Applications*. Springer, 1999.

[48] J. Burns. Measuring spectrum efficiency - the art of spectrum utilization metrics. In *25th IEEE Conference on Getting the Most Out of Spectrum*, 2002.

[49] Yulong Zou and Yu-Dong Yao. Spectrum efficiency of cognitive relay transmissions with cooperative diversity in cognitive radio networks. In *Proceeding of Second International Conference on Communication Systems, Networks and Applications (ICCSNA)*, pages 59 – 62, 2010.

[50] Zhengchao Wang. Power control for secondary users in cognitive radio based on spectrum utilization efficiency. In *Proceeding of Second International Conference on Networks Security Wireless Communications and Trusted Computing (NSWCTC)*, pages 162 – 167, 2010.

[51] M. Taki and F. Lahouti. Spectral efficiency optimized adaptive transmission for interfering cognitive radios. In *IEEE International Conference on Communications Workshops (ICC Workshops 2009)*, 2009.

[52] Max Planck. The genesis and present state of development of the quantum theory (Nobel Lecture). June 1920.

[53] Zhu Ji and K.J.R. Liu. Multi-stage pricing game for collusion-resistant dynamic spectrum allocation. *IEEE Journal on Selected Areas in Communications*, 26(1):182 – 191, January 2008.

[54] Beibei Wang, Zhu Ji, K. Liu, and T.C. Clancy. Primary-prioritized markov approach for dynamic spectrum allocation. *IEEE Transactions on Wireless Communications*, 8(4):1854 – 1865, April 2009.

[55] Zhijin Zhao, Zhen Peng, Shilian Zheng, and Junna Shang. Cognitive radio spectrum allocation using evolutionary algorithms. *IEEE Transactions on Wireless Communications*, 8(9):4421 – 4425, September 2009.

Appendix B

Publication B

Yuehong Gao and Yuming Jiang; Performance Analysis for a Cognitive Radio with Imperfect Spectrum Sensing; IEEE INFOCOM 2010 Workshop on Cognitive Wireless Communications and Networking, March 2010.

Abstract

In this paper, we use stochastic network calculus to analyze a cognitive radio network, where influences of imperfect spectrum sensing and different retransmission schemes are considered. In particular, stochastic arrival curves for spectrum sensing error processes are derived firstly, based on which stochastic service curve for each class of users is obtained under different retransmission schemes, including without retransmission, retransmission until success and maximum-N-time retransmission. Then backlog and delay bounds for primary and secondary users are derived. Finally, numerical results are shown for different types of traffic, where the influence of different retransmission schemes is further discussed.

Keywords

Cognitive radio, Sensing error, Retransmission scheme, Performance analysis, Stochastic network calculus

B.1 Introduction

Cognitive radio is a newly proposed wireless communication technology [1], which can improve spectrum utilization and thus increasing communication demands can be better fulfilled. In a cognitive radio network, there are two types of users, namely primary users (PU) and secondary users (SU). Secondary users have the ability to sense and use available spectrum holes when primary users do not transmit data on the assigned spectrum. However, spectrum sensing errors may happen due to uncertainty of wireless channels and unpredictable interference, and imperfect spectrum sensing can influence system performance. It is hence important to conduct performance analysis for both primary and secondary users taking into consideration the impact of imperfect sensing.

Among existing analysis tools, queueing theory has been proved to be a useful method to deal with queueing problems in communication networks. It has also been employed in performance analysis of cognitive radio networks [2–4], where some results have been obtained, such as packet waiting time in queue and delay. However, the focus has mainly been on average values in steady states. In addition, the current analysis mostly assumes $M/G/1$ model with Poisson arrival, and the obtained results cannot be easily extended to other types of arrival processes. Moreover, the influence of imperfect spectrum sensing is not well considered either; although some results based on Monte Carlo simulation are reported in [2], no analytical research is known. Besides, the impact of retransmission is not found.

The objective of this paper is to conduct performance analysis of cognitive radio network by considering both spectrum sensing and retransmission. Specifically, stochastic network calculus [5–8] is employed to analyze performance distribution bounds. First, we obtain stochastic arrival curve for the sensing error process, followed by the derivation of stochastic service curve for both primary users and secondary users under different retransmission schemes. Then performance analysis is conducted based on stochastic network calculus, where expressions for backlog and delay bounds are shown. Lastly, numerical results under various configurations are presented for both Poisson traffic and (σ, ρ)-constrained traffic, where the impact of retransmission is also discussed.

The paper is organized as following. The cognitive radio network is modeled as a preemptive queueing system in Section B.2, where basic assumptions, traffic and server models as well as retransmission schemes are described. In Section B.3, we first derive stochastic arrival curves and bounding functions for the sensing error process based on different retransmission mechanisms. We then analyze stochastic service guarantees provided to the primary and secondary users, followed by general expressions of performance bounds. Numerical results are shown in Section B.4, from which more insights are discussed. The summary is given in the last section.

B.2 System Model

B.2.1 System Model

In this paper, we consider a cognitive radio network with two input flows as illustrated in Fig.B.1, where fl^P and fl^S represent the aggregated flows from primary users and secondary users, respectively. For ease of expression and with the focus on the impact of sensing error and retransmission, the wireless channel is assumed to be error free with a constant service rate C. The analysis can be easily extended to consider stochastical channel that can be expressed with a stochastic service curve. The primary users' flow has preemptive priority over the secondary users' flow. If one packet arrives into the system and cannot be transmitted immediately, it will be stored in the corresponding buffer in a First-In-First-Out (FIFO) manner, where the buffer is assumed to be large enough and therefore no packet will be dropped.

The system is supposed to be synchronized and the time is divided into slots with length T and indexed by $[0, 1, ..., s, ..., t, ...]$. At the beginning of each slot, secondary users will try to sense the spectrum to decide whether it is idle or busy. In this paper, it is assumed that the time period used for spectrum sensing is small and its effect is not considered. It is also assumed that PUs and SUs can negotiate respectively among themselves before transmitting so no collision will happen between PUs or between SUs. But spectrum sensing errors may occur and they can have significant influence on the system performance.

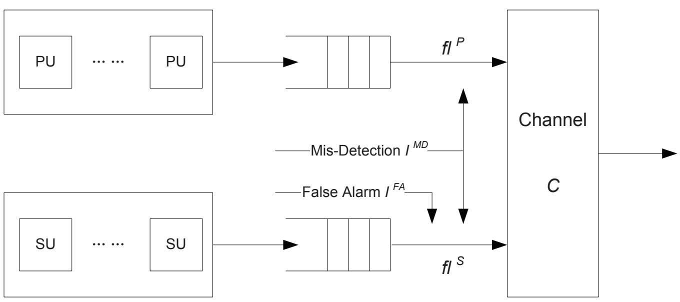


Figure B.1: System Model

Typically, spectrum sensing errors can be classified into two types [9], i.e., mis-detection (MD) and false alarm (FA). Mis-detection means that the spectrum is occupied by PUs but the spectrum sensing result says it is available for SUs, which will result in transmission collision and influence both PUs' and SUs' current transmission. However, false alarm occurs in the opposite way, when SUs believe that the spectrum is being used by PUs but actually the spectrum is idle, which will waste transmission opportunities for SUs. Let p_e denote the average probability that a sensing error (either MD or FA) happens in one time slot. Let ϕ be the probability that this error is a mis-detection. Then the average probability in one time slot for MD and FA can be respectively expressed as $p_e^{MD} = p_e \cdot \phi$ and $p_e^{FA} = p_e \cdot (1 - \phi)$.

Generally, packet arrivals and sensing errors are stochastic processes, which will only lead to stochastic service guarantees. While the system model described above has already been overly simplified, to the best of our knowledge, it is difficult (if not impossible) to obtain explicit results from the traditional queueing theory, particularly when the involving s-tochastic processes are not Poisson or with exponentially distributed rates. To address this problem, we resort to stochastic network calculus.

B.2.2 Stochastic Network Calculus Basics

Stochastic network calculus theory is a newly developed queueing theory for service guarantee analysis (e.g., [6–8] and references therein), which contains two fundamental concepts, stochastic arrival curve and stochastic service curve.

In stochastic network calculus, a stochastic arrival curve (SAC) is used to describe the stochastic characteristics of a flow. There are several definitions for SAC, and in this paper the following definition is used, which explores the *virtual backlog property* of deterministic arrival curve [7].

Definition 1. ***(Stochastic Arrival Curve).*** *A flow $A(t)$ is said to have a virtual-backlog-centric (v.b.c) stochastic arrival curve $\alpha \in F$[1] with bounding function $f \in \bar{F}$[2], denoted by $A(t) \sim_{sac} \langle f, \alpha \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\left\{\sup_{0 \leq s \leq t} \{A(s,t) - \alpha(t-s)\} > x\right\} \leq f(x).$$

[1] F: the set of non-negative wide-sensing increasing functions
[2] $\bar{F}$: the set of non-negative wide-sensing decreasing functions

In Definition 1, $A(s,t)$ denotes the cumulative amount of traffic of the flow during period $(s,t]$, $A(t) = A(0,t)$, and $\alpha(t)$ is a non-decreasing function. While the stochastic arrival curve describes the traffic, the stochastic service curve shows the service guarantee provided by a server. Similarly with SAC, stochastic service curve (SSC) can also be defined in different ways. In this paper, we use the following one [7].

Definition 2. ***(Stochastic Service Curve).*** *A system S is said to provide a stochastic service curve $\beta \in F$ with bounding function $g \in \bar{F}$, denoted by $S \sim_{ssc} \langle g, \beta \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\{A \otimes \beta(t) - A^*(t) > x\} \leq g(x).$$

Here, $A \otimes \beta(t) \equiv \inf_{0 \leq s \leq t} \{A(s) + \beta(t-s)\}$, and $A^*(t)$ denote the cumulative output traffic amount up to time t.

Given SAC and SSC, the following bounds have been derived under stochastic network calculus [7]:

Theorem 1. *Consider a system S with input A. Suppose the input has a v.b.c stochastic arrival curve as $A \sim_{sac} \langle f, \alpha \rangle$; and server S provides the input with a stochastic service guarantee as $S \sim_{ssc} \langle g, \beta \rangle$. Then for any $t \geq 0$ and $x \geq 0$, the backlog $B(t)$ and delay $D(t)$ is bounded by:*

$$\begin{aligned} P\{B(t) > x\} &\leq [f \otimes g(x - \alpha \oslash \beta(0))]_1 \\ P\{D(t) > h(\alpha + x, \beta)\} &\leq [f \otimes g(x)]_1 \end{aligned}$$

where $\alpha \oslash \beta(0) = sup_{u \geq 0}\{\alpha(u) - \beta(u)\}$, $h(\alpha + x, \beta) = sup_{s \geq 0}\{inf\{\tau \geq 0 : \alpha(s) + x \leq \beta(s+\tau)\}\}$ and $[\cdot]_1$ denotes $max(min(\cdot, 1), 0)$.

In order to apply stochastic network calculus results to performance analysis of cognitive radio network, a critical challenge is to find stochastic service curve for both PUs and SUs. Major contribution of this paper is in finding out stochastic service curve provided to both PUs and SUs, considering sensing errors and retransmission schemes, which are presented below.

B.2.3 Retransmission Schemes

As discussed above, transmission collisions may happen due to sensing errors, which may significantly influence the system performance. Retransmission technology is a commonly used method to deal with transmission errors, and different schemes can result in different outcomes. In this paper, the following three retransmission schemes will be discussed, where the first two are extreme cases and the third one is a tradeoff.

WithOut-ReTransmission (WO-RT)

In this scheme, it is assumed that there is no physical layer retransmission. In other words, when one packet is transmitted through the wireless channel, it will be removed from waiting queue no matter it will be received correctly or not. Therefore, sensing error process will not influence backlog and delay, but will affect transmission error.

ReTransmission until Success (RT-S)

Compared with WO-RT, RT-S goes to the other extreme. One packet will be removed from the waiting queue only if it has been received by the receiver successfully. Otherwise, it will be backlogged in buffer as long as needed. Therefore, no transmission error will occur. However, spectrum sensing impairments will lead to larger backlogged queue and longer waiting time.

Max-N-time ReTransmission (Max-N-RT)

This scheme is a tradeoff between WO-RT and RT-S, in which one packet can be retransmitted at most N times. After that, the packet will be removed from the queue no matter it has been received correctly or not. It can be expected from Max-N-RT that the transmission error can be reduced to some extend as compared with WO-RT, while the backlog and delay can be better guaranteed as compared with RT-S.

B.3 Performance Analysis

In this section, performance analysis of the considered cognitive radio network is conducted. The focus is on finding probabilistic bounds on delay and backlog of both primary users (PUs) and secondary users (SUs), and the theoretical tool we rely on is stochastic network calculus. As highlighted in Sec.II.B, in order to apply Theorem 1, it is essential to find stochastic service curves for both PUs and SUs which take into consideration sensing errors and consider different retransmission schemes. To achieve this, we present in the following an analytical approach. First, we study the impact of the sensing error process under different retransmission schemes. Particularly, if a collision happens due to sensing error *and* the collided packets need retransmission, we say the corresponding amount of service (i.e., the corresponding time slot) has been *wasted*, and we shall characterize the *wasted* service with stochastic arrival curve. Then, we treat the *wasted* service process as an interference process, and establish the relationship between the interference process and the stochastic service curve. Based these, we conclude stochastic service curves for both PUs and SUs, where for SUs, we further treat the arrivals from PUs as an interference process to SUs. Finally, we present delay and backlog bounds for both PUs and SUs based on stochastic network calculus results. Throughout the analysis, the three different retransmission schemes are studied.

B.3.1 Impact of Sensing Error

In this subsection, we study the impact of sensing error. Particularly, we focus on the sensing error impact on the amount of service that will otherwise be delivered to the users successfully. We shall say such service is *wasted* to the sender in the sense that it has not helped in reducing the number of packets in the sending queue. We shall characterize the *wasted* service process using stochastic arrival curve.

For WO-RT, interestingly, there is no *wasted* service to the sending queue. This is due to that even though the packet under transmission is collided, the sending queue (no matter whether it belongs to PUs or SUs) does not care and the packet is still removed from the corresponding buffer. As a result, from the sending queue viewpoint, it works just as if there had been no error, and hence the *wasted* service to the sending queue is zero. However, if retransmission takes place due to sensing error, some service will be *wasted* as seen by the sending queue, since no packet is moved out of the queue in a *wasted* service slot.

Assume the sensing error probability is the same on each time slot, denoted by p_e. The average number of errors during any time period $(s, t]$ will be $p_e(t-s)$. Let $I_n(s, t)$ denote the number of sensing errors during $(s, t]$, and $\gamma_n(s, t) = K \cdot p_e(t-s)$, where $K > 0$ is a constant parameter facilitating later analysis in obtaining performance bounds.

For RT-S, the equivalent amount of *wasted* service in $(s, t]$ can be expressed as $I(s,t) = I_n(s,t) * CT$. Let $M_X(\theta)$ denote the moment generating function of random variable X, i.e., $M_X(\theta) = E[e^{\theta X}]$ for any $\theta > 0$. Then, we have:

Lemma 1. *The* wasted *service process I under RT-S has a stochastic arrival curve $\langle f^I, \gamma^I \rangle$, where*

$$\begin{aligned}
\gamma^I(s,t) &= K \cdot p_e(t-s) \cdot CT \\
f^I(x) &= e^{-\frac{\theta x}{CT}} \frac{e^{-\theta K p_e}(p_e e^{\theta} + 1 - p_e)}{1 - e^{-\theta K p_e}(p_e e^{\theta} + 1 - p_e)}
\end{aligned}$$

for any $\theta > 0$.

Proof. By using definition of stochastic arrival curve, Boole's inequation and Chernoff bound, we have:

$$\begin{aligned}
& P\{\sup_{0 \le s \le t} \{I(s,t) - \gamma^I(s,t)\} > x\} \\
= \; & P\{\sup_{0 \le s \le t} \sum_{k=s}^{k=t-1} [I_n(k,k+1) - \gamma_n(k,k+1)] > \frac{x}{CT}\} \\
\le \; & e^{-\frac{\theta x}{CT}} \sum_{s=0}^{s=t-1} E[e^{\theta \sum_{k=s}^{k=t-1} [I_n(k,k+1) - \gamma_n(k,k+1)]}] \\
\le \; & e^{-\frac{\theta x}{CT}} \sum_{i=1}^{i=+\infty} E[e^{\theta [I_n(1) - \gamma_n(1)]}]^i \\
= \; & e^{-\frac{\theta x}{CT}} \frac{M_{I_n(1) - \gamma_n(1)}(\theta)}{1 - M_{I_n(1) - \gamma_n(1)}(\theta)} \equiv f^I(x).
\end{aligned}$$

Since $\gamma_n(1) = K p_e$ and $I_n(1)$ is a Bernoulli distributed random variable, its moment generating function can be obtained as $E[e^{\theta [I_n(1) - \gamma_n(1)]}] = E[e^{\theta I_n(1)}] E[e^{-\theta K p_e}] = e^{-\theta K p_e}(p_e e^{\theta} + 1 - p_e)$. Then, the process I has a stochastic arrival curve as $\langle f^I, \gamma^I \rangle$. □

Similarly, the stochastic arrival curve characterization of the corresponding *wasted* service processes due to mis-detection I^{MD} and false alarm I^{FA} can be found in the same way by replacing p_e in Lemma 1 respectively with ϕp_e and $(1-\phi) p_e$:

$$I^{MD} \sim_{sac} \langle f^{MD}, \gamma^{MD} \rangle, I^{FA} \sim_{sac} \langle f^{FA}, \gamma^{FA} \rangle$$

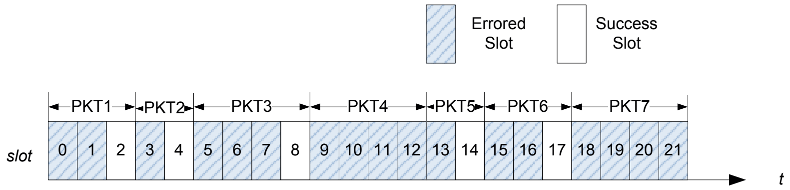


Figure B.2: An Example of Max-N-Time Transmission

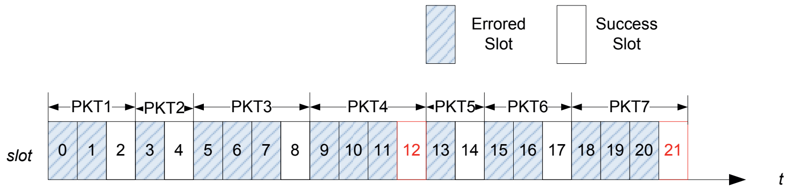


Figure B.3: Equivalent Model of Max-N-Time Transmission

For Max-N-RT, before presenting the result, let us consider an example as shown in Fig.B.2 to see how it works.

Fig.B.2 shows an example of this scheme with $N = 3$. It is shown that Packet 4 and Packet 7 are not transmitted successfully after 4 times transmission, but they will be removed from the buffer and the next packet will be served. In other words, the result of the 3rd retransmission will not influence the operation of the sending queue. Therefore, we can use Fig.B.3 as an equivalent model, in which only the shadowed slots will influence the delay and backlog of the sending queue and are considered as *wasted*, the 3rd retransmission slots (such as slot 12 and 21) can be treated as a succeeded slot like slot 2 and 4.

Let us call all of the shadowed slots in equivalent model as "wasted slots", and let $I'_n(s,t)$ denote the number of such slots in an equivalent system during time period $(s,t]$. It is easy to find that there are at least $\frac{1}{N+1}(t-s)$ succeeded slots during $(s,t]$. Therefore, we can know that $I'_n(s,t) \leq \frac{N}{N+1} I_n(s,t)$. In addition, the average number of failed slots will become $\frac{N}{N+1} p_e(t-s)$, and γ'_n can be expressed as $\frac{N}{N+1}\gamma_n$. Let $\eta = \frac{N}{N+1}$ denote the retransmission efficiency. Then, we have

$$
\begin{aligned}
& P\left\{\sup_{0\leq s\leq t}\{I'(s,t)-\gamma^{I'}(s,t)\} > x\right\} \\
\leq\ & P\left\{\sup_{0\leq s\leq t}\{\eta I_n(s,t)-\eta\gamma_n^I(s,t)\} > \frac{x}{CT}\right\} \\
\leq\ & e^{-\frac{\theta x}{CT}} \frac{M_{\eta I_n(1)-\eta\gamma_n^I(1)}(\theta)}{1-M_{\eta I_n(1)-\eta\gamma_n^I(1)}(\theta)} \equiv f^{I'}(x)
\end{aligned}
$$

where $\langle f^{I'}, \gamma^{I'} \rangle$ is the stochastic arrival curve characterization of the *wasted* service process due to sensing error. Formally, we have proved the following result:

Lemma 2. *The* wasted *service process I under Max-N-RT has a stochastic arrival curve* $\langle f^{I'}, \gamma^{I'} \rangle$*, where*

$$\begin{aligned} \gamma^{I'}(s,t) &= \eta \cdot K \cdot p_e(t-s) \cdot CT \\ f^{I'}(x) &= e^{-\frac{\theta x}{CT}} \frac{e^{-\eta\theta K p_e}(p_e e^{\eta\theta} + 1 - p_e)}{1 - e^{-\eta\theta K p_e}(p_e e^{\eta\theta} + 1 - p_e)} \end{aligned}$$

for any $\theta > 0$*.*

In the same way, the stochastic arrival curve characterization of the corresponding *wasted* service processes due to mis-detection $I^{MD'}$ and false alarm $I^{FA'}$ under the Max-N-RT scheme can be found by replacing p_e in Lemma 2 respectively with ϕp_e and $(1-\phi)p_e$:

$$I^{MD'} \sim_{sac} \langle f^{MD'}, \gamma^{MD'} \rangle, I^{FA'} \sim_{sac} \langle f^{FA'}, \gamma^{FA'} \rangle.$$

B.3.2 Stochastic Service Curves of Users

For ease of expression, we have assumed that the channel is error-free. Under this assumption, the channel provides a constant strict service guarantee, i.e., $\hat{\beta} = Ct$ (for all $t \geq 0$), which can be considered as a special form of stochastic service curve with bounding function $\hat{g}(x) = 0$ for any $x \geq 0$. In the following, we apply a concept called *interference process* [6, 7] to facilitate finding stochastic service curves for PUs and SUs. Particularly, in a system with interference, the interfered service will be treated as *wasted* or cannot be used by the sender. For the considered cognitive radio network, both the performance of PUs and SUs is influenced by some interference processes. Particularly, for PUs, the *wasted* service process due to sensing error is an interference process. For SUs, in addition to the *wasted* service process due to sensing error, the packet arrival process of PUs can also be treated as an interference process.

The following result establishes the link between the interference process and the stochastic service guarantee.

Theorem 2. *For the considered cognitive radio network, if the interference process I to an input flow F (either* fl^p *or* fl^s*) has a stochastic arrival curve*

$\langle g^I, \beta^I(t)\rangle$, *then the network provides to the flow a stochastic service curve* $\langle g^I, Ct - \beta^I(t)\rangle$.

Proof. Let $R(t)$ and $R^*(t)$ denote the sum of inputs and outputs from flow F and interference process I, respectively, i.e., $R(t) = F(t) + I(t)$ and $R^*(t) = F^*(t) + I^*(t)$. Since the output traffic will not be larger than the input traffic, we have $F^*(t) \leq F(t)$, $I^*(t) \leq I(t)$ and $R^*(t) \leq R(t)$. It is easy to find that, for any $s \geq 0$,

$$\begin{aligned}
& F(s) \otimes \left(\hat{\beta}(s) - \beta^I(s)\right) - F^*(s) \\
= \;& (R(s) - I(s)) \otimes \left(\hat{\beta}(s) - \beta^I(s)\right) - (R^*(s) - I^*(s)) \\
= \;& \inf_{0\leq u\leq s} \left[R(u) + \hat{\beta}(s-u) - \beta^I(s-u) - I(u)\right] \\
& - (R^*(s) - I^*(s)) \\
\leq \;& \inf_{0\leq u\leq s} \left[R(u) + \hat{\beta}(s-u)\right] - R^*(s) \\
& + I(s) - \inf_{0\leq u\leq s} \left[\beta^I(s-u) + I(u)\right] \\
\leq \;& \sup_{0\leq u\leq s} \left[I(u,s) - \beta^I(s-u)\right]
\end{aligned}$$

where the last step is due to that for a constant rate server with rate C, it has been shown in the literature (e.g., see [7]) that $R \otimes \hat{\beta}(t) \leq R^*(t)$ for $\hat{\beta}(t) = C \cdot t$ which is the case here.

Since the interference process I has stochastic arrival curve $\langle g^I, \beta^I(t)\rangle$, then based on the definition of SAC, we have:

$$\begin{aligned}
& P\left\{F \otimes (\hat{\beta} - \beta^I) - F^*(s) > x\right\} \\
\leq \;& P\left\{\sup_{0\leq u\leq s} [I(u,s) - \beta^I(s-u)] > x\right\} \leq g^I(x)
\end{aligned}$$

which ends the proof. $\square$

Based on this theorem and different retransmission schemes, we can get the service guarantees provided to users in each scenario as summarized in Table B.1.

Table B.1: Stochastic Service Guarantee

RT Scheme	PUs Flow	SUs Flow
WO-RT	$\langle 0, \hat{\beta} \rangle$	$\langle f^P, \hat{\beta} - \alpha^P \rangle$
RT-S	$\langle f^{MD}, \hat{\beta} - \gamma^{MD} \rangle$	$\langle f^P \otimes f^I, \hat{\beta} - \alpha^P - \gamma^I \rangle$
MAX-N-RT	$\langle f^{MD'}, \hat{\beta} - \gamma^{MD'} \rangle$	$\langle f^P \otimes f^{I'}, \hat{\beta} - \alpha^P - \gamma^{I'} \rangle$

B.3.3 Performance Bounds

With the stochastic service curves obtained above, performance bounds for each type of users can be immediately obtained from Theorem 3 and are presented below.

Theorem 3. *For the considered network, suppose respectively flow fl^P and flow fl^S have stochastic arrival curves as $A^P(t) \sim_{sac} \langle f^P, \alpha^P \rangle$ and $A^S(t) \sim_{sac} \langle f^S, \alpha^S \rangle$. For the stochastic service curves received by them, we denote by $S^P_{RT}(t) \sim ssc\langle g^P_{RT}, \beta^P_{RT} \rangle$ and $S^S_{RT}(t) \sim ssc\langle g^S_{RT}, \beta^S_{RT} \rangle$, where $RT \in \{WO-RT, RT-S, MAX-N-RT\}$, and for each combination, the corresponding bounding function and service curve are found in Table B.1. Then the backlog $B(t)$ and delay $D(t)$ distribution bounds can be expressed as:*

$$P\{B^U(t) > x\} \le [f^U \otimes g^U_{RT}(x - \alpha^U \oslash \beta^U_{RT}(0))]_1$$
$$P\{D^U(t) > h(\alpha^U + x, \beta^U_{RT})\} \le [f^U \otimes g^U_{RT}(x)]_1$$

where $U \in \{P, S\}$.

B.4 Numerical Results

In Section B.3, the stochastic service curve for each flow is obtained. Given the stochastic arrival curve, performance bounds can be derived by using Theorem 3. In this section, we consider two types of input traffic: Poisson traffic and $(\sigma(\epsilon), \rho(\epsilon))$-constrained traffic. The $(\sigma(\epsilon), \rho(\epsilon))$-constrained traffic model is a general traffic and many types traffic can be represented by it [5], which include exponential on-off, Markov modulated process and effective bandwidth. Numerical results under different configurations are shown and discussed.

B.4.1 Numerical Results for Poisson Traffic

(Poisson Traffic.) Suppose all packets of a flow have the same size L and they arrive according to a Poisson process with mean arrival rate λ. Then the flow has a stochastic arrival curve $A(t) \sim_{sac} \langle f_{Pois}, rt \rangle$ for any $r > \lambda L$ with bounding function [8]:

$$f_{Pois}(x) = 1 - (1-a) \sum_{i=0}^{k} \left[\frac{[a(i-k)]^i}{i!} e^{-a(i-k)} \right]$$

where $a = \frac{\lambda L}{r}$ and $k = \lceil \frac{x}{L} \rceil$.

First, the influence of average rate r in Poisson arrival is studied, as Fig.B.4 shows, where system capacity C is set as $500kbps$, packet length for both PUs and SUs flows is set to be 8kbits, and the arrival rate is 10 packets per second. Spectrum sensing error probability in each slot, i.e., p_e, is 1%, and the probability that this is a mis-detection is 50%. Parameter K is set as 2.1 and θ is 1.3 for both sensing error process and mis-detection process. Retransmission is not considered. When applying Theorem 3 in this case, it is required that the average rate for PUs and SUs flow should fulfill $r^P \leq C$ and $r^P + r^S \leq C$, otherwise no guarantee will be provided. Therefore, r^P is set to be within $[200, C]$ and r^S set to $C - r^P$.

It is obvious that the results for the PUs flow locate lower than SUs flow, which means the PUs flow is provided with better QoS guarantee, since it is assigned with high priority. In addition, it is found that smaller bounds will be provided to the PUs flow when increasing r^P, which is straightforward since r^P denotes average service rate. On the other hand, bounds for SUs flows become smaller first and then larger when increasing r^S. This is due to the fact that the PUs' bounding function also has effect on SUs' bounds. Higher r^S means smaller r^P, more backlog in PUs buffer, and more time slots will be occupied by PUs, which will lead to more backlog in the SUs flow. Therefore, optimal bounds for the SUs flow are tradeoff between PUs' and SUs' service guarantees. Under the current configuration, the balance point locates in the middle where $r^P = r^S = \frac{C}{2}$. As for delay, similar results can be obtained which are not shown here due to limited space.

Fig.B.5 shows backlog bounds and delay bounds for PUs and SUs considering sensing errors and retransmission at the same time. When we consider the results for one flow, it is found that the distribution bounds of

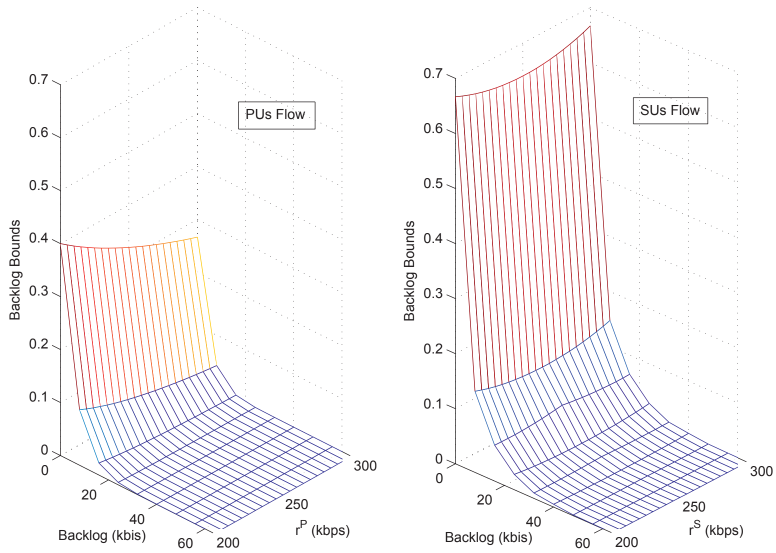


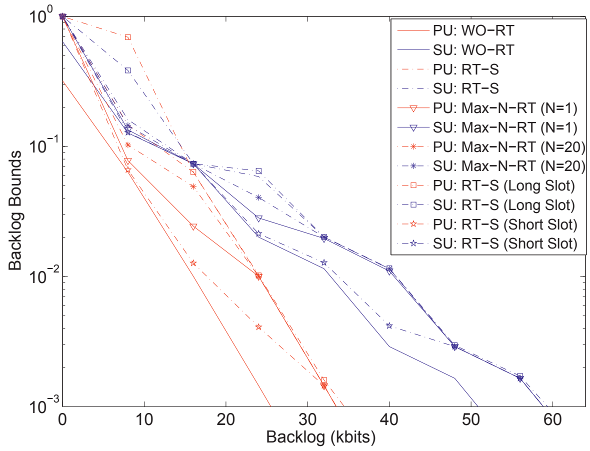
Figure B.4: Backlog Bounds for Poisson Arrivals with Different r^P

WO-RT scheme are smaller than RT-S scheme, and results of Max-N-RT locate between them, which is consistent with what we can expect from the description of each scheme. However, WO-RT scheme and Max-N-RT scheme may introduce transmission errors or retransmissions in higher layers. In addition, it is straightforward that Max-N-RT scheme converges to RT-S when the maximum retransmission time N is set to infinity; and here results for Max-N-RT scheme with $N = 20$ are also plotted, which lie closer to the RT-S scheme's results than Max-N-RT scheme's results when $N = 1$.

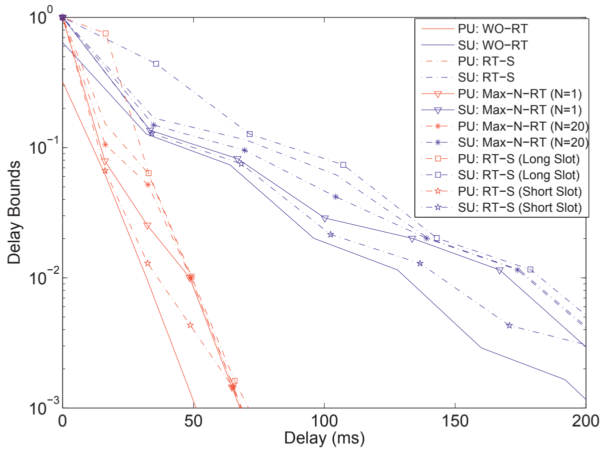
Furthermore, slot length also has significant influences on the system performance. If slot is longer, collision caused by mis-detection will lead to larger backlog and delay because one collision wastes more time; while on the other side, short slot results in better performance guarantee. The results when $T = 2.5ms$ and $T = 1.5ms$ are shown, where results with $T = 1.5ms$ give smaller distribution bounds.

B.4.2 Numerical Results for $(\sigma(\varepsilon), \rho(\varepsilon))$-Constrained Traffic

($(\sigma(\varepsilon), \rho(\varepsilon))$-Constrained Traffic.) If a flow is $(\sigma(\varepsilon), \rho(\varepsilon))$ upper constrained, then it has a stochastic arrival curve $\alpha(t) = \rho(\varepsilon) \cdot t + \sigma(\varepsilon)$ with bounding function $f(x) = e^{-\varepsilon x}$ [7].



(a) Backlog



(b) Delay

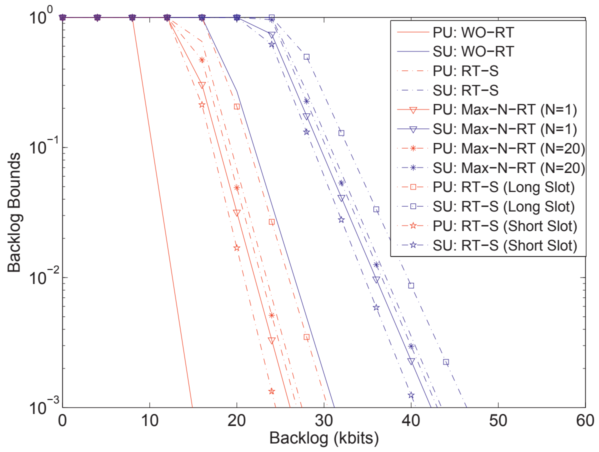
Figure B.5: Numerical Results for Poisson Traffic

This model can be used to model Markov modulated processes but cannot be easily analyzed by using traditional queueing theory. However, it is possible to use stochastic network calculus to obtain distribution bounds [5, 7].

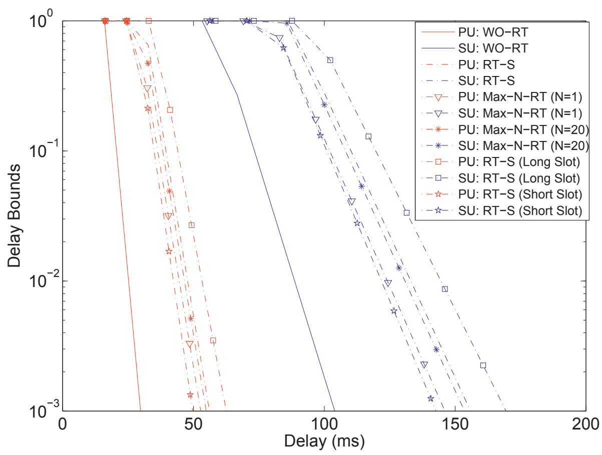
In the following, average arrival rate $\rho(\varepsilon)$ and maximum burst $\sigma(\varepsilon)$ for both PUs and SUs flow are set to fixed values as 200kbps and 8kbits, respectively; while ε in the bounding function is set to 1. In addition, the channel capacity is 500kbps and the slot length is 2ms. Numerical results are shown in Fig.B.6, where similar trends as Poisson traffic can be found. In short, WO-RT scheme provides better physical layer backlog and delay guarantee; while smaller slot length and less retransmission time in RT-S and MAX-N-RT schemes can improve the bounds.

B.5 Conclusion

In this paper, performance analysis for a cognitive radio network has been conducted. The network is modeled as a preemptive priority queueing system, with imperfect spectrum sensing and different retransmission schemes. The spectrum sensing error process is modeled as a combination of two processes, mis-detection and false alarm, for which stochastic arrival curves of the corresponding "wasted" service process are found. Three physical layer retransmission schemes are discussed, including without retransmission, retransmission until success and maximum-N-time retransmission. Stochastic service curves provided to both PUs and SUs are proved under different retransmission schemes together with performance bounds on backlog and delay, which can be applicable to many types of traffic. Numerical results are shown for two types of input traffic, Poisson traffic and (σ, ρ)-constrained traffic, where further discussions are made. We believe that these results will shed light on deeper understanding of cognitive radio networks, and on the design of optimal retransmission schemes in such networks.



(a) Backlog



(b) Delay

Figure B.6: Numerical Results for $(\sigma(\theta), \rho(\theta))$-Constrained Traffic

Bibliography

[1] J. III Mitola and G.Q. Jr. Maguire. Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6:13–18, August 1999.

[2] I. Suliman and J. Lehtomaki. Queueing analysis of opportunistic access in cognitive radios. In *Proc. 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management*, pages 153 – 157, Aalborg, Denmark, May 2009.

[3] Caoxie Zhang, Xinbing Wang, and Jun Li. Cooperative cognitive radio with priority queueing analysis. In *Proc. IEEE International Conference on Communications (ICC '09)*, pages 1 – 5, Dresden, Germany, 14–18, 2009.

[4] Junghyun Heo, Jungchae Shin, Jihee Nam, Yutae Lee, Joon Goo Park, and Ho-Shin Cho. Mathematical analysis of secondary user traffic in cognitive radio system. In *Proc. IEEE 68th Vehicular Technology Conference (VTC 2008-Fall)*, pages 1 – 5, Calgary, Canada, September 21–24, 2008.

[5] Cheng-Shang Chang. *Performance Guarantees in Communication Networks*. Springer, 2000.

[6] Yuming Jiang. A basic stochastic network calculus. In *Proc. of the conference on Applications, technologies, architectures, and protocols for computer communications*, pages 123–134, September 2006.

[7] Yuming Jiang and Yong Liu. *Stochastic Network Calculus*. Springer, 2008.

[8] Y Jiang, Q Yin, Y Liu, and S Jiang. Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks. *Computer Networks*, 53(12):2011–2021, August 2009.

[9] Shensheng Tang and Brian L. Mark. Modeling and analysis of opportunistic spectrum sharing with unreliable spectrum sensing. *IEEE Trans. on Wireless Communications*, 8(4):1934–1943, April 2009.

Appendix C

Publication C

Yuehong Gao, Yuming Jiang, Tao Lin and Xin Zhang; Performance Bounds for a Cognitive Radio Network with Network Calculus Analysis; 2010 International Conference on Network Infrastructure and Digital Content, September 2010.

Abstract

In this paper, we use stochastic network calculus to analyze the performance bounds for a cognitive radio network with two classes of input traffic. First, stochastic service curves for primary users and secondary users are obtained based on the system model and stochastic network calculus. Then, we derive the general expressions of backlog and delay bounds for both primary and secondary users under two methods, i.e., min-plus convolution and independent case analysis. Finally, numerical results as well as simulation results are compared and discussed.

Keywords

Cognitive radio, Performance bound, Stochastic network calculus

C.1 Introduction

Spectrum is valuable as well as scarce resource in wireless communication systems. Currently, the spectrum is assigned in a fixed manner, which has been found to be quite inefficient, since a large portion of spectrum is underutilized. Therefore, cognitive radio network [1] was proposed, in which the cognitive users (or secondary users, SUs) try to make use of the spectrum holes when the licensed users (or primary users, PUs) do not transmit data on the assigned spectrum. By employing these spectrum holes, secondary users can utilize some unused spectrum resource to transmit data and thus system utilization can be improved. But some efforts are needed in order to figure out what service guarantee can be provided to both primary users and secondary users. It is hence important to conduct performance analysis for both primary and secondary users.

Queuing theory has been employed in performance analysis of cognitive radio networks [2–5], where some results have been obtained, such as packet waiting time in queue and delay. However, the focus has mainly been on average values in the steady states. Some frameworks for performance evaluation have been proposed. For example, a framework for performance evaluation of cognitive radios is proposed in [6], where the performance metrics, utility function and methodology are discussed. Another framework for performance evaluation of cognitive radio networks in heterogeneous environments is presented in [7]. These frameworks show the evaluation methods, but they don't provide clear clue about how to obtain the evaluation results.

The objective of this paper is to conduct the performance analysis of a cognitive radio network from a new viewpoint, which lists out the specific steps, theorems and results. Specifically, stochastic network calculus [8–11] is used to analyze performance distribution bounds. We believe our work is a strong supplement to the existing references.

The paper is organized as follows. First, the cognitive radio network is modeled as a preemptive queueing problem in Section C.2, where basic assumptions and stochastic network calculus basics are presented. Then, in Section C.3, we analyze stochastic service guarantees provided to the primary and secondary users, followed by general expressions of performance bounds. Lastly, numerical results are shown in Section C.4 and the summary is given in Section C.5.

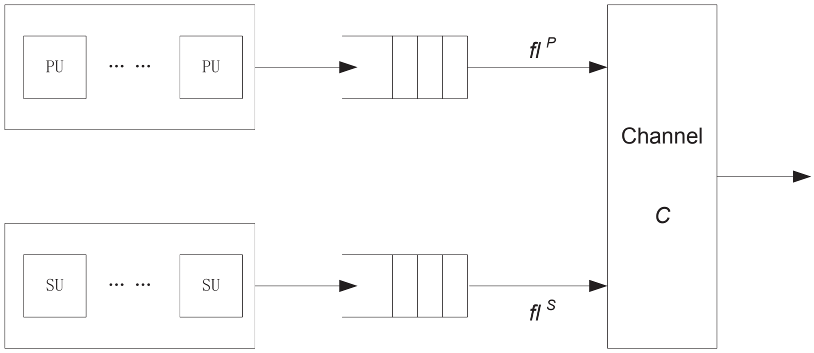


Figure C.1: System Model

C.2 System Model

C.2.1 System Model

In this paper, we consider a cognitive radio network with two input flows as illustrated in Fig.C.1, where fl^P and fl^S represent the aggregated flows from primary users and secondary users, respectively. For ease of expression, the wireless channel is assumed to be error free with a constant service rate C, which can be thought as a special case of stochastical channel. The analysis can be easily extended to consider a real stochastical channel that can be expressed with a stochastic service curve. The primary users' flow has preemptive priority over the secondary users' flow. If one packet arrives into the system and cannot be transmitted immediately, it will be stored in the corresponding buffer in a First-In-First-Out (FIFO) manner, where the buffer is assumed to be large enough and therefore no packet will be dropped.

Generally, packet arrivals are stochastic processes, which will only lead to stochastic service guarantees. While the system model described above has already been overly simplified, to the best of our knowledge, it is difficult (if not impossible) to obtain explicit results from the traditional queueing theory, particularly when the involving stochastic processes are not Poisson or with exponentially distributed rates. To address this problem, we resort to stochastic network calculus.

C.2.2 Stochastic Network Calculus Basics

Stochastic network calculus theory is a newly developed queueing theory for service guarantee analysis (e.g., [8–10] and references therein), which contains two fundamental concepts, stochastic arrival curve (SAC) and stochastic service curve (SSC).

In stochastic network calculus, a stochastic arrival curve is used to describe the stochastic characteristics of a flow. There are several definitions for SAC, and in this paper the following definition is used, which explores the virtual backlog property of deterministic arrival curve [10].

Definition 1. ***(Stochastic Arrival Curve).*** *A flow $A(t)$ is said to have a virtual-backlog-centric (v.b.c) stochastic arrival curve $\alpha \in F$ (F: the set of non-negative wide-sensing increasing functions) with bounding function $f \in \bar{F}$ ($\bar{F}$: the set of non-negative wide-sensing decreasing functions), denoted by $A(t) \sim_{sac} \langle f, \alpha \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\left\{ \sup_{0 \leq s \leq t} \{A(s,t) - \alpha(t-s)\} > x \right\} \leq f(x).$$

In Definition 1, $A(s,t)$ denotes the cumulative amount of traffic of the flow during period $(s,t]$, $A(t) = A(0,t)$, and $\alpha(t)$ is a non-decreasing function.

While the stochastic arrival curve describes the traffic, the stochastic service curve shows the service guarantee provided by a server. Similarly with SAC, stochastic service curve can also be defined in different ways. In this paper, we use the following one [10].

Definition 2. ***(Stochastic Service Curve).*** *A system S is said to provide a stochastic service curve $\beta \in F$ with bounding function $g \in \bar{F}$, denoted by $S \sim_{ssc} \langle g, \beta \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\{A \otimes \beta(t) - A^*(t) > x\} \leq g(x).$$

Here, $A \otimes \beta(t) \equiv \inf_{0 \leq s \leq t} \{A(s) + \beta(t-s)\}$, and $A^*(t)$ denotes the cumulative amount of output traffic up to time t.

Given SAC and SSC, the following bounds have been derived under stochastic network calculus [10]:

Theorem 1. ***(Performance Distribution Bounds).*** *Consider a system S with input A. Suppose the input has a v.b.c stochastic arrival curve as*

$A(t) \sim_{sac} \langle f, \alpha \rangle$; and server S provides the input with a stochastic service guarantee as $S \sim_{ssc} \langle g, \beta \rangle$. Then for any $t \geq 0$ and all $x \geq 0$, the backlog $B(t)$ and delay $D(t)$ is bounded by:

$$P\{B(t) > x\} \leq [f \otimes g(x - \alpha \odot \beta(0))]_1$$
$$P\{D(t) > h(\alpha + x, \beta)\} \leq [f \otimes g(x)]_1$$

where $\alpha \odot \beta\,(0) = \sup_{u \geq 0} \{\alpha\,(u) - \beta\,(u)\}$, $h(\alpha + x, \beta) = \sup_{s \geq 0}\{\inf\{\tau \geq 0 : \alpha(s) + x \leq \beta(s + \tau)\}\}$ and $[\cdot]_1$ denotes $\max\,(\min\,(\cdot, 1)\,, 0)$. This method is referred as *min-plus convolution analysis* "$\otimes$" in this paper.

Furthermore, if the arrival process A and the service process S are independent with each other, then the backlog and delay bounds can be expressed as:

$$P\,\{B\,(t) > x\} \leq 1 - \overline{f} * \overline{g}\,(x - \alpha \odot \beta\,(0))$$
$$P\,\{D\,(t) > h\,(\alpha + x, \beta)\} \leq 1 - \overline{f} * \overline{g}\,(x)$$

where $\overline{f}\,(x) = 1 - [f\,(x)]_1$, $\overline{g}\,(x) = 1 - [g\,(x)]_1$ and $\overline{f} * \overline{g}\,(x) = \int_0^x \overline{g}(x - y) d\overline{f}\,(y)$. We use *independent case analysis* "$*$" to represent this method.

In order to apply stochastic network calculus results to performance analysis of cognitive radio network, a critical challenge is to find stochastic service curve for both PUs and SUs. Major contribution of this paper is in finding out stochastic service curve provided to both PUs and SUs, which are presented below.

C.3 Performance Analysis

In this section, performance analysis of the considered cognitive radio network is conducted. The focus is on finding probabilistic bounds on backlog and delay of both primary users (PUs) and secondary users (SUs), and the theoretical tool we rely on is stochastic network calculus. As highlighted in Section C.2.2, in order to apply Theorem 1, it is essential to find stochastic service curves for both PUs and SUs, as presented below. Then we obtain backlog and delay bounds for both PUs and SUs based on stochastic network calculus results.

C.3.1 Stochastic Service Curves of Users

For ease of expression, we have assumed that the channel is error-free. Under this assumption, the channel provides a constant strict service guarantee, i.e., $\widehat{\beta} = Ct$ (for all $t \geq 0$), which can be considered as a special form of stochastic service curve with bounding function $\widehat{g}(x) = 0$ for any $x \geq 0$. For the considered cognitive radio network, the performance of PUs is only influenced by the channel capacity due to the preemptive priority, i.e., flow fl^P is provided with a constant service guarantee $S^P \sim \left\langle 0, \beta^P = \widehat{\beta} \right\rangle$. While for SUs, in addition to the channel capacity, the packet arrival process of PUs can also influence the performance of SUs.

The following result establishes the link between the stochastic arrival process of PUs and the stochastic service guarantees provided to SUs.

Theorem 2. *For the considered cognitive radio network, if flow fl^P has a v.b.c. stochastic arrival curve $\langle f^P, \alpha^P \rangle$, then the network provides to flow fl^P a stochastic service curve $\left\langle f^P, \widehat{\beta} - \alpha^P \right\rangle$.*

Proof. Let $A(t)$ and $A^*(t)$ denote the sum of inputs and outputs from PUs and SUs, respectively, i.e., $A(t) = A_P(t) + A_s(t)$ and $A^*(t) = A^*_P(t) + A^*_s(t)$. Since the output traffic will not be larger than the input traffic, we have $A(t) \leq A^*(t)$, $A^*_P(t) \leq A_P(t)$ and $A^*_S(t) \leq A_S(t)$. It is easy to find that, for any $s \geq 0$,

$$\begin{aligned}
& A_S \otimes \left(\widehat{\beta} - \alpha^P\right) - A^*_S(s) \\
= \; & (A - A_P) \otimes \left(\widehat{\beta} - \alpha^P\right)(s) - (A^* - A^*_P)(s) \\
= \; & \inf_{0 \leq u \leq s} \left[A(u) + \widehat{\beta}(s-u) - \alpha^P(s-u) - A_P(u)\right] \\
& - (A^*(s) - A^*_P(s)) \\
\leq \; & \inf_{0 \leq u \leq s} \left[A(u) + \widehat{\beta}(s-u)\right] - A^*(s) + A^*_P(s) \\
& - \inf_{0 \leq u \leq s} \left[\alpha^P(s-u) + A_P(u)\right] \\
= \; & \left[A \otimes \widehat{\beta}(s) - A^*(s)\right] + \sup_{0 \leq u \leq s} \left[A^*_P(u,s) - \alpha^P(s-u)\right] \\
\leq \; & \sup_{0 \leq u \leq s} \left[A^*_P(u,s) - \alpha^P(s-u)\right]
\end{aligned}$$

where the last step is due to that for a constant rate server with rate C, it has been shown in the literature (e.g., see [10]) that $A \otimes \widehat{\beta}(t) \leq A^*(t)$ for $\widehat{\beta} = Ct$ which is the case here.

Since the arrival process of the flow fl^P has stochastic arrival curve $\langle f^P, \alpha^P \rangle$, then based on the definition of SAC, we can obtain:

$$\begin{aligned} & P\left\{A_S \otimes \left(\widehat{\beta} - \alpha^P\right) - A_S^* (s) > x\right\} \\ \leq & P\left\{\sup_{0\leq u\leq s} \left[A_P^* (u,s) - \alpha^P (s-u)\right] > x\right\} = f^P (x) \end{aligned}$$

which ends the proof. □

C.3.2 Performance Bounds for PUs and SUs Flow

With stochastic service curves obtained above, performance bounds for both PUs and SUs can be immediately obtained from Theorem 1, which is further summarized in Theorem 3 presented below.

Theorem 3. *For the considered network, suppose respectively flow fl^P and flow fl^S have stochastic arrival curves as $A^P \sim \langle f^P, \alpha^P \rangle$ and $A^S \sim \langle f^S, \alpha^S \rangle$. As it is proved above, the network provides flow fl^P and flow fl^S a stochastic service curve as $S^P \sim \langle g^P = 0, \beta^P = Ct \rangle$ and $S^S \sim \langle g^S = f^P,$ $\beta^S = Ct - \alpha^P \rangle$, respectively. Then backlog $B(t)$ and delay $D(t)$ distribution bounds can be expressed as:*

$$\begin{aligned} P\left\{B^u (t) > x\right\} &\leq \left[f^u \otimes g^u \left(x - \alpha^u \odot \beta^u (0)\right)\right]_1 \\ P\left\{D^u (t) > h\left(\alpha^u + x, \beta^u\right)\right\} &\leq \left[f^u \otimes g^u (x)\right]_1 \end{aligned}$$

where $u \in \{P, S\}$. Furthermore, the bounds can be improved when the PUs flow and SUs flow are independent:

$$\begin{aligned} P\left\{B^u (t) > x\right\} &\leq 1 - \overline{f^u} * \overline{g^u} \left(x - \alpha^u \odot \beta^u (0)\right) \\ P\left\{D^u (t) > h\left(\alpha^u + x, \beta^u\right)\right\} &\leq 1 - \overline{f^u} * \overline{g^u} (x) \end{aligned}$$

C.4 Results Analysis

In Section C.3, the stochastic service curve for each flow is obtained. Given the stochastic arrival curve, performance bounds can be derived by using

Theorem 3. In this section, due to the limited space, we only consider one traffic model, Poisson traffic, which has been widely used in many literatures. However, the performance bounds obtained above also can be employed to other types of traffic, such as $(\sigma(\theta), \rho(\theta))$-constrained traffic model, which is a general traffic and many types of traffic can be represented by it [11], including exponential on-off, Markov modulated process and effective bandwidth. Numerical results as well as simulation results are presented and discussed below.

C.4.1 Poisson Traffic Model

Suppose all packets of a flow have the same size L and they arrive according to a Poisson process with mean arrival rate λ. Then the flow has a v.b.c stochastic arrival curve $A(t) \sim_{sac} \langle f_1(x), \alpha_1(t)\rangle$ for any $r > \lambda L$, $\alpha_1(t) = rt$ with bounding function [9]:

$$f_1(x) = 1 - (1-a)\sum_{i=0}^{k}\left[\frac{[a(i-k)]^i}{i!}e^{-a(i-k)}\right]$$

where $a = \lambda L/r$ and $k = ceiling(x/L)$.

However, it is easy to know that the bounding function $f_1(x)$ is not continuous and thus non-differentiable, which prevents to use the independent case analysis. Therefore, we rely on the effective bandwidth and obtain an approximation modeling for Poisson traffic as shown below [11, 12].

Suppose all packets of a flow have the same size L and they arrive according to a Poisson process with mean arrival rate λ. Then the flow has a v.b.c stochastic arrival curve $A(t) \sim_{sac} \langle f_2(x), \alpha_2(t)\rangle$ for any $\theta > 0$, $\theta_1 \geq 0$, and $\alpha_2(t) = \lambda t\left(e^{\theta L} - 1\right)/\theta + \theta_1 t$, with bounding function

$$f_2(x) = e^{-\theta\theta_1}e e^{-\theta x}$$

where θ and θ_1 are free parameters and need to be optimized.

In the results presented below, the channel capacity is supposed to be $500kbps$, the time slot is set as $10ms$, the arrival rate for both PUs flow and SUs flow is set to 20 packets per second, and the packet length is $5kbits$. Hence, one packet can be served in one time slot. Furthermore, θ and θ_1 are optimized so that the bounding function is the tightest.

C.4.2 Results for PUs Flow

For PUs flow, the channel provides a deterministic service guarantee, and therefore min-plus convolution and independent case analysis result in the

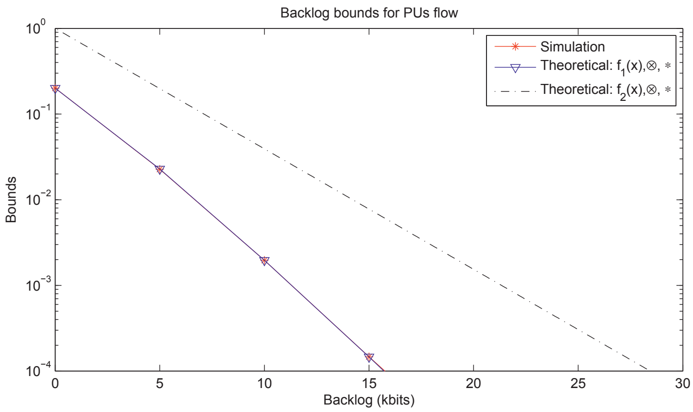


Figure C.2: Backlog Bounds for PUs Flow

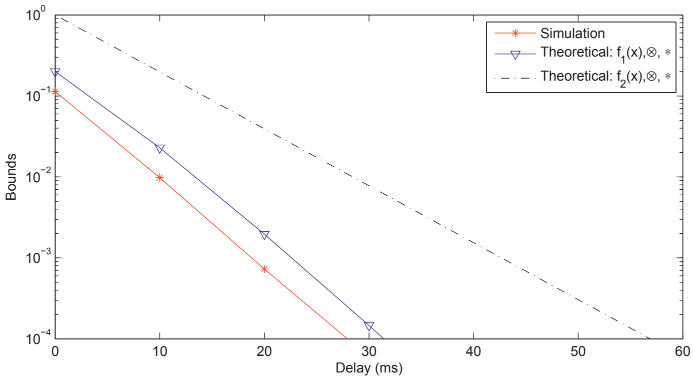


Figure C.3: Delay Bounds for PUs Flow

same performance distribution bounds. In addition, the theoretical analysis results closely match the simulation results. Particularly, the theoretical backlog bounds obtained from $f_1(x)$ coincides with the simulation results. The results from $f_2(x)$ are relatively loose, since $f_2(x)$ is just an approximation. Fig.C.2 and Fig.C.3 summarize the backlog bounds and delay bounds, respectively.

C.4.3 Results for SUs Flow

Due to the existence of PUs flow, the service provided to SUs flow is a stochastic process. Therefore, the theoretical results will not be consistent

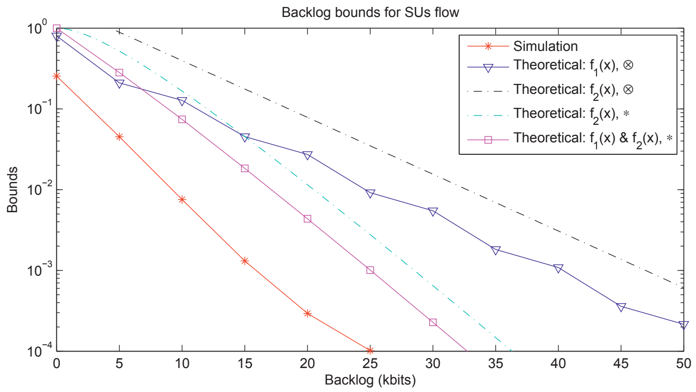


Figure C.4: Backlog Bounds for SUs Flow

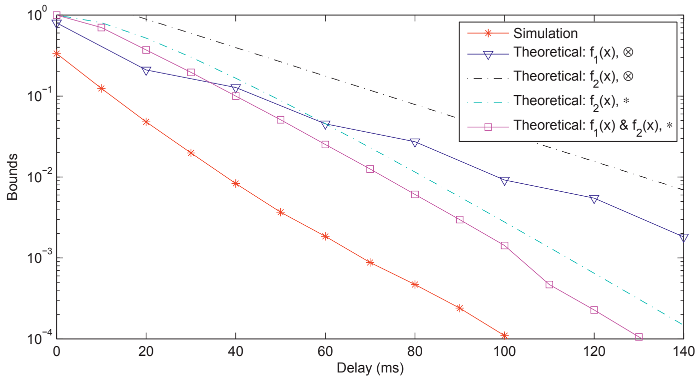


Figure C.5: Delay Bounds for SUs Flow

with the simulation results. Different v.b.c stochastic arrival curves can be used for the two input flows. When we use the stochastic arrival curve $\langle f_1(x), \alpha_1(x)\rangle$ for both PUs flow and SUs flow, then only the min-plus convolution method can be employed, which gives a medium bound as shown in Fig. C.4 and Fig.C.5. While when the $\langle f_2(x), \alpha_2(x)\rangle$ is used for the two flows, both the min-plus convolution and independent case analysis can be obtained, and the independent case analysis shows better results. Besides, a mixture of $\langle f_1(x), \alpha_1(x)\rangle$ and $\langle f_2(x), \alpha_2(x)\rangle$ is also possible for the independent case analysis, which shows the best bound as we can notice from Fig.C.4 and Fig.C.5.

Although these bounds are not exactly the same as the simulation results, they tell us the upper bounds, where the error is controlled within one order of magnitude.

C.5 Conclusions

In this paper, performance analysis for a cognitive radio network has been conducted by using stochastic network calculus under two methods, i.e., min-plus convolution and independent case analysis. The network is modeled as a preemptive priority queueing system with two input flows. First, the stochastic service curves provided to PUs and SUs are proved, which can be applicable to many types of traffic. Then performance bounds for both PUs and SUs are obtained. Numerical results and simulation results are shown and discussed for Poisson traffic. It is found that the network calculus analysis gives tight (or exact) distribution bounds for the PUs flow. As for the SUs flow, the independent case analysis, under a mixture usage of Poisson traffic model, presents better results than the other configurations.

Bibliography

[1] J. III Mitola and G.Q. Jr. Maguire. Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6:13–18, August 1999.

[2] I. Suliman and J. Lehtomaki. Queueing analysis of opportunistic access in cognitive radios. In *Proc. 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management*, pages 153 – 157, Aalborg, Denmark, May 2009.

[3] Caoxie Zhang, Xinbing Wang, and Jun Li. Cooperative cognitive radio with priority queueing analysis. In *Proc. IEEE International Conference on Communications (ICC '09)*, pages 1 – 5, Dresden, Germany, 14–18, 2009.

[4] Junghyun Heo, Jungchae Shin, Jihee Nam, Yutae Lee, Joon Goo Park, and Ho-Shin Cho. Mathematical analysis of secondary user traffic in cognitive radio system. In *Proc. IEEE 68th Vehicular Technology Conference (VTC 2008-Fall)*, pages 1 – 5, Calgary, Canada, September 21–24, 2008.

[5] M. Rashid, E. Hossain M. Hossain, and V. Bhargava. Opportunistic spectrum scheduling for multiuser cognitive radio: a queueing analysis. *IEEE Transactions on Wireless Communications*, 8(10):5259 – 5269, 2009.

[6] J. Neel Y. Zhao, S. Mao and J. Reed. Performance evaluation of cognitive radios: Metrics, utility functions, and methodology. In *Proceedings of the IEEE*, pages 642 – 659, Vienna, VA, April 2009.

[7] J. Neel Y. Zhao, S. Mao and J. Reed. Framework for performance evaluation of cognitive radio networks in heterogeneous environments. In *Canadian Conference on Electrical and Computer Engineering*, pages 199 – 203, 2009.

[8] Yuming Jiang. A basic stochastic network calculus. In *Proc. of the conference on Applications, technologies, architectures, and protocols for computer communications*, pages 123–134, September 2006.

[9] Y. Liu Y. Jiang, Q. Yin and S. Jiang. Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks. *Computer Networks*, 53(12):2011–2021, August 2009.

[10] Yuming Jiang and Yong Liu. *Stochastic Network Calculus*. Springer, 2008.

[11] Cheng-Shang Chang. *Performance Guarantees in Communication Networks*. Springer, 2000.

[12] M. Kelly. Notes on effective bandwidths. *Stochastic Networks: Theory and Applications, Royal Statistical Society Lecture Notes Series*, 4, 1996.

Appendix D

Publication D

Yuehong Gao, Jinxing Yang, Xin Zhang and Yuming Jiang; Capacity Limits for a Cognitive Radio Network under Fading Channel; Springer Lecture Notes in Computer Science: IFIP Networking 2011 workshops, pp. 42 - 51, May 2011.

Abstract

In this paper, performance evaluation of a cognitive radio network is conducted. The analysis is based on stochastic network calculus. The system is supposed to work in a Time Division Multiple Access (TDMA) mode with fixed slot length. The wireless channel is modeled as a Gilbert-Elliott (GE) fading channel, where the channel quality transits between state *ON* and state *OFF* according to a Markov chain. Spectrum sensing errors, which can be classified into mis-detection and false-alarm, are taken into consideration. Particularly, a stochastic arrival curve for spectrum sensing error process, and a stochastic service curve for GE channel, are derived. In addition, performance distribution bounds are obtained based on stochastic network calculus. Furthermore, numerical calculations are made to show the capacity limits under delay constraints.

Keywords

Capacity, Cognitive radio, GE channel, Performance bound, Stochastic network calculus

D.1 Introduction

Nowadays, cognitive radio has become a promising technology, since it provides a solution to improve the spectrum utilization efficiency. In a cognitive radio network, the secondary users (SUs) sense the spectrum before transmitting on it, and if they find available spectrum holes, they will make use of those resources. The sensing results, however, may not exactly match with the real condition. In other words, spectrum sensing error happens sometimes, which leads to collision between the primary transmission and the secondary transmission or waste of transmission opportunities for secondary users. Therefore, physical layer re-transmission is needed in order to deal with such collisions.

In this paper, we consider a cognitive radio network with two classes of input traffic, the aggregated flow from primary users and the one from secondary users, as shown in Fig.D.1. The system works in a slotted mode with fixed slot length T. The flow from PUs has higher priority over SUs flow to be served. Secondary users try to sense the channel and act based on the sensing results. Sensing errors may happen and will affect the performance.

How to analyze the performance guarantee for each class of users is a key issue in cognitive radio networks. Some queueing theory based studies have been made such as in [1], where delay and queue related parameters are derived. In [2–4], we made some analysis by using network calculus to derive the backlog and delay distribution bounds for a simplified system model. Network calculus is an approach to deal with flow problems in communication networks, which was introduced by Cruz in 1991 [5]. After about 20-year development, network calculus has evolved into two branches: the deterministic branch and the stochastic branch. In this study, stochastic network calculus is employed. In stochastic network calculus, stochastic

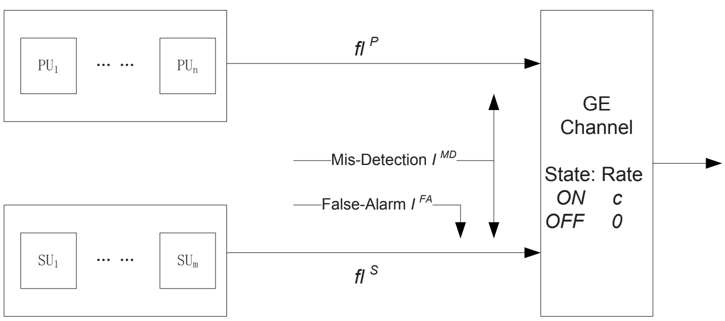


Figure D.1: System Model

arrival curve (sac) and stochastic service curve (ssc) are used to describe the characteristics of a flow and a server, respectively. Based on sac and ssc, performance bounds can be derived.

In [2–4], the wireless channel was assumed to be a constant error-free channel. However, this assumption is not practical in real systems, because the essential characteristic of a wireless channel is its fading nature. In this paper, this shortage is overcome by considering the Gilbert-Elliott (GE) channel model [6, 7], which has two states, Good (G) and Bad (B). In the discrete time model, these two states transit between each other according to a Markov chain. In addition, the sensing error process is re-modeled compared with the model in [2], so that tighter performance bounds can be obtained.

The paper is organized as follows. In Section D.2, stochastic network calculus basics are introduced, where the stochastic arrival curve for sensing error process and the stochastic service curve for GE fading channel are derived, and the delay bound for each flow is also obtained. Section D.3 discusses the numerical results and capacity limits, followed by a summary in Section D.4.

D.2 Stochastic Network Calculus Analysis

D.2.1 Traffic Modeling

Stochastic arrival curve can be defined from different aspects [8]. Here, we explore the *virtual-backlog-centric* based definition as follows.

Definition 1. ***(Stochastic Arrival Curve).*** *A flow $A(t)$ is said to have a virtual-backlog-centric (v.b.c) stochastic arrival curve $\alpha \in F^1$ with bounding function $f \in \bar{F}^2$, denoted by $A(t) \sim_{sac} \langle f, \alpha \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\left\{\sup_{0\leq s\leq t}\{A(s,t)-\alpha(t-s)\} > x\right\} \leq f(x), \tag{D.1}$$

where $A(s,t)$ denotes the cumulative amount of traffic during period $(s,t]$, $A(0,t)$ is written as $A(t)$ for short, and $\alpha(t)$ is a non-decreasing function.

[1] F: the set of non-negative wide-sensing increasing functions
[2] $\bar{F}$: the set of non-negative wide-sensing decreasing functions

Stochastic arrival curves of many traffic models have been derived, such as in [8]. Therefore, we just employ the models directly in this paper, and put our efforts on other aspects.

D.2.2 Modeling of Spectrum Sensing Error Process

In cognitive radio networks, secondary users sense the spectrum and utilize the available white spaces for their transmissions. However, sensing errors may happen, which lead to *transmission collision* or *opportunity waste*. To be specific, sensing errors can be classified into two types, mis-detection (MD) and false-alarm (FA). Mis-detection means that the spectrum is occupied by PUs but the spectrum sensing result says it is available for SUs, which will result in transmission collision and influence both PUs' and SUs' current transmission. However, false alarm occurs in the opposite way, when SUs believe that the spectrum is being used by PUs but actually the spectrum is idle. As a result, SUs will miss those transmission opportunities.

Based on the facts described above, the error process can be considered as a special type of input traffic, which also competes for the transmission resource and has the highest priority.In this part, the stochastic arrival curve for sensing error process will be derived.

Here, we consider a slotted system with fixed slot length T, and the probability that sensing error happens in one time slot is supposed to be p. By further assuming the independency between the appearances of sensing errors in adjacent slots, the impairment arrival process $I(t)$ is a Lévy process, where $I(t)$ denotes the number of sensing errors during slot $(0, t]$. Then, according to Lemma 1 in the appendix, process $I(t)$ has a v.b.c stochastic arrival curve, denoted by $I(t) \sim_{sac} \langle f^I, \alpha^I \rangle$, where

$$\begin{aligned} f^I(x) &= e^{-\theta\theta_1} e^{-\theta x} && \text{(D.2)} \\ \alpha^I(t) &= \left[\frac{1}{\theta} \log E[e^{\theta I(1)}] + \theta_1\right] \cdot t \equiv [\rho^I(\theta) + \theta_1] \cdot t && \text{(D.3)} \end{aligned}$$

for free parameters $\forall\theta_1 \geq 0$ and $\forall\theta > 0$.

In each slot, the happening of sensing error has a Bernoulli distribution with parameter p. Therefore, $\rho^I(\theta)$ in Eq.(D.3) can be expressed as

$$\rho^I(\theta) = \frac{1}{\theta} \log(1 - p + pe^{\theta\sigma}), \tag{D.4}$$

where σ denotes the number of packets that are not transmitted successfully in a slot due to a sensing error. Furthermore, mis-detection process and false-alarm process have the same characteristic as the sensing error process, and the only difference is the happening probability. In later parts, the following notations are used to represent the stochastic arrival curves of mis-detection process and false-alarm process:

$$I^{MD}(t) \sim_{sac} \langle f^{MD}, \alpha^{MD} \rangle, \; I^{FA}(t) \sim_{sac} \langle f^{FA}, \alpha^{FA} \rangle, \tag{D.5}$$

where f^{MD} and f^{FA} have the same form as in Eq.(D.2), α^{MD} and α^{FA} can be obtained by replacing the probability p in Eq.(D.3) with p^{MD} and p^{FA}, respectively.

D.2.3 Server Modeling

Similar to the concept of stochastic arrival curve, stochastic service curve is defined to describe the service guarantee that a server can provide, and several different definitions have been proposed. Here, we employ the following one [8].

Definition 2. ***(Stochastic Service Curve).*** *A system S is said to provide a stochastic service curve $\beta \in F$ with bounding function $g \in \bar{F}$, denoted by $S \sim_{ssc} \langle g, \beta \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\{A \otimes \beta(t) - A^*(t) > x\} \leq g(x). \tag{D.6}$$

Here, $A \otimes \beta(t) \equiv \inf_{0 \leq s \leq t} \{A(s) + \beta(t-s)\}$, and $A^*(t)$ denotes the cumulative amount of output traffic up to time t.

The Gilbert-Elliott channel model is named after the originators, which can be further classified into discrete-time and continuous-time model. In this paper, the discrete time model is considered, since it matches well with the slotted system model. Fig.D.2 shows a two-state GE channel, where the channel can either be in *ON* state (state 1), in which data can be decoded error-free (if no collision happens during the transmission), or in state *OFF* (state 0), in which the channel quality is too bad to transmit any data. The channel state transits among the two states as a Markov process with transition matrix of Q, where q_{ij} denotes the transition probability from state i to state j $(i, j \in \{0, 1\})$.

Let $S(t)$ denote the service provided by the channel during $(0, t]$. Then, there are two cases.

- Case 1: t is not within any backlogged period. In this case, there is no backlog in the system at time t, which means that all traffic that arrived up to time t has left the server. Hence, $A^*(t) = A(t)$ and consequently $A \otimes \beta(t) - A^*(t) = A(t) + \beta(0) - A^*(t) = 0$.

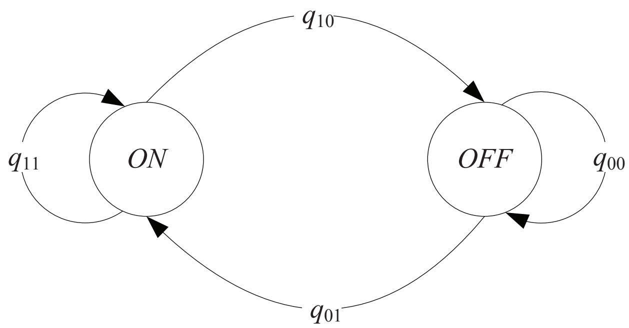


Figure D.2: Discrete-time two-state Gilbert-Elliott channel model

- Case 2: t is within a backlogged period $(t_0, t_b]$, where t_0 is the start point of the backlogged period. Then, $A^*(t_0) = A(t_0)$ and

$$\begin{aligned} & A \otimes \beta(t) - A^*(t) \leq A(t_0) + \beta(t - t_0) - A^*(t) && \text{(D.7)} \\ = \; & \beta(t - t_0) + A^*(t_0) - A^*(t) = \beta(t - t_0) - S(t_0, t) && \text{(D.8)} \end{aligned}$$

Then, we have:

$$\begin{aligned} & P\{A \otimes \beta(t) - A^*(t) > x\} \leq P\{\beta(t - t_0) - S(t_0, t) > x\} \\ \leq \; & P\{e^{\theta[\beta(t-t_0)-S(t_0,t)]} > e^{\theta x}\} \leq e^{-\theta x} E[e^{\theta[\beta(t-t_0)-S(t_0,t)]}] \\ \leq \; & e^{-\theta x} E[e^{\theta[\mu(\theta)\cdot(t-t_0)-S(t_0,t)]}] \\ = \; & e^{-\theta x} E[e^{\theta[\mu(\theta)\cdot\tau-S(\tau)]}] \leq e^{-\theta x} \end{aligned}$$

where the third step is known as Chernoff bound, and the fourth step is due to that $S(t)$ is stationary, and

$$\mu(\theta) \equiv -\frac{1}{\theta\tau} \log E[e^{-\theta S(\tau)}]$$

which is known as the *effective bandwidth* of process S in the literature [9] [10]. For the two-state Markov chain of the considered GE channel, its effective bandwidth has an explicit form [10], which is adopted in this paper as

$$\begin{aligned} & \hat{\mu}(\theta) \\ = \; & \frac{1}{-\theta} \log \left(\frac{q_{00} + q_{11}e^{-c\theta} + \sqrt{(q_{00} + q_{11}e^{-c\theta})^2 - 4(q_{11} + q_{00} - 1)e^{-c\theta}}}{2} \right) \end{aligned}$$

By combining both cases, a stochastic service curve of GE channel has been found as:

$$S(t) \sim_{ssc} \langle g(x) = e^{-\theta x}, \beta(t) = \mu(\theta) \cdot t \rangle. \qquad \text{(D.9)}$$

D.2.4 Delay Bound

Previous work in [2] has discussed how to obtain the stochastic service curve that can be effectively provided to each input traffic. Here, by using the following notations for input traffic,

$$fl^P\colon\ A^P(t) \sim_{sac} \langle f^P(x), \alpha^P(t)\rangle \tag{D.10}$$
$$fl^S\colon\ A^S(t) \sim_{sac} \langle f^S(x), \alpha^S(t)\rangle, \tag{D.11}$$

and by further assuming Re-Transmission until Success (RT-S) scheme, the stochastic service curve for PUs' traffic and SUs' traffic can be expressed as:

$$fl^P\colon \qquad S^P(t) \sim_{ssc} \langle g^P(x), \beta^P(t)\rangle \tag{D.12}$$
$$\text{with} \qquad g^P(x) = f^{MD} \otimes g(x), \beta^P(t) = \beta - \alpha^{MD}(t) \tag{D.13}$$
$$fl^S\colon \qquad S^S(t) \sim_{ssc} \langle g^S(x), \beta^S(t)\rangle \tag{D.14}$$
$$\text{with} \qquad g^S(x) = f^I \otimes g \otimes f^P(x), \beta^S(t) = \beta - \alpha^I - \alpha^P(t) \tag{D.15}$$

Then, based on the performance bound theorem in [2], the delay distribution bound can be summarized as:

Theorem 1. ***(Delay Bound)***

$$P\{D^U(t) > h(\alpha^U + x, \beta^U)\} \leq [f^U \otimes g^U(x)]_1, \tag{D.16}$$

where $U \in \{P, S\}$, $h(\alpha + x, \beta) = sup_{s\geq 0}\{inf\{\tau \geq 0 : \alpha(s) + x \leq \beta(s+\tau)\}\}$ *and* $[\cdot]_1$ *denotes* $max(min(\cdot, 1), 0)$

D.3 Numerical Results

In previous sections, traffic model, server model as well as the considered cognitive radio network model are described with the delay bound theorem as an ending. In this section, specific parameters and configurations will be substituted into the deduction above in order to obtain the capacity limit under certain delay constraints.

The input packet arrival, fl^P and fl^S, are assumed to be Poisson flow. And the stochastic arrival curve for Poisson traffic is defined as follows.

Definition 3. ***(Poisson Traffic).*** *Suppose all packets of a flow have the same size L and they arrive according to a Poisson process with mean*

Table D.1: QoS Requirements for Different Services in LTE System

Traffic	Delay Budget	Packet Loss Prob
VoIP	$50ms$	10^{-2}
TCP	$250ms$	10^{-4}

arrival rate λ. Then the flow has a stochastic arrival curve $A(t) \sim_{sac} \langle f_{Pois}, rt \rangle$ for any $r > \lambda L$ with bounding function [8]:

$$f_{Pois}(x) = 1 - (1-a) \sum_{i=0}^{k} \left[\frac{[a(i-k)]^i}{i!} e^{-a(i-k)} \right]$$

where $a = \frac{\lambda L}{r}$ and $k = \lceil \frac{x}{L} \rceil$.

The network is supposed to be a LTE system using OFDM technology with slot length of $0.5ms$. In each slot, there are 7 OFDM symbols in time domain, 50 resource blocks (RB) in frequency domain with 12 sub-carriers in each RB. $16QAM$ and $1/3-rate$ Turbo code are used as the modulation and coding scheme. Then, the packet length for Poisson arrival is set as the effective bits transmitted in an LTE slot, i.e., $5.6kbits$. Based on this assumption, the parameter σ in error process and c in channel model are all equal to 1 packet per slot. State transition probability q_{01} and q_{10} for GE channel are set as 1 and 0.11, respectively. The free parameters, such as θ, are optimized numerically with a tradeoff between acceptable accuracy and tolerable complexity.

Primary traffic flow is supposed to be a VoIP session, which belongs to the Guaranteed Bit Rate (GBR) bearer in LTE system. While secondary traffic flow is set as TCP interaction service, which is non-GBR bearer because of the lower priority in the whole network. Table.D.1 lists the QoS Class Identifier (QCI) requirements.

In the system model considered here, re-transmission until success mechanism is employed, which means no packet is dropped because of collision or deep channel fading. Packet loss only happens when the sojourn delay exceeds the delay budget. Therefore, the delay constraints can be written as:

$$\textit{Constraint 1}: \quad P(Delay^P > 100slots) \leq 10^{-2} \tag{D.17}$$

$$\textit{Constraint 2}: \quad P(Delay^S > 500slots) \leq 10^{-4} \tag{D.18}$$

In order to fulfill the delay constraints, there exist an *upper bound* on the *arrival rate* λ of input traffic, which is defined as the *capacity limit* in this paper.

The capacity limit of PUs flow can be expressed as:

$$C^P = max\{\lambda^P, \text{subject to } \textit{Constraint 1}\} \tag{D.19}$$

Fig.D.3(a) shows the delay distribution of PUs input flow calculated from Theorem 1. We can notice that, there is still some capacity margin when the arrival rate of PUs traffic is 1600 packets per second; while delay constraint cannot be met when the arrival rate is increased to 1720 packets per second. The maximum arrival rate of primary traffic, also called capacity limit C^P, is 1690 packets per second when the delay constraint can be guaranteed at the same time.

As for the secondary traffic, it can be transmitted when there is no primary traffic. Therefore, the maximum arrival rate of the secondary network has close relationship with the load η of primary network, which is defined as the ratio of actual arrival rate over the capacity limit, i.e., $\eta = \lambda^P/C^P$. Then, the capacity limit of SUs flow can be expressed as:

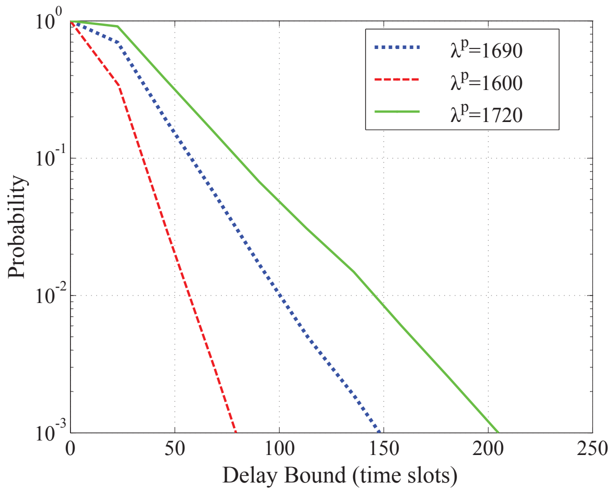
$$C^S = max\{\lambda^S|\eta, \text{subject to } \textit{Constraint 2}\} \tag{D.20}$$

Fig.D.3(b) shows three delay distribution bounds when η is set as 0. We can notice that $\lambda^S = 1738$ packets per second is the capacity limit C^S under delay constraint 2.

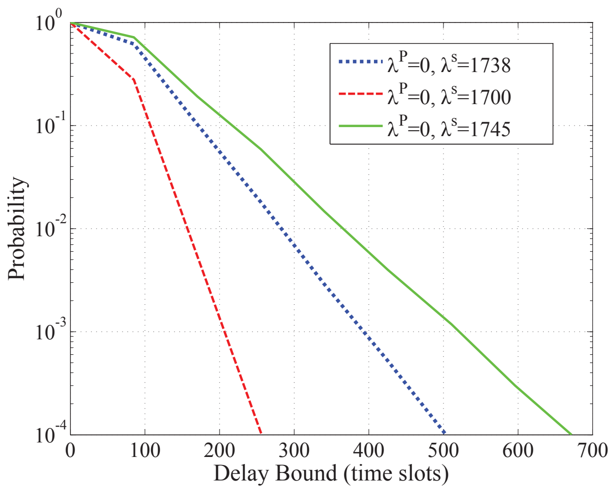
If we define $C^P = 1690$ and $C^S = 1738$ packets per second as 100% load of the primary network and secondary network, respectively, Fig.D.4 provides the admissible capacity region of the system, given Poisson arrivals. It is shown that the maximum arrival rate of secondary flow decreases when the load of primary flow increases. Particularly, for any point below the curve, which corresponds to a load of primary traffic and a load of secondary traffic, the system can guarantee the delay requirement and the required loss probability.

D.4 Conclusion and Discussion

In this paper, capacity limits, defined as the maximum arrival rate, for both primary and secondary traffics in a cognitive radio network are obtained under delay constraints. Stochastic network calculus is relied on to derive the delay distribution bounds, which includes two fundamental concepts: stochastic arrival curve and stochastic service curve. The spectrum sensing error process is analyzed with stochastic arrival curve first. Secondly, Gilbert-Elliott is used to model the fading channel, and its stochastic service curve is derived. And then, specific expressions for delay distribution bounds are obtained. Parameters and configurations in LTE network are



(a) Primary Traffic



(b) Secondary Traffic

Figure D.3: Delay Tail Distribution

used to calculate the numerical results, where the capacity limit of primary traffic and the capacity limit of secondary traffic under different traffic load are discussed.

In this paper, we have only considered Poisson arrival due to space limitation. However, the analysis can be easily extended to other types

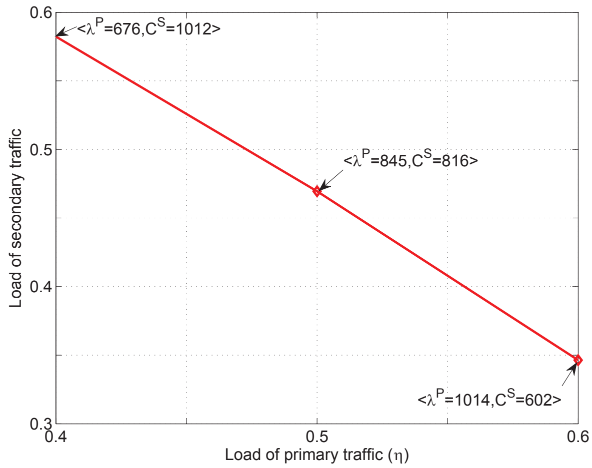


Figure D.4: Capacity Region

of arrivals. Particularly, for many types of traffic, their stochastic arrival curves have been found (e.g. see [11]), with which, the corresponding delay bounds and capacity/throughput regions are readily obtained.

Bibliography

[1] Suliman, I., Lehtomaki, J.: Queueing analysis of opportunistic access in cognitive radios. In: 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management, pp. 153–157. Aalborg (2009)

[2] Gao, Y., Jiang, Y.: Performance Analysis for a Cognitive Radio with Imperfect Spectrum Sensing. In: IEEE INFOCOM 2010 Workshop on Cognitive Wireless Communications and Networking. IEEE Press, San Diego (2010)

[3] Gao, Y., Jiang, Y., Lin, T., Zhang, X.: Performance Bounds for a Cognitive Radio Network with Network Calculus Analysis. In: International Conference on Network Infrastructure and Digital Content. Beijing (2010)

[4] Gao, Y., Jiang, Y., Zhang, X.: On the Backlog and Delay Distributions of Primary and Secondary Users in a Cognitive Radio Network. China Communications, vol.7, no.5, pp 23-28. (2010)

[5] Cruz, R. L.: A calculus for network delay, Part I: Network elements in isolations, and Part II: Network Analysis. In: IEEE Transaction of Information Theory, vol. 37(1), pp: 114–141. IEEE Press (1991)

[6] Gilbert, E. N.: Capacity of a bursty-noise channel. Bell System Technical Journal, vol. 39(5), pp. 1253–1265. (1960)

[7] Elliott, E. O.: Estimates of error rates for codes on bursty-noise channels. Bell System Technical Journal, vol. 42(9), pp. 1977–1997. (1963)

[8] Jiang, Y., Liu, Y.: Stochastic Network Calculus. Springer. (2008)

[9] Kelly, F.: Notes on effective bandwidths. In: Stochastic Networks: Theory and Applications, Royal Statistical Society Lecture Notes Series, 4. Oxford University Press (1996)

[10] Chang, C.-S.: Performance Guarantees in Communication Networks. Springer-Verlag. (2000)

[11] Jiang, Y., Yin, Q., Liu, Y., Jiang, S.: Fundamental Calculus on Generalized Stochastically Bounded Bursty Traffic for Communication Networks. Computer Networks, vol. 53, no. 12, pp. 2011-2021. Elsevier Science. (2009)

Appendix

Lemma 1. ***(v.b.c Stochastic Arrival Curve)*** *If an arrival process $A(t)$ has independent stationary increments, then it has a v.b.c stochastic arrival curve $\alpha(t) = [\rho(\theta) + \theta_1] \cdot t$ with bounding function $f(x) = e^{-\theta\theta_1}e^{-\theta x}$ for $\forall\theta_1 \geq 0$, and for $\forall\theta > 0$ and $\rho(\theta) = \frac{1}{\theta}\log E\left[e^{\theta A(1)}\right]$.*

Proof. Define a sequence of non-negative random variables $\{V_s\}$ as

$$V_s = e^{\theta A(t-s,t)-\theta[\rho(\theta)+\theta_1]\cdot s}. \tag{D.21}$$

Since $A(t)$ has independent stationary increments, we then have,

$$\begin{aligned} V_{s+1} &= e^{\theta A(t-s-1,t)-\theta[\rho(\theta)+\theta_1]\cdot(s+1)} && \text{(D.22)} \\ &= e^{\theta\sum_{k=t-s}^{t} X_k-\theta[\rho(\theta)+\theta_1]\cdot(s+1)} && \text{(D.23)} \\ &= V_s \cdot e^{\theta X_{t-s}-\theta[\rho(\theta)+\theta_1]} && \text{(D.24)} \end{aligned}$$

where $X_k = A(k-1,k)$ is used to simplify the notations. In addition, it is easy to know that X_{t-s} is independent of $X_t, X_{t-1}, ..., X_{t-s+1}$, and it has stationary increments, there holds:

$$\begin{aligned} & E[V_{s+1}|V_1, V_2, ..., V_s] = E[V_{s+1}|X_t, X_{t-1}, ..., X_{t-s+1}] && \text{(D.25)} \\ = \; & E[V_s \cdot e^{\theta X_{t-s}-\theta[\rho(\theta)+\theta_1]}|X_t, X_{t-1}, ..., X_{t-s+1}] && \text{(D.26)} \\ = \; & E[V_s|X_t, X_{t-1}, ..., X_{t-s+1}] \cdot E[e^{\theta X_{t-s}-\theta[\rho(\theta)+\theta_1]}] && \text{(D.27)} \\ = \; & V_s \cdot \frac{E[e^{\theta X_1}]}{\theta\rho(\theta)+\theta\theta_1} \leq V_s && \text{(D.28)} \end{aligned}$$

Hence, $V_1, V_2, ..., V_t$ form a non-negative supermartingale. Then based on an inequality for supermartingale, Doob's inequality, the definition of $\rho(\theta)$ and $A(t)$ has stationary increments, there holds:

$$
\begin{aligned}
& P\left\{\sup_{0\leq s\leq t}\{A(s,t)-[\rho(\theta)+\theta_1]\cdot(t-s)\}>x\right\} && \text{(D.29)}\\
= \;& P\left\{\sup_{0\leq s\leq t}\{e^{A(s,t)-[\rho(\theta)+\theta_1]\cdot(t-s)}\}>e^{x}\right\} && \text{(D.30)}\\
= \;& P\left\{\sup_{1\leq s\leq t}V_s>e^{\theta x}\right\}\leq P\{V_1>e^{\theta x}\} && \text{(D.31)}\\
\leq \;& e^{-\theta x}E[e^{\theta A(t-1,t)-\theta[\rho(\theta)+\theta_1]}]\leq e^{-\theta x}e^{-\theta\theta_1} && \text{(D.32)}
\end{aligned}
$$

which ends the proof. □

Appendix E

Publication E

Yuehong Gao and Yuming Jiang; Analysis on the capacity of a cognitive radio network under delay constraints; IEICE Transactions on Communications; Vol. E95-B, No. 04, 2012.

Abstract

In this paper, performance analysis of a cognitive radio network is conducted. In the network, there is imperfect sensing and the wireless channel is a Gilbert-Elliott channel. The focus is on the network's capacity in serving traffic with delay constraints. Specifically, the maximum traffic arrival rates of both primary users and secondary users, which the network can support with guaranteed delay bounds, are investigated. The analysis is based on stochastic network calculus. A general relationship between delay bounds, traffic patterns and important characteristics such as spectrum sensing errors and channel fading of the cognitive radio network is derived. This relationship lays a foundation for finding the capacity under different traffic scenarios. Two specific traffic types are exemplified, namely periodic traffic and Poisson traffic. Analytical results are presented in comparison with simulation results. The comparison shows a good match between them, validating the analysis.

Keywords *Cognitive radio network, Stochastic network calculus, Capacity, Delay-constrained*

E.1 Introduction

Cognitive radio is a promising wireless communication technology used to increase the spectrum efficiency [1]. In a cognitive radio network, there are two types of users, namely primary users (PUs) and secondary users (SUs). The essential concept is to allow SUs to access available spectrum resource when PUs are silent. Particularly, secondary users have the ability to sense and use available spectrum holes when primary users do not transmit data on the assigned spectrum. As a result, an increased capacity in accommodating communication demands can be expected. A fundamental and challenging question is hence how much data traffic a cognitive radio network can serve, particularly when the traffic has delay requirement. To the best of our knowledge, this is an open question and few results are available in the literature.

The aim of this paper is to investigate the capacity of a cognitive radio network in serving data traffic, where the sensing is imperfect and the wireless channel is a Gilbert-Elliott channel. Particularly, we are interested in finding the *maximum traffic arrival rates* of both primary users and secondary users, which the network can support with guaranteed delay bounds. Here, a guaranteed delay bound is in the probabilistic setting and represented by a delay bound and a violation probability. It reads: the probability that the system delay is larger than the delay bound is bounded by the violation probability. This probabilistic definition of delay bound has been widely adopted (e.g. [2]).

In this paper, a newly developed theory - *stochastic network calculus* - is relied on to obtain performance bounds of primary users and secondary users in the considered cognitive radio network. In stochastic network calculus, stochastic arrival curve and stochastic service curve are defined to model traffic and service, respectively. The contribution of the paper is as follows. First, a stochastic arrival curve for sensing error process and a stochastic service curve for Gilbert-Elliott fading channel are derived. With this, stochastic service curves for primary users and secondary users are obtained. Further on, based on existing stochastic network calculus results, general relationships between delay bounds, data traffic processes, spectrum sensing error processes, and channel fading process is established for both PUs and SUs. For two specific traffic types, namely periodic traffic and Poisson traffic, the corresponding maximum traffic arrival rates, which represent the capacity of the network under delay constraints, are exemplified with numerical results. Comparisons with simulation results are also provided, showing good match with analytical results and hence validating the analysis.

The rest of this paper is organized as follows. In Section E.2, we review related work. Then, the considered cognitive radio network is modeled in Section E.3. Section E.4 introduces the basics of stochastic network calculus. Section E.5 shows the detailed analysis that includes the derivations of stochastic arrival curves for sensing error processes, stochastic service curves for both primary users and secondary users, and delay bounds as functions of the data traffic processes, spectrum sensing error processes, and channel fading process. Section E.6 shows numerical results and simulation results demonstrating and validating the analytical results. Finally, the conclusion is given in the last section.

E.2 Related Work

Cognitive radio network has gained a lot of attention since it was proposed. Significant effort has been put in this area. Among the various research issues, performance analysis and evaluation play an important role.

In the literature, the classic queueing theory has been used to conduct performance analysis of cognitive radio networks. In [3], an $M/D/1$ priority queueing system model is used to derive the average waiting times and average queueing lengths in a cognitive radio network with perfect spectrum sensing. The authors of [4] relied on the $M/G/1$ preemptive priority queue to obtain analytical forms of average delay and throughput for both PUs and SUs. $M/M/1$ queueing model is employed to analyze the average queueing time of secondary users in [5]. In these works, the impact of spectrum sensing errors on the system performance is not well studied. In addition, they only provide results in terms of average values with little investigation on probabilistic delay bounds.

The Markov chain model has also been relied on to conduct performance analysis. Considering the cognitive radio scenario, the state space of the Markov chain can be defined in two ways: (1) based on the channel occupancy state (that is, whether a channel is free, occupied by PU, occupied by SU or collision), and (2) the number of PUs and the number of SUs in the system. In most literatures, the second is used for indices of the Markov state space, such as in [6–8], because, with it, the dimensionality and complexity of the Markov model (especially in the multi-channel case) can be more easily reduced as highlighted by the authors of [7]. The Markov chain based analysis helps to derive the blocking probability for SUs, average number of users in the system as well as throughput. However, to the best of our knowledge, no delay-related results are available from this analysis.

In all the related works, the authors assume Poisson arrival and most of them also assume exponentially distributed service time, so that existing queueing theory results, particularly $M/G/1$ priority queue results, can be directly applied, and the Markov chain model can be established. However, these assumptions are too restrictive for modern wireless communication networks, where the traffic can be of different types and the channel capacity can vary over time.

E.3 The System Model

In this paper, we consider a cognitive radio network as shown in Fig.E.1, where there are two classes of users, primary users (PUs) and secondary users (SUs), sharing a wireless channel. Primary users are authorized users and have priority to use the spectrum over secondary users. Secondary users are equipped with the ability to sense the channel, and transmit data on the channel only when the channel is sensed idle, indicating that primary users do not transmit data at the moment.

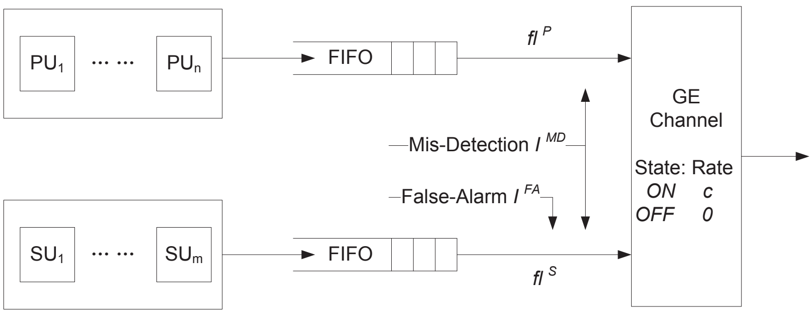


Figure E.1: System Model

We assume Time Division Multiplexing (TDM) for the channel. The network is supposed to be synchronized and the time domain is divided into slots with length T and indexed by $[0, 1, 2, \dots]$. It is assumed that some coordination within the same class of users exists such that their data are queued and served in the first-in-first-out (FIFO) manner, as indicated by two parallel FIFO buffer in Fig.E.1, therefore, no collision will happen between the same class of users[1].

However, potential sensing errors between different classes of users are considered. At the beginning of each slot, secondary users will try to sense the spectrum to decide whether the current slot is idle or busy. For ease of exposition, we assume that the time used for spectrum sensing is small and

[1]These assumptions imply no wasted service due to sensing or collision between the same class of users, and hence allow to explore the maximum capacity of the network in serving data traffic.

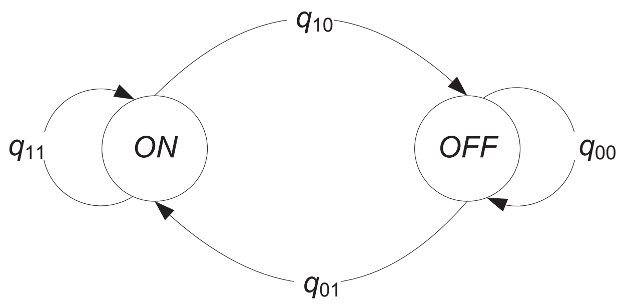


Figure E.2: Discrete-time two-state Gilbert-Elliott channel model

its effect is not counted. When the sensing result is different from the real scenario, a sensing error happens. Typically, spectrum sensing errors can be classified into two types [9], i.e., mis-detection (MD) and false alarm (FA). Mis-detection means that the channel is occupied by PUs but the sensing result says it is available for SUs, which will result in transmission collision between PUs' and SUs' current transmission. However, false alarm occurs in the opposite way, when SUs believe that the channel is being used by PUs but actually it is idle, which means that SUs will miss the transmission slot.

An important feature in wireless systems is channel fading, which can have great impact on the system performance. In this paper, a widely adopted fading channel model, the Gilbert-Eliott (GE) channel, is considered. Fig.E.2 shows the two-state GE channel, where the channel can either be in *ON* state (state 1), in which data can be decoded error-free (if no collision happens during the transmission), or in state *OFF* (state 0), in which the channel quality is too bad to transmit any data. The channel state transits between the two states as a Markov process with transition matrix Q, where q_{ij} denotes the transition probability from state i to state j ($i, j \in \{0, 1\}$), and it is at the steady-state from the start.

Note that packet arrivals, sensing errors and the channel service are generally stochastic processes, which will only lead to stochastic service guarantees. While the system model described above has already been overly simplified, to the best of our knowledge, it is difficult (if not impossible) to obtain explicit results from the classic queueing theory, particularly when the involving stochastic processes are not Poisson or with exponentially distributed rates. Furthermore, when the concern is probabilistic guarantees, such as the capacity under probabilistic delay constraints studied in this paper, the difficulty becomes even more challenging. To address this problem, we resort to stochastic network calculus, and its basics and related

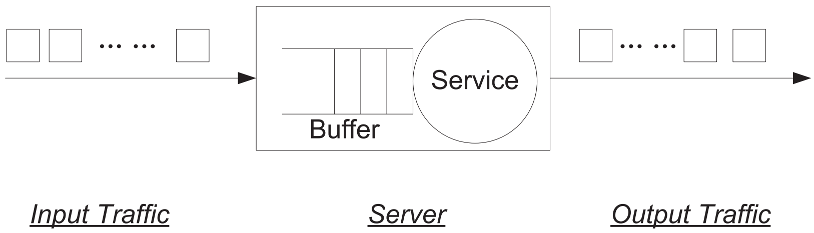


Figure E.3: Basic Elements of a Traffic Serving System

results are introduced in the next section.

E.4 Stochastic Network Calculus Basics

Network calculus is a newly developed queueing theory for service guarantee analysis. Since its introduction in early 1990s [10, 11], network calculus has evolved into two branches: the deterministic branch and the stochastic branch. In our research, stochastic network calculus is employed because stochastic processes are involved in the considered network.

Like the classic queueing theory, network calculus also has its foundation on three basic elements, namely input traffic, server and output traffic, as shown in Fig.E.3. Network calculus provides a set of results, based on its defined models for these basic elements, that establish relationships between these models and various performance parameters such as delay and backlog. In stochastic network calculus, stochastic arrival curve and stochastic service curve are the essential concepts and models in representing the traffic and service respectively. Specifically, the traffic of a flow is modeled by a stochastic arrival curve (sac), while a server is described by a stochastic service curve (ssc), in stochastic network calculus. Given the sac of the input traffic and the ssc of the server or system, probabilistic delay bound and backlog bound can be derived.

While several definition variations of stochastic arrival curve are available, we adopt the following one, which explores the *virtual-backlog-property* of a deterministic arrival curve [2]. This model was originally proposed in [12] and generalized in [13], and is now also known in the literature as *sample path envelope* [14].

Definition 1. ***(Stochastic Arrival Curve).*** *A flow $A(t)$ is said to have a* *virtual-backlog-centric (v.b.c) stochastic arrival curve* $\alpha \in F^2$ *with bounding function* $f \in \bar{F}^3$, *denoted by* $A(t) \sim_{sac} \langle f, \alpha \rangle$, *if for all* $t \geq 0$ *and all* $x \geq 0$ *there holds:*

$$P\left\{\sup_{0 \leq s \leq t} \{A(s,t) - \alpha(t-s)\} > x\right\} \leq f(x), \tag{E.1}$$

where $A(s,t)$ denotes the cumulative amount of traffic of an input flow during period $(s,t]$, $A(t) = A(0,t)$.

For some commonly used traffic models, their stochastic arrival curves have been found, (see, e.g. [13]). Here, the stochastic arrival curves of periodic traffic and Poisson traffic are introduced below, which will be used later in the paper.

Periodic Traffic: A periodic source produces an amount of workload, denoted by δ, at times $\{U\tau + n\tau, n = 0, 1, 2, ...\}$, where τ is the period time length and U is uniformly distributed on the interval $[0,1]$. It has a stochastic arrival curve $A(t) \sim_{sac} \langle f_{Peri}(x), \alpha_{Peri}(t) \rangle$ as:

$$f_{Peri}(x) = 0 \tag{E.2}$$

$$\alpha_{Peri}(t) = \delta + \frac{\delta}{\tau}t, \tag{E.3}$$

where the bounding function $f_{Peri}(x) = 0$ means that the stochastic arrival curve regresses to a deterministic arrival curve, also indicates the traffic generation of a Periodic source is quite smooth.

Poisson Traffic: Suppose all packets of a flow have the same size L and they arrive according to a Poisson process with mean arrival rate λ. Then the flow has a stochastic arrival curve $A(t) \sim_{sac} \langle f_{Pois}, \alpha_{Pois} \rangle$ for any $r > \lambda L$ with bounding function [2]:

$$f_{Pois}(x) = 1 - (1-a)\sum_{i=0}^{k}\left[\frac{[a(i-k)]^i}{i!}e^{-a(i-k)}\right] \tag{E.4}$$

$$\alpha_{Pois}(t) = rt, \tag{E.5}$$

where $a = \frac{\lambda L}{r}$ and $k = \lceil \frac{x}{L} \rceil$.

Similar to the concept of stochastic arrival curve, stochastic service curve is defined to describe the service guarantee that a server can provide. While several definition variations are available [2] [14], we adopt the following one in this paper.

[2] F: the set of non-negative wide-sensing increasing functions

[3] $\bar{F}$: the set of non-negative wide-sensing decreasing functions

Definition 2. ***(Stochastic Service Curve).*** *A system S is said to provide a stochastic service curve $\beta \in F$ with bounding function $g \in \bar{F}$, denoted by $S \sim_{ssc} \langle g, \beta \rangle$, if for all $t \geq 0$ and all $x \geq 0$ there holds:*

$$P\{A \otimes \beta(t) - A^*(t) > x\} \leq g(x). \tag{E.6}$$

Here, $A \otimes \beta(t) \equiv \inf_{0 \leq s \leq t} \{A(s) + \beta(t-s)\}$, $A^*(t)$ denotes the cumulative amount of output traffic up to time t.

When the input traffic has been described by a stochastic arrival curve, and the service provided by a server is characterized by a stochastic service curve, various performance bounds have derived in the stochastic network calculus literature (see e.g. [2]). In this paper, we are particularly interested in the delay bound summarized in the following theorem, whose proof can be found e.g. from [2] and is omitted for conciseness.

Theorem 1. ***(Delay Bound).*** *Consider a server S with input A. Suppose that the input has a v.b.c stochastic arrival curve denoted as $A \sim_{sac} \langle f, \alpha \rangle$, and the server provides to the input a stochastic service curve written as $S \sim_{ssc} \langle g, \beta \rangle$, then for any $t \geq 0$ and $x \geq 0$, the delay of traffic arriving at t, denoted as $D(t) = \inf\{\tau \geq 0 : A(t) \leq A^*(t+\tau)\}$, is bounded by*

$$P\{D(t) \geq h(\alpha + x, \beta)\} \leq f \otimes g(x), \tag{E.7}$$

where $h(\alpha + x, \beta)$ is the maximum horizontal distance between functions $\alpha(t) + x$ and $\beta(t)$ and can be expressed as

$$h(\alpha + x, \beta) = \sup_{t \geq 0} \{\inf\{\tau \geq 0 : \alpha(t) + x \leq \beta(t+\tau)\}\}.$$

E.5 The Analysis

In this section, we provide an analysis on the capacity of the considered cognitive radio network. The key idea is to make use of Theorem 1 to establish a relationship between the required probabilistic delay guarantee, the traffic process, the sensing error processes, and the channel fading process. With this relationship, a capacity bound is readily obtained.

As highlighted in the previous section, in order to apply Theorem 1, it is essential to find stochastic service curves for both PUs and SUs which take into consideration sensing errors and two-state GE channel. To achieve this, we present in the following an analytical approach. First, we study the stochastic service curve characterization of the GE channel. Then, we characterize the sensing error processes using the stochastic arrival curve concept. Third, we investigate the impact of the sensing error processes on the service provided to PUs and that to SUs, and derive stochastic service curves for both PUs and SUs, with which Theorem 1 is applied to establish the desired relationship. Finally, the capacity of the network in serving traffic under delay constraints is studied, where two specific traffic types, namely periodic traffic and Poisson traffic, are exemplified in defining the capacity.

E.5.1 Stochastic Service Curve of the GE Channel

Considering the GE channel described in Section E.3, the following lemma summarizes its stochastic service curve.

Lemma 1. *For a two-state GE channel with state ON (state 1) and OFF (state 0), let q_{ij} denote the transition probability from state i to state j ($i, j \in \{0, 1\}$), and c denote the transmission rate when the channel is ON. Then, the service provided by the channel during $(0, t]$, denoted by $S(0, t)$ ($S(t)$ for short), has a stochastic service curve $\hat{\beta}(t)$ with bounding function $\hat{g}(x)$, denoted by $S(t) \sim \langle \hat{g}(x), \hat{\beta}(t)\rangle$, where*

$$\hat{g}(x) = e^{-\theta x}, \hat{\beta}(t) = \hat{\mu}(\theta) \cdot t \tag{E.8}$$

for $\forall \theta > 0$ and

$$\begin{aligned} \hat{\mu}(\theta) &\equiv -\frac{1}{\theta\tau} \log E[e^{-\theta S(\tau)}] \\ &= -\frac{1}{\theta} \log \Bigg(\frac{q_{00}}{2} + \frac{q_{11}}{2} e^{-c\theta} + \\ &\frac{\sqrt{(q_{00} + q_{11}e^{-c\theta})^2 - 4(q_{11} + q_{00} - 1)e^{-c\theta}}}{2} \Bigg) \end{aligned} \tag{E.9}$$

Proof. To prove, let us consider any time $t \geq 0$. There are two cases.

Case 1: t is not within any backlogged period. In this case, there is no backlog in the system at time t, which means that all traffic that arrived up to time t has left the server. Hence, $A^*(t) = A(t)$ and consequently $A \otimes \hat{\beta}(t) - A^*(t) = A(t) + \hat{\beta}(0) - A^*(t) = 0$.

Case 2: t is within a backlogged period $(t_0, t_b]$, where t_0 is the start point of the backlogged period. Then, $A^*(t_0) = A(t_0)$ and

$$\begin{aligned} & A \otimes \hat{\beta}(t) - A^*(t) \leq A(t_0) + \hat{\beta}(t - t_0) - A^*(t) \\ = \quad & \hat{\beta}(t - t_0) + A^*(t_0) - A^*(t) = \hat{\beta}(t - t_0) - S(t_0, t), \end{aligned}$$

where $S(t_0, t) = A^*(t_0) - A^*(t)$ represents the amount of service provided by the channel during the backlogged period $(t_0, t]$. Then, for the service process $S(s, t)$ with stationary increments, by using Chernoff bound and by the definition of $\hat{\beta}(t)$, we have:

$$\begin{aligned} & P\{A \otimes \hat{\beta}(t) - A^*(t) > x\} \leq P\{\hat{\beta}(t - t_0) - S(t_0, t) > x\} \\ = \quad & P\{e^{\theta[\hat{\beta}(t-t_0)-S(t_0,t)]} > e^{\theta x}\} \\ \leq \quad & e^{-\theta x} E[e^{\theta[\hat{\beta}(t-t_0)-S(t_0,t)]}] \\ = \quad & e^{-\theta x} E[e^{\theta[\hat{\mu}(\theta)\cdot(t-t_0)-S(t_0,t)]}] \\ = \quad & e^{-\theta x} E[e^{\theta[\hat{\mu}(\theta)\cdot\tau-S(\tau)]}] = e^{-\theta x}, \end{aligned}$$

where $\hat{\mu}(\theta) = -\frac{1}{\theta\tau} \log E[e^{-\theta S(\tau)}]$ is known as the *effective bandwidth* of process S in the literature [15] [16]. For the two-state Markov chain of the considered GE channel, its effective bandwidth has an explicit form [16] as expressed by (E.9) above.

Combining both cases, the stochastic service curve of the GE channel is obtained. □

E.5.2 Modeling of Sensing Error Processes

We assume independent sensing in each slot. For each slot, the sensing gives three possible outcomes, namely mis-detection (MD), false-alarm (FA) and correct sensing (CS). Let $C(t)$ and $W(t)$ denote the real channel occupancy state and the sensed occupancy state in slot t, respectively, and 1 means the channel is occupied, 0 not occupied. Then, the probabilities in each slot of the three possible sensing outcomes are defined as:

$$\begin{aligned} p^{MD} &= P\{W(t)=0|C(t)=1\}P\{C(t)=1\} \\ p^{FA} &= P\{W(t)=1|C(t)=0\}P\{C(t)=0\} \\ p^{CS} &= P\{W(t)=1|C(t)=1\}P\{C(t)=1\} \\ &+ P\{W(t)=0|C(t)=0\}P\{C(t)=0\}. \end{aligned}$$

It is worth highlighting that MD and FA have different effects on PUs and SUs. Specifically, while MD affects both PUs and SUs, FA does not affect PUs at all. Note that if mis-detection happens in a slot, a collision between PUs' transmission and SUs' transmission will happen. Throughout this paper, we assume that after collision, retransmission will take place until the transmission is successful[4]. In addition, p^{MD} and p^{FA} are also correlated with the channel fading state. But in the following deductions of this subsection, the channel is supposed to be always in state *ON*, which leads to an upper bound of sensing error process.

From the PUs' perspective, the sensing only has two states: MD or not, and the corresponding probabilities are p^{MD} and $\bar{p}^{MD} = 1 - p^{MD}$. Due to this, for PUs, we are only interested in the MD process that is defined to be $MD(t) \equiv \sum_{s=0}^{t-1} MD(s, s+1)$ where $MD(s, s+1)$ denotes the amount of wasted service due to mis-detection in slot $s+1$, which is denoted as σ in the following. Note that, the mis-detection event in a slot only has two states: state *HAPPEN* with probability p^{MD} and σ amount of wasted service, and state *NOT-HAPPEN* with probability $1 - p^{MD}$, where there is no wasted service. Therefore, the moment generating function (MGF) $E[e^{\theta MD(1)}]$ can be obtained as $p^{MD}e^{\theta\sigma} + 1 - p^{MD}$.

For SUs, both MD and FA take effect, so we combine both MD and FA at each slot and simply use sensing error (SE) to denote the combined event. It is clear that in each slot, SE is also a two-state event as MD and the probability that SE happens is simply $p^{SE} = p^{MD} + p^{FA}$. For SUs, we are interested in this combined sensing error process and denote it by $SE(t) \equiv \sum_{s=0}^{t-1} SE(s, s+1)$ where $SE(s, s+1)$ denotes the wasted amount of service due to sensing error in slot $s+1$. Similarly, the MGF $E[e^{\theta SE(1)}]$ has the same form as that of $MD(1)$, where probability parameter p^{SE} is used.

[4]Various policies may be employed when collisions happen. In [17], we discussed three retransmission policies and studied their impacts on the services provided to both PUs and SUs. Results in [17] can be incorporated in the analysis in this paper, but for ease of exposition and without loss of the principle of the introduced analytical approach, we assume retransmission until success in this paper.

The following lemma summarizes the stochastic arrival curves of MD process and SE process.

Lemma 2. *Consider an interference process* $I(t) \equiv \sum_{s=0}^{t-1} I(s, s+1)$*,* $I \in \{MD, SE\}$*, where* $I(s, s+1)$ *has two states in each slot:* HAPPEN *with probability* p^I *and* σ *amount of wasted service, and* NOT HAPPEN *with probability* $1 - p^I$ *and no wasted service. In addition,* $I(s, s+1)$ *is independent of* $I(k, k+1)$ $(\forall s \neq k)$*. Then, the process* $I(t)$ *has a stochastic arrival curve* $\alpha^I(t)$ *with bounding function* $f^I(x)$*, denoted as* $I(t) \sim_{sac} \langle f^I(x), \alpha^I(t)\rangle$*, where*

$$f^I(x) = e^{-\theta x} \tag{E.10}$$
$$\alpha^I(t) = \frac{1}{\theta}\log(1 - p^I + p^I e^{\theta\sigma}) \cdot t, \tag{E.11}$$

for any $\theta > 0$*.*

Proof. To prove, let us define a sequence of non-negative random variables $\{V_s\}$ as

$$V_s = e^{\theta I(t-s,t) - \theta\rho(\theta)\cdot s},$$

where $\rho(\theta) = \frac{1}{\theta}\log(1 - p^I + p^I e^{\theta\sigma})$.

Since $I(t)$ is the cumulation of independent identically distributed random variables and has independent stationary increments, we then have,

$$\begin{aligned} V_{s+1} &= e^{\theta I(t-s-1,t) - \theta\rho(\theta)\cdot(s+1)} \\ &= e^{\theta\sum_{k=t-s}^{t} X_k - \theta\rho(\theta)\cdot(s+1)} \\ &= V_s \cdot e^{\theta X_{t-s} - \theta\rho(\theta)}, \end{aligned}$$

where $X_k = I(k-1, k)$ is used to simplify the notations. In addition, it is easy to know that X_{t-s} is independent of $X_t, X_{t-1}, ..., X_{t-s+1}$, because we have assumed that the channel is always in ON state and so the correlation of ON and OFF is ignored. There holds:

$$\begin{aligned} & E[V_{s+1}|V_1, V_2, ..., V_s] = E[V_{s+1}|X_t, X_{t-1}, ..., X_{t-s+1}] \\ = \; & E[V_s \cdot e^{\theta X_{t-s} - \theta \cdot \rho(\theta)}|X_t, X_{t-1}, ..., X_{t-s+1}] \\ = \; & E[V_s|X_t, X_{t-1}, ..., X_{t-s+1}] \cdot E[e^{\theta X_{t-s} - \theta \cdot \rho(\theta)}] \\ = \; & V_s \cdot \frac{E[e^{\theta X_1}]}{e^{\theta \cdot \rho(\theta)}} \le V_s, \end{aligned}$$

where the last step is due to that $X_1 = I(0,1)$ and

$$E[e^{\theta I(0,1)}] = 1 - p^I + p^I e^{\theta\sigma} = e^{\theta\rho(\theta)}.$$

Hence, $V_1, V_2, ..., V_t$ form a non-negative supermartingale. Then based on Doob's inequality for supermartingales [20], the definition of $\rho(\theta)$, and that $I(t)$ has independent stationary increments, we obtain:

$$\begin{aligned} & P\left\{\sup_{0 \le s \le t} \{I(s,t) - \rho(\theta) \cdot (t-s)\} > x\right\} \\ = \; & P\left\{\sup_{0 \le s \le t} \{e^{I(s,t) - \rho(\theta) \cdot (t-s)}\} > e^x\right\} \\ = \; & P\left\{\sup_{1 \le s \le t} V_s > e^{\theta x}\right\} \le e^{-\theta x} E[e^{\theta I(t-1,t) - \theta\rho(\theta)}] \\ = \; & e^{-\theta x} E[e^{\theta I(0,1) - \theta\rho(\theta)}] \\ = \; & e^{-\theta x}. \end{aligned}$$

This ends the proof. □

E.5.3 Stochastic Service Curves to PUs and SUs

Having obtained the stochastic service curve of the channel and the stochastic arrival curves of the sensing error processes, we are now ready to present stochastic service curves for the services provided to PUs and SUs.

Let us denote by $S^{PU}(t)$ and $S^{SU}(t)$ the service processes that the network successfully provide to PUs and SUs respectively. In addition, throughout the rest of this paper, we suppose the traffic of PUs, denoted by $A^{PU}(t)$, has a stochastic arrival curve and so does the traffic of SUs, denoted by $A^{SU}(t)$. Specifically, there are:

$$\begin{aligned} A^{PU}(t) &\sim_{sac} \langle f^{PU}(x), \alpha^{PU}(t)\rangle && \text{(E.12)} \\ A^{SU}(t) &\sim_{sac} \langle f^{SU}(x), \alpha^{SU}(t)\rangle. && \text{(E.13)} \end{aligned}$$

Then, we have the following result:

Theorem 2. *For the considered cognitive radio network, the channel is modeled as a GE channel with stochastic service curve of $\langle \hat{g}(x), \hat{\beta}(t)\rangle$ given in Lemma 1, while mis-detection process and sensing error process are characterized by stochastic arrival curves $\langle f^{MD}(x), \alpha^{MD}(t)\rangle$ and $\langle f^{SE}(x), \alpha^{SE}(t)\rangle$ in Lemma 2, then*
(i) The service successfully provided to PUs has a stochastic service curve $\beta^{PU}(t)$ with bounding function $g^{PU}(x)$, denoted by

$$S^{PU}(t) \sim_{ssc} \langle g^{PU}(x), \beta^{PU}(t)\rangle,$$

where

$$\begin{aligned} \beta^{PU}(t) &= \hat{\beta}(t) - \alpha^{MD}(t) && \text{(E.14)} \\ g^{PU}(x) &= \hat{g} \otimes f^{MD}(x). && \text{(E.15)} \end{aligned}$$

(ii) The service successfully provided to SUs has a stochastic service curve $\beta^{SU}(t)$ with bounding function $g^{SU}(x)$, denoted by

$$S^{SU}(t) \sim_{ssc} \langle g^{SU}(x), \beta^{SU}(t)\rangle,$$

where

$$\begin{aligned} \beta^{PU}(t) &= \hat{\beta}(t) - \alpha^{SE}(t) - \alpha^{PU}(t) && \text{(E.16)} \\ g^{SU}(x) &= \hat{g} \otimes f^{SE} \otimes f^{PU}(x). && \text{(E.17)} \end{aligned}$$

Proof. We first prove Part (i) for PUs. Let $R(t)$ denote the sum of the amount of traffic from PUs $A(t)$ and the amount of service of the time slots up to time t where mis-detection sensing errors happened, which is denoted by $I(t)$. The channel is characterized by its stochastic service curve $\langle g(x), \beta(t)\rangle$. From the channel viewpoint, $R(t)$ is its input that consumes service from the channel. Correspondingly, we denote by $R^*(t)$ the sum of the amount of successfully transmitted traffic from PUs, denoted by $A^*(t)$, and the amount of actually collided service due to mis-detection, denoted by $I^*(t)$. Again from the channel viewpoint, $R^*(t)$ is its output that has received service from the channel.

$$R(t) = A(t) + I(t), R^*(t) = A^*(t) + I^*(t). \tag{E.18}$$

In real scenarios, the output will not exceed the input. Therefore, the following inequalities hold for any time $t \geq 0$:

$$R(t) \geq R^*(t), A(t) \geq A^*(t), I(t) \geq I^*(t). \tag{E.19}$$

Let us now consider any time $s \geq 0$. We have:

$$\begin{aligned}
& A(s) \otimes \big(\beta(s) - \alpha^I(s)\big) - A^*(s) \\
= \; & (R(s) - I(s)) \otimes \big(\beta(s) - \alpha^I(s)\big) \\
& - (R^*(s) - I^*(s)) \\
= \; & \inf_{0 \leq u \leq s} \big[R(u) + \beta(s-u) - I(u) - \alpha^I(s-u)\big] \\
& - (R^*(s) - I^*(s)) \\
\leq \; & \inf_{0 \leq u \leq s} [R(u) + \beta(s-u)] - R^*(s) + I(s) \\
& - \inf_{0 \leq u \leq s} \big[I(u) + \alpha^I(s-u)\big] \\
= \; & \inf_{0 \leq u \leq s} [R(u) + \beta(s-u)] - R^*(s) \\
& + \sup_{0 \leq u \leq s} \big[I(s,u) - \alpha^I(s-u)\big].
\end{aligned}$$

Then, the distribution probability can be derived as:

$$\begin{aligned}
& P\left\{A(s) \otimes \left(\beta(s) - \alpha^I(s)\right) - A^*(s) > x\right\} \\
\leq\ & P\left\{R \otimes \beta(s) - R^*(s) > y\right\} \\
& + P\left\{\sup_{0 \leq u \leq s}\left[I(s,u) - \alpha^I(s-u)\right] > x - y\right\} \\
\leq\ & g(y) + f^I(x-y),
\end{aligned}$$

which holds for $\forall y, (0 \leq y \leq x)$. Hence, we have

$$\begin{aligned}
& P\left\{A(s) \otimes \left(\beta(s) - \alpha^I(s)\right) - A^*(s) > x\right\} \\
\leq\ & \inf_{0 \leq y \leq x}\left\{g(y) + f^I(x-y)\right\} \\
=\ & g \otimes f^I(x),
\end{aligned}$$

which ends the proof. □

Note that in the above proof, we have intentionally dropped out the superscript PU in $A(t)$, $A^*(t)$, MD in $I(t)$ and $I^*(t)$. This is because the above proof steps can also be followed to prove Part (ii). Particularly, by treating $A(t)$ and $A^*(t)$ as from SUs and process $I(t)$ as the integrated process of the arrival process of PUs and the combined sensing error process, the second part is similarly obtained and the details are omitted.

E.5.4 Delay-Constrained Capacity

Suppose the required probabilistic delay constraints for PUs and SUs are respectively represented as:

$$\text{Constraint 1 (C1): } P\{D^{PU}(t) > d^1 \leq \epsilon^1 \tag{E.20}$$

$$\text{Constraint 2 (C2): } P\{D^{SU}(t) > d^2 \leq \epsilon^2. \tag{E.21}$$

Denote by R^{PU} and R^{SU} the average traffic rates of PUs and SUs respectively, which are:

$$R^{PU} = \lim_{t\to\infty} \frac{A^{PU}(t)}{t} \tag{E.22}$$

$$R^{SU} = \lim_{t\to\infty} \frac{A^{SU}(t)}{t}. \tag{E.23}$$

Definition 3. *The capacity of the cognitive radio network is the maximum R^{PU} and R^{SU} that the network can support subject to constraints (C1) and (C2).*

It is worth highlighting that finding the exact delay-constrained capacity under bursty traffic condition is not easy in general, which indeed has been identified as a critical challenge in the field of network information theory [18].

With the analysis so far above, we are able to present an analytical limit on the capacity. Particularly, having derived the stochastic service curves provided by the network to both PUs and SUs, the probabilistic delay guarantees of both PUs and SUs immediately follows from Theorem 1, which are summarized here:

$$P\left\{D^U \geq h(\alpha^U + x, \beta^U)\right\} \leq f^U \otimes g^U(x), \tag{E.24}$$

where $U \in \{PU, SU\}$ and α^U, β^U, f^U and g^U are found from (E.12)–(E.17).

Finally, we have the following result for the capacity of the considered cognitive radio network under delay constraints.

Theorem 3. *The considered cognitive radio network has a guaranteed capacity within which the delay constraints (C1) and (C2) are met, which is decided by rates Λ^{PU} and Λ^{SU} with*

$$\Lambda^{PU} = \max\left\{r^{PU} : \begin{array}{c} h(\alpha^{PU} + x, \beta^{PU}) \leq d^1 \\ f^{PU} \otimes g^{PU}(x) \leq \epsilon^1 \end{array}\right\} \tag{E.25}$$

$$\Lambda^{SU} = \max\left\{r^{SU} : \begin{array}{c} h(\alpha^{SU} + x, \beta^{SU}) \leq d^2 \\ f^{SU} \otimes g^{SU}(x) \leq \epsilon^2 \end{array}\right\}, \tag{E.26}$$

where $r^U \equiv \lim_{t\to\infty} \frac{\alpha^U(t)}{t}$, $u \in \{PU, SU\}$.

E.6 Numerical and Simulation Results

In this section, specific parameters and configurations are substituted into the analysis above to obtain numerical results. In addition, simulations were conducted to find the capacity of the cognitive radio network. Simulation results are obtained and compared with numerical results. The focused results are in terms of capacity under various settings.

E.6.1 System Configurations

The network is supposed to work in the Long Term Evolution (LTE) mode using OFDM technology with slot length of $0.5ms$. In each slot, there are 7 OFDM symbols in time domain, 50 resource blocks (RB) in frequency domain with 12 sub-carriers in each RB. $16QAM$ and $1/3-rate$ Turbo code are used as the modulation and coding scheme. Based on these configurations, at most $5.6kbits$ can be transmitted in one slot. Here, it is supposed that all input packets have fixed length of $5.6kbits$. State transition probability q_{01} and q_{10} of GE channel are set as 1 and 0.11, respectively. In Section E.6.3 and E.6.4, mis-detection probability p^{MD} and false-alarm probability p^{FA} are set to 0.5%, which leads to an overall sensing error probability p^{SE} to be 1%. The free parameters, such as θ, are optimized numerically with a tradeoff between acceptable accuracy and tolerable complexity.

In LTE standard, pre-defined QoS classes are specified into two categories: Guaranteed Bit Rate (GBR) and Non-Guaranteed Bit Rate (Non-GBR). Each class is assigned with several parameters, such as delay budget and packet loss probability [19]. Table E.1 lists four examples of the predefined QoS classes.

Table E.1: QoS Requirements of Different Services in LTE System

Resource Type	Packet Delay Budget	Packet Loss Rate
GBR	$< 50ms$	High (e.g. 10^{-1})
GBR	$50ms$	Medium (e.g. 10^{-2})
Non-GBR	Low ($\sim 50ms$)	10^{-3}
Non-GBR	Medium ($\sim 250ms$)	10^{-4}

In this research, primary users are supposed to use GBR, while secondary users use Non-GBR. It is assumed that physical layer retransmission is made until a packet is received successfully, and buffer size is assumed to be large enough so that no packet will be dropped because of overflow. Therefore, packet is lost only when the delay exceeds the required delay budget. Particularly, the corresponding delay constraints are as:

$$\text{Constraint 1 (C1): } P\{D^{PU} > 50ms\} \leq 10^{-2} \tag{E.27}$$
$$\text{Constraint 2 (C2): } P\{D^{SU} > 250ms\} \leq 10^{-4}. \tag{E.28}$$

Two specific traffic models are used, which are periodic traffic and Poisson traffic. Their stochastic arrival curves have been exemplified in Section

E.4. For these two types of traffic, we have

$$\{R^U, r^U\} \sim \frac{1}{\tau^U}$$

for periodic traffic, where τ^U denotes the periodic time length, $U \in \{PU, SU\}$;

$$\{R^U, r^U\} \sim \lambda^U$$

for Poisson traffic, where λ^U denotes the average packet arrival rate of the Poisson process, $U \in \{PU, SU\}$.

E.6.2 PUs' Capacity Limits

An important design goal of cognitive radio networks is that secondary users should not introduce interference to primary users. By setting $p^{MD} = 0$, the capacity limits in perfect sensing case can be obtained as shown in Table.E.2.

Table E.2: Capacity Limits without Mis-detection Errors

Traffic Model	Theoretical Results (packets per second)	Simulation Results (packets per second)
Periodic	1786	1815
Poisson	1715	1765

In practical scenarios, however, mis-detection is not avoidable, which will degrade PUs' service guarantee and decrease the capacity limit. Fig.E.4 presents the relationship between PUs' capacity limit and mis-detection probability p^{MD} for both Periodic traffic and Poisson traffic, where the y-axis is normalized by the corresponding capacity listed in Table.E.2 when there is no mis-detection error. It is shown that the difference between theoretical results and simulation results is not significant, which validates the tightness of delay distribution bound.

It is intuitive that higher mis-detection probability leads to smaller capacity limit. For example, by considering the theoretical results, the capacity is reduced by about 12% for Periodic traffic and 16% for Poisson traffic, when mis-detection probability in each slot is 10%. Furthermore, an upper bound of mis-detection probability can be found conversely based on the maximum capacity loss of PUs' traffic. In Fig.E.4, 10% capacity deterioration corresponds to a maximum mis-detection probability of $p^{MD} = 8.6\%$ for Periodic traffic and $p^{MD} = 6\%$ for Poisson traffic, which can be used as an upper bound on p^{MD} when designing spectrum sensing algorithms. Simulation results are also plotted in Fig.E.4. It is seen that the theoretical analysis gives tight bound compared with simulation results for Periodic traffic. For Poisson traffic, while the theoretical results still match with simulation results, they are more conservative than for Periodic traffic.

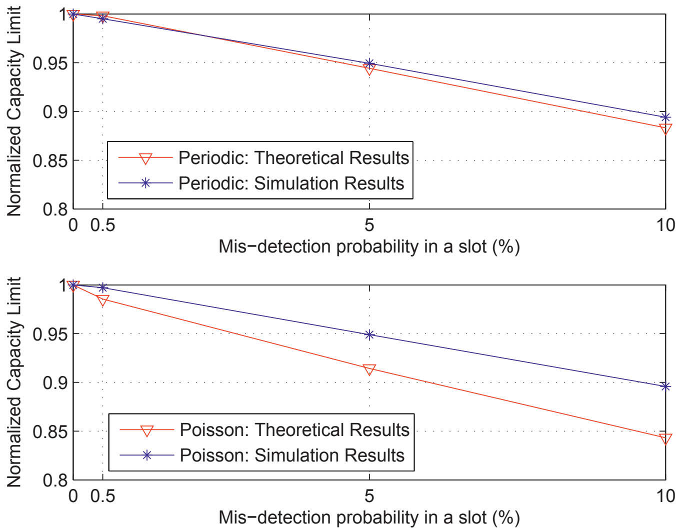


Figure E.4: Impact of Mis-detections on PUs' Capacity Limit

E.6.3 SUs' Capacity Limits

Some SUs' traffic can be accommodated by the network only when there is still some capacity margins. Therefore, SUs' capacity limits depend on PUs' actual input load η. When there is no primary user, then the secondary users can make use of the whole radio resource, and the maximum capacity is obtained and used for normalization purpose in plotting the results.

By increasing the traffic load from PUs, the supportable capacity limit of SUs traffic decreases as shown in Fig.E.5 for Periodic traffic and in Fig.E.6 for Poisson traffic, where the capacity limits are normalized with respect to the case that there is no traffic from PUs and no mis-detection. The area below the curve forms the admissible capacity region of the network. Particularly, for any point below the curve, which corresponds to a load of primary traffic and a load of secondary traffic, the system can guarantee the delay requirement and the required loss probability. In addition, both theoretical results and simulation results are plotted. It can be noticed that the theoretical results locate closely to the simulation results for Poisson traffic, and they are almost the same with the simulation results for Periodic traffic, which indicates the tightness of the capacity limits obtained through the delay analysis of stochastic network calculus.

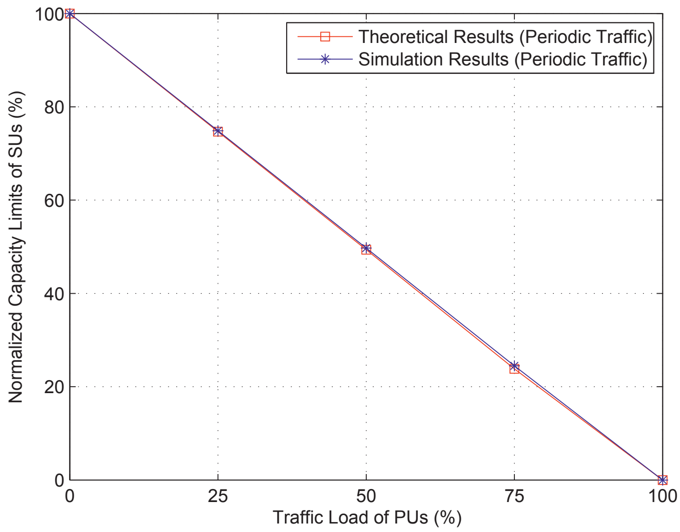


Figure E.5: Normalized Capacity of SUs (Periodic Traffic)

E.6.4 Slow Fading Channel

In the previous results, the state transition probability q_{01} and q_{10} of GE channel are set to 1 and 0.11, respectively, which indicates fast fading speed. In this part, the capacity limits under slow fading are studied. Particularly, we now set $q_{01} = 0.1$ and $q_{10} = 0.011$.

First, the capacity limits of PUs' traffic by theoretical analysis are summarized and compared in Table.E.3, where mis-detection probability is set to zero. It is found that capacity limit decreases when fading speed is slow. This is because the GE channel is likely to stay in state *OFF* for a longer time in slow fading, which may lead to a large delay for packets arrived during this period, weakening the ability in meeting the delay constraints.

The capacity limits of SUs traffic with Periodic model are plotted in Fig.E.7. It is seen that the capacity limits are reduced when the fading is slow. In addition, the gap between theoretical results and simulation

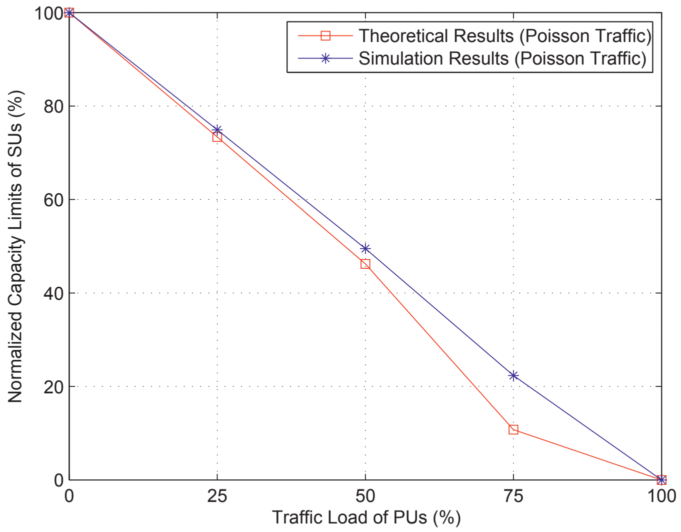


Figure E.6: Normalized Capacity of SUs (Poisson Traffic)

Table E.3: Capacity Limits of PUs under Fast and Slow Fading Channel

Traffic Model	Fading Speed	Capacity Limit (packets per second)	Capacity Loss (%)
Periodic	Fast	1786	6.55
	Slow	1669	
Poisson	Fast	1715	22.6
	Slow	1328	

results increases under slow fading, especially when the primary network is fed with heavy load. Fig.E.8 shows the results for Poisson traffic, which indicate the same trend. In all cases, the analytical capacity results match well with the simulation results.

E.7 Conclusion and Discussion

In this paper, capacity analysis of a cognitive radio network under delay constraints is conducted. Stochastic network calculus is applied to obtain

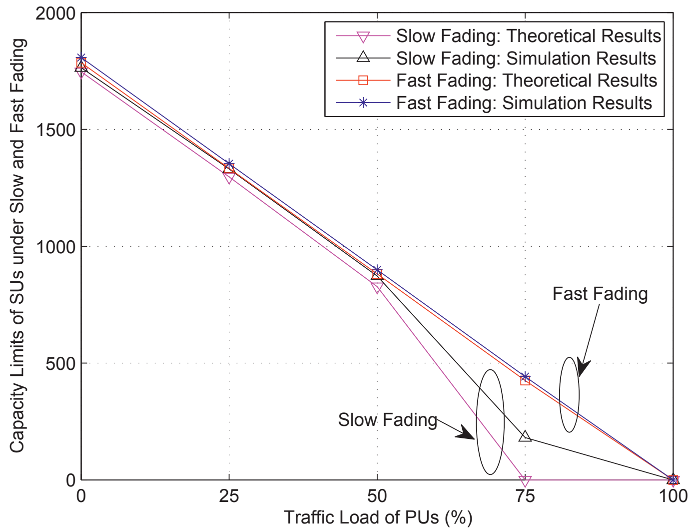


Figure E.7: Capacity Limits of SUs under Slow and Fast Fading (Periodic Traffic)

probabilistic delay bounds for both PUs and SUs, based on which analytical capacity limits are obtained. A crucial part is to find stochastic service curves for both PUs and SUs which take into consideration sensing errors and two-state GE channel. The analysis builds on several steps. First, we study the stochastic service curve characterization of the GE channel. Then, we characterize the sensing error processes using the stochastic arrival curve concept. Third, we investigate the impact of the sensing error processes on the service provided to PUs and that to SUs, based on which stochastic service curves for both PUs and SUs are then derived. Finally, relationships between delay, traffic processes, sensing error processes and two-state GE channel process are established for both PUs and SUs, with which, analytical results on the capacity are presented. Both numerical and simulation results are presented and discussed by considering an LTE parameter setting. Two specific traffic types, namely periodic traffic and Poisson traffic, are used to exemplify the results. The comparison between analytical and simulation results shows a good match between them, indicating the effectiveness of using stochastic network calculus analysis to find capacity limits. These results are ready to be extended to other channel models, such as Finite State Markov Channel, by substituting the stochastic service curve of the concerned channel into $\langle \hat{g}, \hat{\beta} \rangle$. In addition, the results obtained by

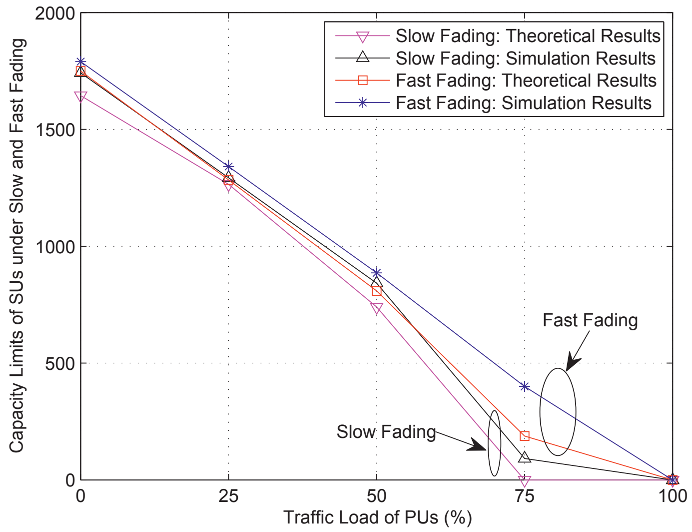


Figure E.8: Capacity Limits of SUs under Slow and Fast Fading (Poisson Traffic)

stochastic network calculus can be further improved by using the concept of stochastic strict server and impairment process, which will form part of our ongoing work.

We stress that in order to obtain tractable analytical results, the cognitive radio network has been greatly simplified. While this will certainly restrict the application of results obtained in this paper, our work provides a first attempt in finding delay-constrained capacity for cognitive radio networks, which also sheds light on how delay-constrained capacity analysis may be conducted for more complex or realistic settings such as multi-channel and none ideal coordination among the same type of users in channel accessing. For such scenarios, additional and likely significant effort will be needed. We leave this as future work.

Acknowledgement

This work was sponsored by BUPT Fund for Young Scholars under grant No.2011RC0114. The authors would like to thank the anonymous reviewers for their insightful comments that helped improve the quality of this paper.

Bibliography

[1] J.I. Mitola and G.J. Maguire, “Cognitive radio: making software radios more personal,” IEEE Personal Communications, vol.6, pp.13–18, Aug. 1999.

[2] Y. Jiang and Y. Liu, Stochastic Network Calculus, Springer, 2008.

[3] I. Suliman and J. Lehtomaki, “Queueing analysis of opportunistic access in cognitive radios,” Proc. 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management, Aalborg, Denmark, pp.153–157, May 2009.

[4] C. Zhang, X. Wang, and J. Li, “Cooperative cognitive radio with priority queueing analysis,” Proc. IEEE International Conference on Communications (ICC ’09), Dresden, Germany, pp.1–5, June 14–18, 2009.

[5] C. Ghosh, S. Pagadarai, D.P. Agrawal, and A.M. Wyglinski, “Queuing theory representation and modeling of spectrum occupancy employing radio frequency measurements,” Proc. IEEE Vehicular Technology Conference (VTC 2009-Fall), Anchorage, USA, pp.1–5, Sept. 20–23, 2009.

[6] C. An, H. Ji, and P. Si, “Dynamic spectrum access with qos provisioning in cognitive radio networks,” Proc. IEEE GLOBAL COMMUNICATIONS CONFERENCE, Miami, USA, pp.1–5, Dec.6-10, 2010.

[7] X. Gelabert, O. Sallent, J. Perez-Romero, and R. Agusti, “Spectrum sharing in cognitive radio networks with imperfect sensing: A discrete-time markov model,” Computer Networks, vol.54, pp.2519–2536, April 2010.

[8] Y.R. Kondareddy, N. Andrews, and P. Agrawal, “On the capacity of secondary users in a cognitive radio network,” IEEE Sarnoff Symposium, Princeton, USA, pp.1–5, March 2009.

[9] S. Tang and B.L. Mark, "Modeling and analysis of opportunistic spectrum sharing with unreliable spectrum sensing," IEEE Trans. on Wireless Communications, vol.8(4), pp.1934–1943, April 2009.

[10] R.L. Cruz, "A calculus for network delay, part I: Network elements in isolation," IEEE Trans. Inf. Theory, vol.37(1), pp.114–131, Jan. 1991.

[11] R.L. Cruz, "A calculus for network delay, part II: Network analysis," IEEE Trans. Inf. Theory, vol.37(1), pp.132–141, Jan. 1991.

[12] Q. Yin, Y. Jiang, S. Jiang, and P.Y. Kong, "Analysis on generalized stochastically bounded bursty traffic for communication networks," Proc. IEEE LCN'02, 2002.

[13] Y. Jiang, Q. Yin, Y. Liu, and S. Jiang, "Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks," Computer Networks, vol.53(12), pp.2011–2021, 2009.

[14] M. Fidler, "A survey of deterministic and stochastic service curve models in the network calculus," IEEE Communications Surveys & Tutorials, vol.12(1), pp.59–86, 2010.

[15] F. Kelly, "Notes on effective bandwidths," Stochastic Networks: Theory and Applications, Royal Statistical Society Lecture Notes Series, 4. Oxford University Press, 1996.

[16] C.S. Chang, Performance Guarantees in Communication Networks, Springer-Verlag, 2000.

[17] Y. Gao and Y. Jiang, "Performance analysis for a cognitive radio with imperfect spectrum sensing," Proc. IEEE INFOCOM 2010 - Workshop on Cognitive Wireless Communications and Networking, pp.1–5, March 2010.

[18] J. Andrews, N. Jindal, M. Haenggi, R.A. Berry, S. Jafar, D. Guo, S. Shakkottai, R.H. Jr, M. Neely, S. Weber, A. Yener, and P. Stone, "Rethinking information theory for mobile ad hoc networkss," IEEE Communications Magazine, Dec 2008.

[19] 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (TS 23.401 v8.0.0)," tech. rep., 2007.

[20] J.L. Doob, Stochastic Processes, Wiley, 1953.

Appendix F

Publication F

Yuehong Gao, Wenting Jiang and Yuming Jiang; Guaranteed Service and Delay-Constrained Capacity of a Multi-Channel Cognitive Secondary Network; 7th International Conference on Cognitive Radio Oriented Wireless Networks (CrownCom); June 2012.

Abstract

In this paper, we consider a multi-channel cognitive radio system serving a primary network and a secondary network, and analyze the quality of service and delay-constrained capacity of the secondary network. Specifically, by assuming that a certain amount of resource is exclusively reserved and used on each channel by the primary network, we derive the traffic transportation capacity that is guaranteed to the secondary network. Based on this, we analyze the traffic delay distribution in the secondary network and derive an upper bound on it, which allows us to further obtain a guaranteed capacity of the secondary network in serving traffic with probabilistic delay requirement. Both numerical and simulation results are presented for an example where the secondary network traffic follows a model taken from 3GPP LTE. The delay distribution, average delay and delay-constrained capacity of the secondary network are compared. The excellent match between analytical results and simulation results validates the theoretical analysis.

F.1 Introduction

Cognitive radio is a promising technique for efficiently making use of wireless spectrum [1]. Its fundamental idea is to allow a secondary network to coexist with the primary network in the system, and the secondary network can also access the system resource (or wireless channels) as far as the performance of the primary network is not affected. In this paper, we consider such a cognitive radio system with focus on the quality of service (QoS) performance of the secondary network. Specifically, we analyze the traffic transportation capacity that can be guaranteed to the secondary network, investigate its delay performance, and obtain its capacity in serving traffic with delay requirement.

In the literature, several attempts have been made to conduct performance analysis of the secondary network. Some of them make use of classical queueing theory [2, 3]. In order to directly apply existing results, typically $M/G/1/$ Priority analysis, Poisson arrival and single channel scenario are assumed with average delay and average queue length as the performance metrics of interest. In some other works, e.g. [4, 5], a classic stochastic process analysis technique is used, which establishes its basis on the states of each channel occupancy, i.e. whether a channel is occupied by which network, and uses a Markov chain to model this process. With such Markov chains, the dropping probability and blocking probability are derived. Although multi-channel is considered in these works, it is often assumed that the arrivals (to each channel) form a Poisson process and the service time (of each channel occupancy) follows some negative exponential distribution in order to ensure the Markov property of the channel occupancy process. In addition, some results on outage/ergodic capacity are available, e.g. [6, 7], under various constraints that include power constraints and peak interference power constraints. However, study on the maximum arrival rate under probabilistic delay constraint (defined as guaranteed delay-constrained capacity in this paper) is very limited.

Furthermore, the problem becomes even challenging when the cognitive radio network is supposed to have multiple parallel channels. For analyzing multi-channel/multi-server systems, another novel approach has been adopted, which lays on the *network calculus* theory [8, 9]. For example, in [10, 11], service guarantee analysis of multi-server Weighted Fair Queueing and multi-server Round Robin scheduling systems have been respectively studied. However, the considered multi-server scenarios therein do not encompass the priority issue, which is typically inherent in cognitive radio systems. To the best of our knowledge, an analysis of multi-channel cognitive radio system with general traffic model and probabilistic delay requirement is yet to be found, which has motivated the present work.

The objective of this paper is to analyze the quality of service and capacity of a cognitive radio secondary network. Specifically, by assuming that a certain amount of resource is exclusively reserved and used on each channel by the primary network, we first derive the traffic transportation capacity that is guaranteed to the secondary network. Then, we analyze the traffic delay distribution in the secondary network and derive an upper bound on it. This delay distribution bound allows us to further obtain a guaranteed capacity of the secondary network in serving traffic with delay requirement. To validate the analysis, both numerical and simulation results are presented by using a 3GPP LTE scenario as an example. The delay distribution, average delay and capacity of the secondary network are compared and discussed. The comparison shows an excellent match between numerical and simulation results.

The rest is organized as follows. Sec.F.2 describes the considered system model. Sec.F.3 presents a backlog period analysis and derives the guaranteed service provided to the secondary network. Then, delay analysis is conducted in Sec.F.4. Sec.F.5 presents numerical results and compares with simulation results. Finally, further discussion and concluding remarks are made in Sec.F.6 and Sec.F.7, respectively.

F.2 The System Model

In this paper, we consider a cognitive radio system with multiple independent channels indexed by i ($1 \leq i \leq N$) as shown in Fig.F.1. In this system, all channels are slotted[1] and synchronized, where T denotes the slot *time* length. Only at the beginning of a slot, scheduling is made and transmission can start. On each channel i, a certain number of slots are reserved[2] *periodically* and *exclusively* for the primary network in order to guarantee its service. Here, by exclusively, we mean that such slots are never used by the secondary network. When there is traffic, the primary network will always try to use such slots first.

Throughout the rest of this paper, we assume that both the primary network and the number of reserved slots are properly planned such that no additional slots are needed for traffic of the primary network. While this assumption is rather conservative, it *guarantees* a certain amount of service available to the secondary network, which we believe is reasonable and can be expected particularly when the secondary network needs to

[1] One slot is defined as the smallest transmission unit in time domain.

[2] The considered reservation works in the time domain, but, we would like to emphasize that the analysis can also be applied/extended when the reservation is made in other domains such as in the frequency domain.

pay the primary network. Under this channel reservation, each channel becomes an *ON-OFF* process from the viewpoint of the secondary network. With this information, we believe channel sensing will work much more effectively and correctly. Due to this, sensing error is ignored in this paper. Recall that, the objective of the paper is to find the traffic transportation capacity that is guaranteed to the secondary network, which also implies perfect sensing.

We suppose each wireless channel has constant transmission rate C_i. We define R_i as the length (in number of slots) of a reservation period on channel i, and R_i^{on} as the *number* of slots reserved by the primary network in each reservation period on channel i. In addition, $\eta_i = \frac{R_i^{on}}{R_i}$ is called as the active factor of the primary network on channel i. Fig.F.1 depicts the aforementioned cognitive radio system, where the traffic generated by the primary and secondary networks are denoted by fl^p and fl^s, respectively.

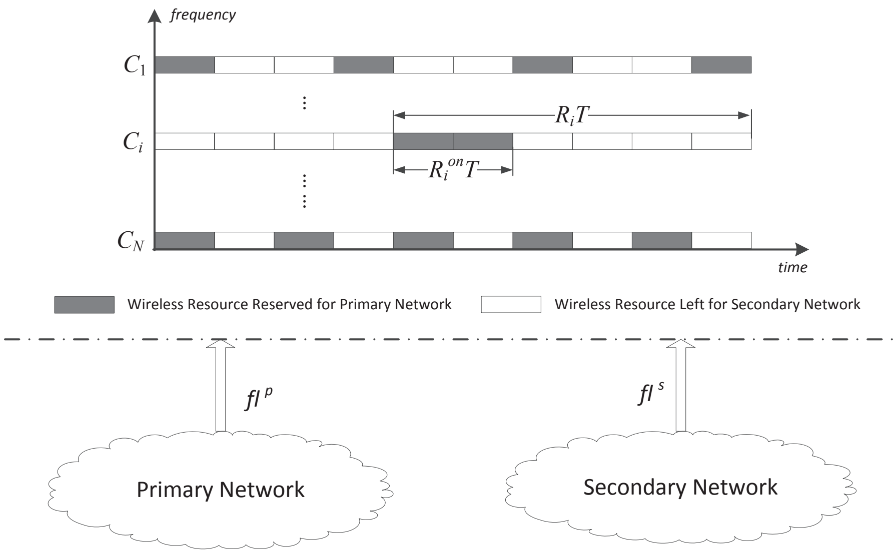


Figure F.1: Considered Cognitive Radio Network

F.3 Guaranteed Service Analysis

By observing the considered system, we notice that every channel provides a deterministic amount of service during any given period. In addition, there also exists an upper bound on the amount of resource reserved for the primary network during any period. Therefore, the amount of service that can be utilized by the secondary network is lower bounded. Intuitively, one may guess the long term average service rate of each channel i, which is available to the secondary network, is $(1-\eta_i)\cdot C_i$ and hence, the total long term average service rate, which the secondary network can maximally get, is $\sum_{i=1}^{N}(1-\eta_i)\cdot C_i$. In this section, we present results on the amount of service that can be provided to the secondary network, which not only leads to a rigorous validation of the average service rate intuition, but also allows us to view this service on short time scale, which is crucial for QoS analysis of the secondary network.

The following theorem presents the main result, which lays the foundation for later delay analysis of the secondary network.

Theorem 1. ***(Guaranteed Service.)*** *For the considered cognitive radio system, the amount of service provided to the secondary network during its any backlogged period* $(\tau, \tau+t]$*, denoted by* $W^s(\tau, \tau+t)$*, satisfies,*

$$C^s \cdot t + U^s \geq W^s(\tau, \tau+t) \geq C^s \cdot (t - L^s)^+$$

where $C^s = \sum_{i=1}^{i=N}(1-\eta_i)C_i$*,* $L^s = \max_{1\leq i\leq N}(2R_i^{on}T + 2T)$*,* $U^s = \sum_{i=1}^{N} C_i$ $(R_i^{on}+1)T$*, and* $x^+ \equiv \max\{x, 0\}$*.*

Before proving Th.1, we first discuss how this result can help prove the long term average service rate of the secondary network. Note that the maximum long term average service rate is achieved when there is always traffic to send from the network, and hence can be written as $\lim_{t\to\infty}\frac{W(\tau,\tau+t)}{t}$. Then, with the first part of Th.1, we get

$$\lim_{t\to\infty}\frac{W(\tau,\tau+t)}{t} \leq \lim_{t\to\infty}\frac{C^s\cdot t + U^s}{t} = C^s$$

and with the second part of Th.1, we get

$$\lim_{t\to\infty}\frac{W(\tau,\tau+t)}{t} \geq \lim_{t\to\infty}\frac{C^s\cdot (t-L^s)^+}{t} = C^s.$$

Summing up, we conclude:

Corollary 1. ***(Long Term Average Service Rate.)*** *The long-term average service rate that the secondary network can maximally provide is* $C^s = \sum_{i=1}^{i=N}(1-\eta_i)C_i$*.*

F.3.1 Proof of Theorem 1

The rest of this section is devoted to the proof of Th.1. In this paper, we assume that the amount service of a slot is delivered or received by a network when and only when the slot ends and it is allocated to this network at the start of this slot. The rationale of this assumption is that in packet-switched networks, a packet is considered to be serviced when and only when its last bit has been serviced.

Since all channels are independent with each other, the analysis on each channel is the same. Hence, we start by considering an arbitrary channel indexed with i, and later, the analysis will be extended to the whole system. Consider any backlog time period $(s, t]$ $(0 \leq s \leq t)$ for the secondary flow fl^s, which means that there is always traffic waiting to be served in the secondary network during this period. Therefore, we have

$$W_i^s(s,t) = W_i(s,t) - W_i^p(s,t), \tag{F.1}$$

where $W_i(s,t)$ denotes the total amount of service that can be provided by channel i, $W_i^p(s,t)$ the amount of service that may maximally occupied by fl^p, and $W_i^s(s,t)$ the amount of service occupied by fl^s. Note that $W_i^p(s,t)$ essentially denotes the amount of service of the reserved slots in $(s,t]$ by the primary network, and due to exclusive reservation, these slots are not used by the secondary network though they may not carry traffic from the primary network.

We first prove the second part of Th.1.

In Eq.(F.1), $W_i(s,t)$ can be easily obtained as $W_i(s,t) \geq C_i \cdot (t-s-2T)^+$. The amount of service provided within two slot $C_i \cdot 2T$ are deducted, because the worse case happens when the time point s (or t) locates just after (or before) a slot starts (or ends), indicating the first slot and the last slot during $(s,t]$ are not complete slots. In addition, $(\cdot)^+$ is due to the fact that the amount of service cannot be negative.

Regarding the specific expression for $W_i^s(s,t)$, there are two scenarios to be analyzed.

- *Scenario 1: The time length $(t-s)$ is no longer than one reservation period, i.e., $t - s \leq R_i T$.*

  In this scenario, service reserved for fl^s is upper bounded by $C_i \cdot R_i^{on} T$ as shown in Fig.F.2(a). Therefore, we have

$$\begin{aligned} & W_i^s(s,t) = W_i(s,t) - W_i^p(s,t) \\ \geq\ & [C_i \cdot (t-s-2T) - C_i \cdot R_i^{on} T]^+ \\ =\ & C_i \cdot (t-s-2T-R_i^{on}T)^+. \end{aligned} \tag{F.2}$$

- *Scenario 2: The time length* $(t-s)$ *lasts longer than one reservation period, i.e.,* $t-s > R_iT$.

  In this scenario, let s' denote the start time of the next period just after s, and t' denote the end time of the latest period just before t, as illustrated by Fig.F.2(b). Then, the amount of service provided to the secondary network consists of three parts:

  $$W_i^s(s,t) = W_i^s(s,s') + W_i^s(s',t') + W_i^s(t',t), \tag{F.3}$$

  where the length of $(s'-s)$ and $(t-t')$ are shorter than one period cycle R_iT, and therefore, $W_i^s(s,s')$ and $W_i^s(t',t)$ fall into the range of *Scenario 1*. Then, there hold:

  $$W_i^s(s,s') \geq C_i \cdot (s'-s-R_i^{on}T-T)^+ \tag{F.4}$$
  $$W_i^s(t',t) \geq C_i \cdot (t-t'-R_i^{on}T-T)^+. \tag{F.5}$$

  Note that, only one slot length is deducted in (F.4) and (F.5) compared with (F.2), because s' and t' are at the edge of a slot.

  Intuitively from the definition, we know the time length between s' and t' is integer times of one period R_iT, and hence, the service left for the secondary network during $[s',t']$ can be obtained deterministically as

  $$W_i^s(s',t') = \frac{t'-s'}{R_iT} \cdot C_i \cdot (R_iT - R_i^{on}T). \tag{F.6}$$

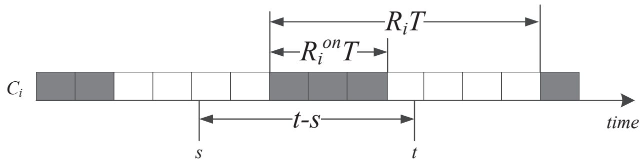
  Combining these scenarios together, we have

  $$\begin{aligned} W_i^s(s,t) &= W_i^s(s,s') + W_i^s(s',t') + W_i^s(t',t) \\ &\geq C_i \cdot \left(s'-s-R_i^{on}T-T\right)^+ \\ &\quad + C_i \cdot \left(1-\frac{R_i^{on}}{R_i}\right)(t'-s') \\ &\quad + C_i \cdot \left(t-t'-R_i^{on}T-T\right)^+ \\ &\geq (1-\eta_i)C_i \cdot (t-s-2R_i^{on}T-2T)^+. \end{aligned}$$

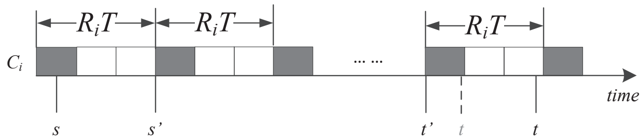
Considering the fact of $R_i^{on} \geq 1$ and $\eta_i \leq 1$, the result of *Scenario 1* and *Scenario 2* can be further merged as

$$\begin{aligned} W_i^s(s,t) &\geq (1-\eta_i)C_i \cdot (t-s-2R_i^{on}T-2T)^+ \\ &\triangleq C_i^s \cdot (t-s-L_i^s)^+ \end{aligned}$$

for any backlog period $(s,t]$.



(a) Scenario 1



(b) Scenario 2

Figure F.2: Illustration of Two Scenarios

Then, an lower bound on the amount of service provided by the whole system to the secondary network can be obtained by making a summation as

$$\begin{aligned} W^s(s,t) &= \sum_{i=1}^{i=N} W_i^s(s,t) \\ &\geq \sum_{i=1}^{i=N} C_i^s(t-s-L_i^s)^+ \\ &\geq (\sum_{i=1}^{i=N} C_i^s)\cdot(t-s-\max_{1\leq i\leq N} L_i^s)^+ \\ &\triangleq C^s\cdot(t-s-L^s)^+ = \beta^s(t-s), \end{aligned} \tag{F.7}$$

which ends the proof of the second part.

For the first part, the prove follows similarly. Particularly, it can be easily verified that $W_i(s,t) \leq C_i \cdot (t-s+T)$. In addition, in any time interval $(s,t]$, the number of reserved slots for the primary network is not smaller that $\lfloor \frac{t-s}{R_iT} \rfloor R_i^{on}$ and hence the corresponding time length not shorter than $(\frac{t-s}{R_iT}-1)^+ \cdot (R_i^{on}T)$. We then have

$$W_i^p(s,t) \geq (\frac{t-s}{R_iT}-1)^+ \cdot (R_i^{on}T)C_i$$

and hence

$$W_i^s(s,t) \;\;\leq\;\; (1-\eta_i)C_i\cdot(t-s) + C_i\cdot T\cdot(R_i^{on}+1)$$

with which the first part is proved by summing up all channels' service.

F.4 Delay Distribution Analysis

Th.1 fundamentally indicates the amount of service that can be guaranteed for the secondary network. With this, the following theorem presents that, if the traffic arrival process of the secondary network is stochastically bounded, the traffic delay (including queueing delay and transmission time) in the secondary network is probabilistically upper bounded.

Theorem 2. ***(Delay Distribution.)*** *For the considered cognitive radio system, if the amount of traffic of the secondary network $A^s(s,t)$ is stochastically bounded by an arrival function $\alpha^s(t) \in F^3$ and a probability distribution function $f(x) \in \bar{F}^4$, i.e., there holds*

$$P\{\sup_{0\leq s\leq t}\{A^s(s,t) - \alpha^s(t-s)\} > x\} \leq f(x), \tag{F.8}$$

then the system delay for any traffic from the secondary network is probabilistically upper bounded by

$$P\{d^s > h(\alpha^s(t) + x, \beta^s(t - L^\sigma))\} \leq f(x), \tag{F.9}$$

where $\beta^s(t) \equiv C^s \cdot (t - L^s)^+$, $h(\alpha^s(t) + x, \beta^s(t - L^\sigma)$ is the maximum horizontal distance between $\alpha^s(t) + x$ and $\beta^s(t - L^\sigma)$, and $L^\sigma = \sigma^s_{max}/C^s$ is the latency of serving the largest unit of traffic denoted as σ^s_{max}.

Here we would like to remark the difference between L^s and L^σ in Th.2. L^s is given in Th. 1, denoting the latency term if the service would have been defined using the Latency Rate server model [12]. However, L^σ is a time length related to serving the largest traffic unit, such as maximum length packet or maximum length file, in the secondary network. This difference is clearly seen from the example given in the next section. In addition, we would like to highlight that the literature has proved that a lot types of traffic satisfy (F.8) and extensive discussion on this can be found from e.g. [9, 13].

Define delay-constrained capacity as the maximum long term traffic rate that can be supported by a network under delay constraint (D, ϵ), denoted

[3] F: the set of non-negative wide-sensing increasing functions
[4] $\bar{F}$: the set of non-negative wide-sensing decreasing functions

by $C_{(D,\epsilon)}$. Specifically, for the secondary network, the delay-constrained capacity $C^s_{(D,\epsilon)}$ is defined as

$$C^s_{(D,\epsilon)} \equiv \max \lim_{t\to\infty} \frac{A(\tau,\tau+t)}{t} \quad \text{such that} \quad P\{d^s > D\} = \epsilon.$$

With Th.2, the following result is immediately obtained:

Corollary 2. ***(Guaranteed Delay-Constrained Capacity.)*** *It is guaranteed that the delay-constrained capacity of the secondary network is not smaller than* $\max \lim_{t\to\infty} \frac{\alpha(t)}{t}$*, where* $x = f^{-1}(\epsilon)$ *is the inverse function of* ϵ *and* $\alpha(t)$ *satisfies*

$$P\{h(\alpha^s(t) + x, \beta^s(t - L^\sigma) \le D\} \le \epsilon.$$

F.4.1 Proof of Theorem 2

Consider any traffic unit σ^s_j that arrives at the secondary network at time t. There exists a time point $0 \le t_0 \le t$ which is the start of the backlog period containing time t. We can always find such t_0, because at least the arrival of itself will start the backlog period. Then, the system delay can be expressed as

$$d^s_j = \inf\{\tau : A^s(t_0, t) \le A^s_{out}(t_0, t+\tau)\}. \tag{F.10}$$

We can prove that for any $x \ge 0$, if $d^s_j > x$, there must be $A^s(t_0, t) > A^s_{out}(t_0, t+x)$, since otherwise if $A^s(t_0, t) \le A^s_{out}(t_0, t+x)$, then $d^s_j \le x$ should hold, which will contradict the condition $d^s_j > x$. To sum up, event $d^s_j > x$ implies event $A^s(t_0, t) > A^s_{out}(t_0, t+x)$. Therefore, it holds:

$$P\{d^s_j > x\} \le P\{A^s(t_0, t) > A^s_{out}(t_0, t+x)\}. \tag{F.11}$$

Note that $A^s_{out}(t_0, t+x)\} \le W^s(t_0, t+x)$, because the system may be busy severing previous traffic unit $\sigma^s_k \le \sigma^s_{max}$ at time t_0, which is the last one sent to the system and empties the buffer before time t_0. Therefore, we have

$$W^s(t_0, t+x) - \sigma^s_{max} \le A^s(t_0, t+x) \le W^s(t_0, t+x). \tag{F.12}$$

Then, the following steps hold:

$$\begin{aligned}
& P\{A^s(t_0,t) > A^s_{out}(t_0,t+x)\} \\
\leq \;& P\{A^s(t_0,t) > W^s(t_0,t+x) - \sigma^s_{max}\} \\
\leq \;& P\{A^s(t_0,t) > \beta^s(t+x-t_0) - \sigma^s_{max}\} \\
\leq \;& P\{A^s(t_0,t) > \beta^s(t+x-t_0 - \frac{\sigma^s_{max}}{C^s})\} \\
= \;& P\{A^s(t_0,t) - \alpha^s(t-t_0) > \\
& \beta^s(t+x-t_0-L^\sigma) - \alpha^s(t-t_0)\},
\end{aligned}$$

where $\alpha^s(t)$ is a non-negative wide-sensing increasing function, and L^σ is the latency term introduced by serving σ^s_{max}. When $\lim_{t\to\infty}\frac{\alpha^s(t)}{t} < \lim_{t\to\infty}\frac{\beta^s(t)}{t}$ holds, there exists a maximum horizontal distance between $\alpha^s(t)+y$ and $\beta^s(t-L^\sigma)$ for $\forall y > 0$, defined as:

$$\begin{aligned}
& h(\alpha^s(t)+y, \beta^s(t-L^\sigma)) \\
= \;& \sup_{t\geq 0}\{\inf\{\zeta \geq 0 : \alpha^s(t)+y \leq \beta^s(t-L^\sigma+\zeta)\}\}.
\end{aligned} \tag{F.13}$$

By setting $x = h(\alpha^s(t)+y, \beta^s(t-L^\sigma))$ and by definition (F.13), we have:

$$\begin{aligned}
& P\{d^s_j > h(\alpha^s(t)+y, \beta^s(t-L^\sigma))\} \\
\leq \;& P\{A^s(t_0,t) - \alpha^s(t-t_0) > y\} \\
\leq \;& P\{\sup_{0\leq t_0\leq t}\{A^s(t_0,t) - \alpha^s(t-t_0)\} > y\}.
\end{aligned} \tag{F.14}$$

where the $\sup_{0\leq t_0\leq t}\{\cdot\}$ in (F.14) is used to remove the randomness of t_0.

When the arrival process $A^s(s,t)$ is stochastically bounded as defined in (F.8), there holds

$$P\{d^s > h(\alpha^s(t)+y, \beta^s(t-L^\sigma))\} \leq f(x), \tag{F.15}$$

which ends the proof.

F.5 Numerical and Simulation Results

In this section, we provide both numerical results and simulation results for the uplink of a FDD LTE system. As specified by [14], the uplink transmissions are organized into radio frames with duration of 10 ms, which is employed as the smallest transmission unit here, i.e., $T = 10$ ms. The number of slots in a reservation period (i.e., R_i) is set to 10 frames. The active factor η_i varies within the range of $[0.1, 0.2, ..., 0.9]$. Along the frequency axis, the system is "grided" into Resource Blocks (RB), and hereafter, each RB is considered as a single channel (i.e., $N = N^{RB}$), which varies depending on the system bandwidth within $\{15, 25, 50, 75, 100\}$ channels. Each RB in frequency domain contains 12 sub-carriers and the channel rate C_i is 224 kbps under 1/3 coding rate and $16QAM$ modulation.

The traffic generated by the secondary network is considered to follow a non-full buffer FTP model as suggested by [15]. Specifically, file is the concerned traffic unit with fixed length of $\sigma^s = 4Mbits$ indicating bursty traffic. The bursty file is divided into small packets to fill each slot when transmitted on channels. The arrival process of files is supposed to be a Poisson process[5]. The *delay* to be presented later is defined as the time length between the arrival of a file and the end point of the slot containing the last bit of this file. This arrival process is a *compound* Poisson process. It can be proved that the considered process is stochastically bounded by Eq.(F.8) with the following setting [16, 17]:

$$\begin{aligned} \alpha^s(t) &= \frac{\lambda^s}{\theta}\left(e^{\sigma^s\theta} - 1\right)t \\ f(x) &= e^{-\theta x} \end{aligned}$$

where $\theta > 0$ is a free parameter and can be used to optimize the results presented later.

By applying Th.1 and the aforementioned configurations into Th. 2, the probabilistic delay distribution bound can be expressed as:

$$P\{d^s > x\} \le f\left(C^s(x - L^s - L^\sigma)^+\right), \tag{F.16}$$

subject to

$$\frac{\lambda^s}{\theta}\left(e^{\sigma^s\theta} - 1\right) \le C^s, \tag{F.17}$$

where $C^s = (1 - \eta_i)NC_i$, $L^s = 2(R_i^{on} + 1)T$ and $L^\sigma = \sigma^s/C^s$.

Fig.F.3 compares the upper bound of delay distribution probabilities obtained by the theoretical analysis with simulation results, where the number of channels is 50, the average file arrival rate from the secondary network is 0.5 files per second and the active factor of the primary network is 0.1 and 0.5, respectively. Though a gap exists between the theoretical and corresponding simulation results, the probabilities obtained by the theoretical analysis are close to and in the same order of magnitude as the simulation results.

[5] Note that, we are talking about the service demand of flow arrival (not packet level) that has Poisson property.

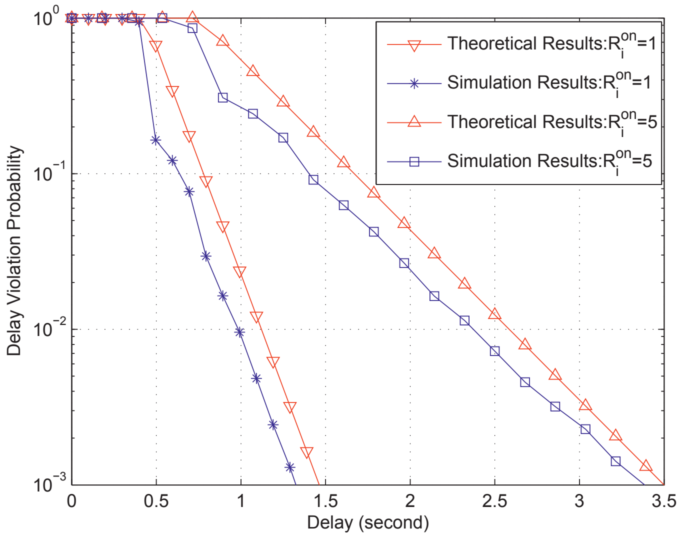


Figure F.3: Delay Distribution Probability

Fig.F.4 plots the average delay under different configurations. Firstly, the influence of R_i^{on} is investigated. It is obvious that the average delay of secondary files increases when more resource is reserved for the primary network. In addition, the average delay goes to infinity when $R_i^{on} = 9$, which is not included in the figure. Similar trend can be found when the average file arrival rate from the secondary network increases. On the contrary, more channels (equivalent to larger bandwidth) will guarantee better delay requirements. In addition, the theoretical results locate close to the simulation results, especially when the system is heavily loaded. Those heavily loaded points are usually utilized to find the guaranteed capacity for admission control.

Finally, Fig.F.5 compares the guaranteed capacity under the constraint that the delay exceeding 3 seconds has a probability less than 1%. The figure shows that the theoretical results match well with the simulation results, which validates the effectiveness of the presented analysis.

F.6 Discussion

We would like to further discuss several key issues in this work. First, the theoretical analysis here finds its root in the area of network calculus [8, 9].

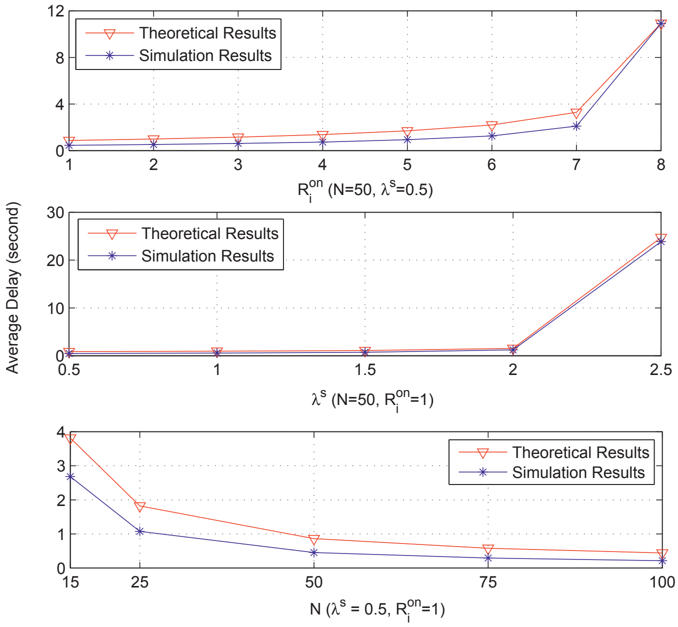


Figure F.4: Average Delay

The guaranteed service $\beta^s(t)$ in Th.1 is indeed the so-called *Latency-Rate Service Curve* [12], which is an important type of service curve in network calculus. In addition, in characterizing the stochastic arrival process of the secondary network for delay analysis, the definition in (F.8) is known as the *virtual-centric-backlog* stochastic arrival curve in stochastic network calculus [9], and $\sup\{\cdot\}$ in (F.8) cannot be omitted. Importantly, while at a first glance, one might think to apply the leftover service property in network calculus to find the service guaranteed to the secondary network, we stress that this cannot be done easily. The fundamental reason is that the available network calculus leftover service property is applicable only to *single server* systems. Furthermore, while much of the existing multi-channel analysis literature (e.g. [10, 11]) is also based on network calculus, the considered systems therein do not match with the considered cognitive radio system. All these have motivated the present work, which also imply the challenge in the analysis.

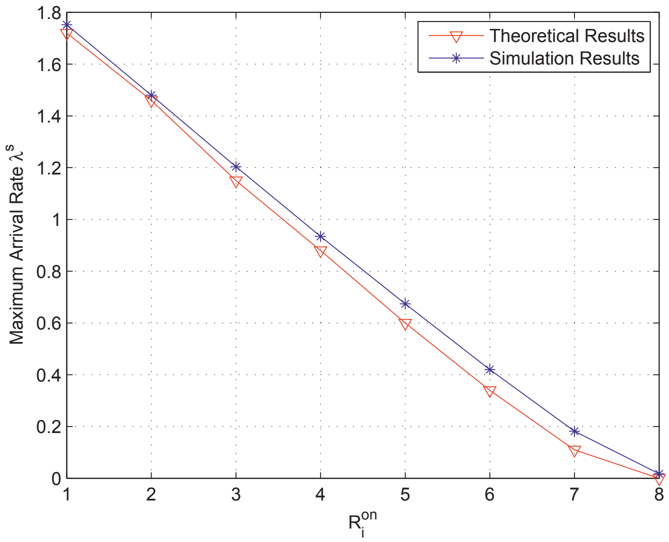


Figure F.5: Delay-Constrained Capacity of Secondary Network

F.7 Conclusion

In this paper, performance analysis of a multi-channel cognitive radio secondary network is conducted. In order to ensure a certain level of service guarantee in the cognitive radio system, it is assumed that some amount of resource is reserved for the primary network. With this assumption, we derive the guaranteed amount of service that can be provided to the secondary network. Then, an upper bound on delay distribution probability in the secondary network is obtained if its traffic arrival process is stochastically bounded. In addition, a delay-constrained capacity of the secondary network is derived. Both numerical and simulation results are presented and discussed by considering an LTE parameter setting. Specifically, delay distribution probabilities, average delay and capacity are compared, which shows a good match between the analytical results and the simulation results, indicating the effectiveness of the theoretical analysis.

We stress that in order to move forward in multi-channel cognitive radio analysis, several assumptions are made. While this will certainly restrict the application of results in this paper, the work still sheds light on how multi-channel cognitive radio systems may be analyzed. For more complex scenarios, such as fading channel and random primary traffic arrival, more effort is needed, which are our on-going work.

Bibliography

[1] J. III Mitola and G.Q. Jr. Maguire. Cognitive radio: making software radios more personal. *IEEE Personal Communications*, 6:13–18, August 1999.

[2] I. Suliman and J. Lehtomaki. Queueing analysis of opportunistic access in cognitive radios. In *Proc. 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management*, pages 153–157, Aalborg, Denmark, May 2009.

[3] C. Ghosh, S. Pagadarai, D.P. Agrawal, and A.M. Wyglinski. Queueing theory representation and modeling of spectrum occupancy employing radio frequency measurements. In *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, pages 1–5, September 2009.

[4] Yogesh R Kondareddy, Nirmal Andrews, and Prathima Agrawal. On the capacity of secondary users in a cognitive radio network. In *IEEE Sarnoff Symposium*, pages 1–5, Princeton, USA, March 2009.

[5] S.M. Kannappa and M. Saquib. Performance analysis of a cognitive network with dynamic spectrum assignment to secondary users. In *IEEE International Conference on Communications (ICC)*, pages 1–5, may 2010.

[6] Xin Kang, Ying-Chang Liang, A. Nallanathan, H.K. Garg, and Rui Zhang. Optimal power allocation for fading channels in cognitive radio networks: Ergodic capacity and outage capacity. *IEEE Transactions on Wireless Communications*, 8(2):940–950, feb. 2009.

[7] S. Stotas and A. Nallanathan. Enhancing the capacity of spectrum sharing cognitive radio networks. *IEEE Transactions on Vehicular Technology*, 60(8):3768–3779, oct. 2011.

[8] J.-Y. Le Boudec and P. Thiran. *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet.* Springer, 2001.

[9] Yuming Jiang and Yong Liu. *Stochastic Network Calculus*. Springer, 2008.

[10] Josep M. Blanquer and Banu Özden. Fair queuing for aggregated multiple links. *ACM SIGCOMM Comput. Commun. Rev.*, 31:189–197, August 2001.

[11] Haiming Xiao and Yuming Jiang. Analysis of multi-server round robin scheduling desciplines. *IEICE Transactions on Communications*, E87-B:3593–3602, December 2004.

[12] D. Stiliadis and A. Varma. Latency rate servers: a general model for analysis of traffic scheduling algorithms. *IEEE/ACM Trans. Networking*, 6(5):611–624, Oct. 1998.

[13] Yuming Jiang and et al. Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks. *Computer Networks*, 53:2011–2021, 2009.

[14] 3GPP. Evolved universal terrestrial radio access (E-UTRA): Physical channels and modulation (release 8). Technical report, March 2008.

[15] 3GPP. Evolved universal terrestrial radio access (E-UTRA): Further advancements for E-UTRA physical layer aspects (release 9). Technical report, March 2010.

[16] F. Kelly. Notes on effective bandwidths. In *Stochastic Networks: Theory and Applications, Royal Statistical Society Lecture Notes Series, 4. Oxford University Press*, 1996.

[17] Yuming Jiang. A note on applying stochastic network calculus. Technical report, 2010.