Application of Game Theory in Wireless Communication Networks

by

Wei Huang

M. Phil., Hong Kong University of Science and Technology, 2007

B. Sc., Zhejiang University, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES
(Electrical and Computer Engineering)

The University of British Columbia
(Vancouver)

February 2012

Abstract

The ability to model independent decision makers whose actions potentially affect other decision makers makes game theory attractive to analyze the performances of wireless communication systems. Recently, there has been growing interest in adopting game theoretic methods to wireless networks for power control, rate adaptation and channel access schemes. This thesis focuses the application of dynamic game theory and mechanism design in cognitive radio networks, Wireless Local Area Networks (WLAN) and Long Term Evolution (LTE) systems. The first part of the thesis aims to optimize the system performance through the transmission rate adaptation among wireless network users. The optimal transmission policy of each user is analyzed by formulating such a problem under general-sum Markov game framework. Structural results are obtained and provably convergent stochastic approximation algorithm that can estimate the optimal transmission policies are proposed. Especially, in the switching control Markov game theoretic rate adaptation formulation, it is proved that the optimal transmission policies are monotone in channel state and there exists a Nash equilibrium at which every user deploys a monotone transmission policy. This

structural result leads to a particularly efficient stochastic-approximation-based adaptive learning algorithm and a simple distributed implementation of the transmission rate control.

This thesis also considers the spectrum allocation problem in an LTE system. By incorporating cognitive capacities into femtocell base stations, the Home Node-Bs (HNBs) can be formulated as cognitive base stations competing for the spectrum resource while trying to minimize the interference introduced to the evolved Node-B (eNB) which is also referred as primary base station. Given the primary base station spectrum occupancy, the spectrum allocation problem among HNBs can be formulated under game theoretic framework. The correlated equilibrium solution of such a problem is being investigated. A distributed Resource Block (RB) access algorithm and a correlated equilibrium Q-learning algorithm are proposed to compute the spectrum allocation solutions under static and dynamic environments, respectively. The last part of the thesis uses mechanism design to design a truth revealing opportunistic scheduling system in a cognitive radio system. A mechanism learning algorithm is provided for users to learn the mechanism and to obtain the Nash equilibrium policy.

Preface

This thesis is completed based on the following publications. In some cases the conference papers contain materials overlapping with the journal papers.[1]

Book Chapters:

- J. W. Huang, H. Mansour, and V. Krishnamurthy, "*Transmission Rate Adaptation in Multimedia WLAN - A Dynamical Games Approach,*" CRC Press, Taylor and Francis Group, 2011.

- J. W. Huang, V. Krishnamurthy, "*Transmission Control in Cognitive Radio as a Markovian Dynamic Game - Structural Result on Randomized Threshold Policies*," Auerbach Publications, CRC Press, Taylor and Francis Group, 2009.

Journal Papers:

- J. W. Huang and V. Krishnamurthy, "Transmission Control in Cognitive Radio as a Markovian Dynamic Game - Structural Result on Randomized Threshold Policies," *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 301-310, February 2010.

[1]The author also publishes under the names of J. W. Huang and Jane Wei Huang.

- J. W. Huang, H. Mansour, and V. Krishnamurthy, "A Dynamical Games Approach to Transmission Rate Adaptation in Multimedia WLAN," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3635-3646, July 2010.

- J. W. Huang and V. Krishnamurthy, "Cognitive Base Stations in LTE/3GPP Femtocells: A Correlated Equilibrium Game Theoretic Approach," *IEEE Transactions on Communications*, December 2011.

- J. W. Huang and V. Krishnamurthy, "Game Theoretical Issues in Cognitive Radio Systems," *Journal of Communications*, vol. 4, no. 10, pp. 790-802, November 2009. (**Invited Paper**)

- J. W. Huang, H. Mansour, and V. Krishnamurthy, "A Dynamical Game Approach to Transmission Rate Adaptation in Wireless Multimedia Networks," *IEEE MMTC e-letter*, vol. 4, no. 8, pp. 11-13, September 2009. (**Invited Paper**)

Conference Papers:

- J. W. Huang and V. Krishnamurthy, "Transmission Control in Cognitive Radio Systems with Latency Constraints as a Switching Control Dynamic Game," in *Proceedings of IEEE CDC*, pages 3823-3828, December 2008.

- J. W. Huang and V. Krishnamurthy, "Rate Adaptation for Cognitive Radio Systems with Latency Constraints," in *Proceedings of IEEE Globecom*, pages 1-5, November-December 2008.

- J. W. Huang and V. Krishnamurthy, "Truth Revealing Opportunistic Scheduling in Cognitive Radio Systems," in *Proceedings of IEEE SPAWC*, June 2009.

- H. Mansour, J. W. Huang, and V. Krishnamurthy, "Multi-User Scalable Video Rate Control in Cognitive Radio Networks as a Markovian Dynamic Game," in *Proceedings of IEEE CDC*, December, 2009.

- J. W. Huang, Q. Zhu, V. Krishnamurthy, and T. Basar, "Distributed Correlated Q-Learning for Dynamic Transmission Control of Sensor Networks," in *Proceedings of IEEE ICASSP*, March 2010.

Table of Contents

List of Tables

List of Figures

Glossary

3GPP 3rd Generation Partnership Project

AWGN Additive White Gaussian Noise

BER Bit Error Rate

CCI Co-Channel Interference

CGS Coarse-Grained Scalability

CIF Common Intermediate Format

CSMA/CA Carrier Sense Multiple Access with Collision Avoidance

DCF Distributed Coordination Function

DDB Drop Dependency Based

DFP Dynamic Frequency Planning

DPB Drop Priority Based

DSL Digital Subscriber Line

eNB Evolved Node-B

HeNB Home Evolved Node-B

HSDPA High Speed Download Package Access

IEEE Institute of Electrical and Electronics Engineers

IUI Inter User Interference

JSVM Joint Scalable Video Model

LMS Least Mean Squares

LTE Long Term Evolution

MAC Medium Access Control

MAD Mean Absolute Difference

Mbps Megabit Per Second

MDP Markov Decision Process

MGS Medium-Grained Scalability

MSE Mean Squared Error

NALU Network Abstraction Layer Unit

ODE ordinary-differential-equation

OFDMA Orthogonal Frequency-Division Multiple Access

PER Packet Error Rate

PSNR Peak Signal-to-Noise Ratio

QAM Quadrature Amplitude Modulation

QoS Quality of Service

QP Quantization Parameter

RB Resource Block

SNR Signal-to-Noise Ratio

SON Self-Organized Network

SPSA Simultaneous Perturbation Stochastic Approximation

SVC Scalable Video Coding

TDMA Time Division Multiple Access

UMTS Universal Mobile Telecommunication System

VCG Vickrey-Clark-Groves

WLAN Wireless Local Area Network

Y-PSNR Luminance Peak Signal-to-Noise Ratio

Acknowledgments

I would like to thank Dr. Krishnamurthy for his tireless efforts to provide me with the most dedicated supervision. I am indebted to Dr. Krishnamurthy for teaching me technical writing and introducing me to several exciting research areas. I am also very grateful for his encouragement, generosity with time and valuable guidance throughout my doctorate program.

I also would like to thank the Enterprise Software Solutions team, Research in Motion (RIM) for the 8-months coop opportunity to work on the SharePoint project. I hugely benefited and wish to thank profusely for what I learned from the project. I also enjoyed the works I have been assigned. It is a pleasure working with the team.

I am grateful to Quanyan Zhu and Dr. Tamer Basar at the University of Illinois at Urbana-Champaign, for their collaboration, as well as the discussions that helped defining some research directions for my degree.

My warmest thanks are to my colleagues, whom I would miss the most when I no longer come to the lab everyday. Together we had a lot of happiness and I forever thankful for their sharing and caring, for the laughters, and the friendship that I treasure. Especially, Sahar Monfared, Raymond

Lee, Kevin Topley, Omid Namvar, Alex Wang and Mustafa Fanaswala have been extremely good friends. I also would like to thank my dearest friend from the wireless communication lab: Wing-Kwan Ng, whom I have known for more than 6 years. I will never forget the lunch time we shared together everyday.

My very heartfelt thanks are to Emad Hajimiri. I appreciate him for being such a good friend during all the good times and bad times. His friendship has been a very important part of my life.

My deepest thoughts are, as always, with my parents, my younger brother and my younger sister. All of life, they have always been there to take care, to teach, to support and to love. I thank my father and my mother for teaching me to be a good person, and for letting me know that their faith in me would never fade.

Dedication

To my family, whom I love the most.

Chapter 1

Introduction

Wireless networks optimization is a vast area of research, in which cross-layer optimization has emerged as a new promising approach. Telecommunication systems such as cellular networks, the Internet, Wireless Local Area Network (WLAN)s, Long Term Evolution (LTE) systems and cognitive radio systems have traditionally been designed using layered architecture based models. A layered architecture allows conceptional layers to be designed independently and hence simplifies the design, implementation and deployment of communication networks. For wireless systems, however, there are other challenges which create dependencies across layers and make the traditional modular design approach inefficient. For example, the wireless channel quality, buffer state, and the resource availability are typically varying with time. This time-varying nature of the wireless communication systems immediately motivates transmission adaptation, resource allocation and opportunistic scheduling, which exploit the statics of the system parameters for transmission, resource or rate control decisions.

Another motivation for cross-layer optimization and design is the fact that the spectrum resource and wireless channel is inherently a shared medium. This issue creates interdependencies among users, and also between the physical and higher layers (e.g., Medium Access Control (MAC) layer). Additionally, the wireless medium also offers some new modalities of communication that the layered architecture do not accommodate [84]. For example, the physical layer of a wireless network may be capable of receiving multiple packets at the same time, or partial collision resolution. In such scenarios, the traditional approach of designing MAC protocol independently of the physical layer will not be adequate. An additional difference between wireless and wired networks that motivates cross-layer approaches is the broadcast nature of the wireless channel allows users to collaborate [84].

The wireless networks considered in this thesis include cognitive radio networks, WLAN multimedia systems and LTE systems. Cognitive radio systems [12, 38] present the opportunity to improve spectrum utilization by detecting unoccupied spectrum holes and assigning them to secondary users. Resource management in cognitive radio networks involves user scheduling and transmission rate adaptation.

Institute of Electrical and Electronics Engineers (IEEE) 802.11 WLAN has been widely accepted as the dominant technology for indoor broadband wireless networking. The most important differences between the WLAN and the MAC protocols of most wireless networks is the application of a Distributed Coordination Function (DCF) for sharing access to the medium based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol. In the CSMA/CA mechanism, a node

listens to the channel before transmission to determine whether someone else is transmitting and an acknowledgement packet is sent a short time interval after the receiving node receives the packet. Chapter 3 considers the scheduling, rate adaptation, and buffer management in a multiuser WLAN where each user transmits scalable video payload. It proposes a modified CSMA/CA channel access mechanism which takes into account the dynamic behaviours of the video variation, as well as the channel quality and transmission delay of each user.

LTE is an enhancement to the Universal Mobile Telecommunication System (UMTS) which will be introduced in 3rd Generation Partnership Project (3GPP) Release 8 [1]. LTE is marketed as the 4th generation cellular technology, aiming at delivering high speed broadband connection to stationary or mobile users. One important feature of 3GPP LTE system, differentiating it from previous generations of cellular systems, is the distributed network architecture, the key issue of which is to deploy femtocells to satisfy a variety of service requirements. These femtocell access points are low-cost, low-power, plug and play cellular base stations that are well-suited to comply with the rising wireless traffic demand. These small base stations are connected to the network operator through the end-user's broadband connection. Home Evolved Node-B (HeNB) can significantly improve the performance of current cellular networks by providing high data rates of the order of several Megabit Per Second (Mbps) and improved coverage inside home and offices [2].

The primary theoretical contributions of the thesis include important analytical results on the structure of optimal transmission policies. The

analytical results are original, and under the given assumptions the optimal transmission policies are monotone in channel states (see Chapters 2, 3). The monotone structural results are of utmost significance in optimization problems as they reduce the optimization dimensionality to the extreme. E.g., in Chapter 2, the transmission policy optimization problem is converted from a function optimization problem to a scalar optimization problem by utilizing the structural result on the Nash equilibrium transmission policy. The conversion allows the application of many gradient-based stochastic approximation algorithms to solve the optimal policy estimation problem [83].

Other contributions of the thesis are the derivation of equilibrium policy learning algorithms. Due to the structural results, the estimation algorithms are very efficient, the equilibrium learning algorithms derived in Chapters 2, 3, 4 can track slowly time-varying systems. The real-time learning algorithms are based on stochastic approximation and do not require a central controller, communication or exchange of information between users, or knowledge of system parameters such as the resource availability, number of active users, or the channel state probability distribution functions.

The derivation of structural results and estimation algorithms in this thesis involves novel use of several powerful mathematical tools, such as game theory, stochastic approximation, mechanism design; and clever application of a number of mathematical concepts such as Nash equilibrium, correlated equilibrium, static games, stochastic Markov games, and supermodularity. The mathematical tools will be outlined later in this chapter.

The remainder of this introductory chapter is organized as follows. The

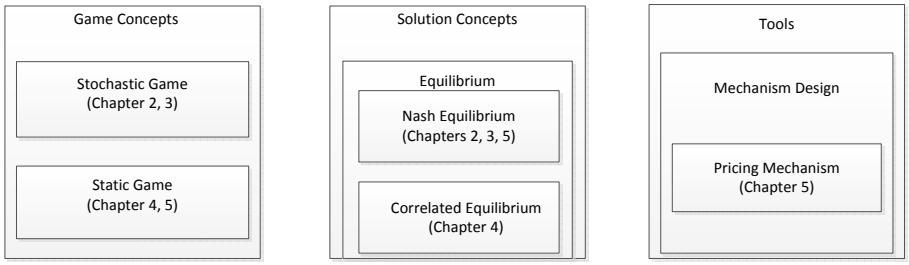

Figure 1.1: Summary of game theoretic topics being covered in this thesis.

next section contains a brief thesis overview. Section 1.2 introduces the major theoretical tools utilized in this thesis, with emphasis on their application in wireless communication systems. A summary of main contributions of this thesis is given in Section 1.3. Lastly, Section 1.4 outlines the thesis organization.

1.1 Thesis Overview

Fig. 1.1 summarizes the game theoretic topics/concepts being covered in this thesis. This thesis consists of three main parts. The first part, consisting of Chapter 2 and Chapter 3, is concerned with transmission adaptation in a cognitive radio system and a WLAN multimedia network using switching control game theoretical approach. The second part of the thesis (Chapter 4) investigates the correlated equilibrium resource allocation solution in an LTE system. The last part of this thesis (Chapter 5) aims to design a truth revealing pricing mechanism to improve the traditional opportunistic scheduling algorithm in a cognitive radio system.

In the transmission adaptation using switching control game theoretical

approach problem, the aim is to analyze the Nash equilibrium transmission policy that optimizes the infinite horizon expected total discounted utility function with the infinite horizon expected total discounted latency constraint. The optimization approach taken is a switching control stochastic game theoretic approach which formulates the Markovian block fading channels as a finite state Markov chain. This stochastic game theoretic framework assumes users are selfish, rational, and aim to maximize their own expected reward. The stochastic game theoretic framework leads to a particularly useful threshold structural result and hence allows users to learn their optimal transmission policies online in a completely decentralized fashion.

Compared to the centralized approach, the stochastic game theoretic formulation yields a very efficient way to optimize the transmission adaptation problem in a cognitive system or a WLAN. In particular, in the game theoretic setting, every user optimizes its transmission policy; then at equilibrium, no user will unilaterally deviate from its policy. Furthermore, instead of considering the myopic solutions, the users aim to optimize the infinite time horizon expected total discounted utility functions with latency constraints. By utilizing the system state transition probability information, it further optimizes the transmission adaptation problem. It will also be shown later in the thesis that the switching control game theoretic transmission adaptation Nash equilibrium solution has a particularly interesting and useful monotone structural result, which substantially simplifies the process of learning the optimal transmission policy. Specifically, it is shown that the Nash equilibrium transmission policy is a randomization of two pure policies, each of these two pure policies is nondecreasing (or nonincreasing) with

the buffer occupancy state. Therefore, to track the switching control game theoretic optimal transmission policy, each user needs to adaptively estimate only the thresholds and the randomization factor. Effectively, the function optimization problem is converted to the searching for few parameters, which can be solved easily and efficiently by many stochastic approximation algorithms. Taking into account the simplicity of the switching control game theoretic model and the implementation of switching control game theoretic solutions, the switching control game approach is practically very useful, especially for large scale wireless networks, where certain statistical properties (e.g., number of active users, channel state transition probabilities) may vary with time.

The second part of this thesis focuses on the correlated equilibrium resource allocation policies in an LTE system. In particular, it considers the spectrum allocation problem in an Orthogonal Frequency-Division Multiple Access (OFDMA) LTE downlink system which consists of a macrocell base station and multiple femtocell base stations. By incorporating cognitive radio capabilities into femtocell base stations, the HeNBs are formulated as cognitive base stations seeking to maximize the spectrum utility while minimizing the interferences to the Evolved Node-B (eNB) (primary base station) in a spectrum overlay LTE system. Given the resource occupancy of the eNB, the competition for the spectrum resources among HeNBs can be formulated in a game theoretic setting. However, instead of computing the Nash equilibrium policy of the formulated game, we seek to characterize and compute the *correlated equilibrium* policy set [7, 8]. We formulate the resource allocation problem among HeNBs in the downlink of an LTE

system under static game theoretic framework, where the correlated equilibrium solutions of the formulated static game is being investigated. A distributed Resource Block (RB) access algorithm is proposed to compute the correlated equilibrium RB allocation policy. The RB access algorithm has distributed implementations that permit them to be implemented in LTE systems.

The third part of the thesis uses mechanism design to improve the performance of the opportunistic scheduling in cognitive radio systems. The conventional opportunistic scheduling algorithm in cognitive radio networks does the scheduling among the secondary users based on the reported state values. However, such opportunistic scheduling algorithm can be challenged in a system where each secondary user belongs to a different independent agent and the users work in competitive way. In order to optimize his own utility, a selfish user can choose not to reveal his true information to the central scheduler. We proposed a pricing mechanism which combines the mechanism design with the opportunistic scheduling algorithm and ensures that each rational selfish user maximizes his own utility function, at the same time optimizing the overall system utility. The proposed pricing mechanism is based on the classic Vickrey-Clark-Groves (VCG) mechanism and had several desirable economic properties. A mechanism learning algorithm is then provided for users to learn the mechanism and to obtain the Nash equilibrium.

1.2 Analytical Tools

This thesis deploys several powerful mathematical tools that come from three research areas that have had many applications in electrical and computer engineering. In particular, we user game theory, which is a universally accepted tool in economics and social science and have been widely used in communications and networking [60]; stochastic approximation, which is a fundamental concept in stochastic control and optimization; and mechanism design, which is used to prevent malicious behaviours from game players. This section attempts to briefly overview these three areas.

1.2.1 Game Theory for Communications

Game theory is a branch of applied mathematics that provides models and tools for analyzing situations where multiple rational agents interact to achieve their goals. The classic book *The Theory of Games and Economic Behavior* by John von Neumann and Oskar Morgenstern published in 1944 [71] laid the first foundations for the research field of game theory, which now has numerous application in economics, social science, engineering and recently computer science. Recent textbooks on game theory include [13, 33, 74]. The treatment of game theory in [33] is very comprehensive and complete. Additionally, [74] and [13] are two very well-written graduate level textbooks in game theory. In communications engineering, game theory is often used for distributed resource management algorithms. They underlying motivation is that game theoretic solutions are naturally autonomous and robust. Examples of applications of game theory in wireless communications include transmission or power control [40, 65, 81], pricing [39], flow

and congestion control [4], and load balancing [34].

Static Game vs. Stochastic Game

Static game framework has been used to compute the Nash equilibrium power allocation policies in cognitive radio networks [38, 81, 94]. Different from a dynamic game where players make sequence of decisions, a static game is one in which all players make decisions one time simultaneously, without knowledge of the strategies of other players. [81] gives an example of a cognitive radio system where each user aims to maximize its information rate subject to the transmission power constraint. A distributed asynchronous iterative water-filling algorithm is used to compute the Nash equilibrium power allocation policy of such a system using static game theoretic approach.

Most games considered in wireless communication systems to date are static games. [73] and [30] are two of such examples which apply static game-theoretic analyses to address the resource allocation problem in cognitive radio networks. Stochastic dynamic game theory is an essential tool for cognitive radio systems as it is able to exploit the correlated channels in the analysis of decentralized behaviours of cognitive radios.

The concept of a stochastic game, first introduced by Lloyd Shapley in early 1950s, is a dynamic game played by one or more players. The elements of a stochastic game include system state set, action sets, transition probabilities and utility functions. It is an extension of the single player Markov Decision Process (MDP) to include the multiple players whose actions all impact the resulting payoffs and next state. A switching control

game [31, 68, 90] is a special type of stochastic dynamic game where the transition probability in any given state depends on only one player. It is known that the Nash equilibrium for such a game can be computed by solving a sequence of Markov decision processes.

In the first part of Chapter 3, we formulate the resource allocation problem among secondary base stations under static environment using static game theoretic framework. In the second part of Chapter 3, such resource allocation problem is considered under dynamic environment and stochastic game theoretic tools are being applied. Chapter 1 and Chapter 2 use switching control game to solve the transmission control problem in a cognitive radio system and a WLAN system, respectively.

Equilibrium and Learning

A major breakthrough in game theory was due to John Nash when he introduced the solution concept of Nash equilibrium in the early 1950's. A strategy combination constitutes a Nash equilibrium if each agent's strategy is optimal against other agent's strategies [74]. As a result, at a Nash equilibrium agents have no motivation to unilaterally deviate from their strategies. Nash equilibrium plays an essential role in analysis of conflict and cooperation in economics and social sciences.

However, Nash equilibrium suffers from limitations, such as non-uniqueness, loss of efficiency, non-guarantee of existence. In game theory, a correlated equilibrium is a solution concept which is more general than the Nash equilibrium [7, 8]. A correlated equilibrium is defined as follows. Each player in a game chooses his action according to his observation of the value of a

signal. A strategy assigns an action to every possible observation a player can make. If no player would deviate from the recommended strategy, the distribution is called a correlated equilibrium. Compared to Nash equilibria, correlated equilibria offer a number of conceptual and computational advantages, including the facts that new and sometimes more "fair" payoffs can be achieved, that correlated equilibria can be computed efficiently for games in standard normal form, and that correlated equilibria are the convergence notion for several natural learning algorithms. Furthermore, it has been argued that the correlated equilibria are the natural equilibrium concept consistent with the Bayesian perspective [8].

In this thesis, Chapter 2 and Chapter 3 look in to the Nash equilibrium solutions of the transmission policy, and Chapter 4 is seeking for the correlated equilibrium resource allocation policies.

While game theory and the equilibrium solution concepts are extremely useful and have numerous applications, the theory of learning in games is yet relatively less complete in the sense that convergence results are hard to establish and may require either restrictive, or hard to verify conditions. [32] contains essential existing results in the field of learning in games together with original work by the authors. The most studied methods of learning in games are fictitious play and best response dynamic [32]. These two methods have similar asymptotic behaviours and convergence to Nash equilibrium, in general, cannot be proved for either of the methods. The exceptions are potential games, where users have a common utility function; then both fictitious play and best response dynamic converge to a Nash equilibrium. Emergence of a Nash equilibrium can be proved for some 2-person games.

Other methods of learning in games include regret-minimizing algorithm. It has been shown in literature that if a game is repeated infinitely many times such that every player plays according to a regret-minimizing strategy, then the empirical frequencies of play converge to a set of correlated equilibrium [36, 52].

1.2.2 Stochastic Approximation Algorithms

Stochastic approximation/optimization algorithms are widely used in electrical computer engineering to recursively estimate the optimum of a function or its root, see for example [54, 83] for excellent expositions of this area. The first papers on the stochastic approximation methods are those by Robbins and Monro [79] and Kiefer and Wolfowitz [49] in the early 1950's. The well-known Least Mean Squares (LMS) adaptive filtering algorithm is a simple example of a stochastic approximation algorithm with a quadratic objective function. Stochastic approximation algorithm has been applied to reinforcement learning for stochastic control problems in [10], learning equilibria games [43, 44], as well as optimization and parametric identification problems (e.g., recursive maximum likelihood and recursive expectation maximization algorithms [50]).

In tracking applications, the step size of a stochastic approximation algorithm is chosen as a small constant. For such constant step size algorithm, one typically proves weak convergence of the iterates generated by the stochastic approximation algorithm. Weak convergence is a generalization of convergence in distribution to a function space. The weak convergence analysis of stochastic approximation algorithms with Markovian noise has

been pioneered by Kushner and co-workers, see [54] and references therein. It was demonstrated in the 1970's that the limiting behaviour of a stochastic approximation algorithm can be modeled as a deterministic ordinary-differential-equation (ODE). This is the basis of the so-called ODE method for convergence analysis of stochastic approximation algorithms.

In wireless networks, as there are usually underlying dynamics (e.g., a correlated fading channel) or factors that can only be measured in noise such as expected system throughput, stochastic approximation algorithms play an important role in optimization. In this thesis, gradient-based stochastic approximation algorithms are used for adaptive and distributed estimation of the optimal transmission policies in Chapters 2, 3.

1.2.3 Mechanism Design

The theory of mechanism design in economics/game theory deals with the design of multi-agent interactions of rational agents [63]. It is fundamental to Economics, as well as to the attempt to design protocols for non-cooperative computational wireless communication environment [15]. The importance of its application in the wireless communication world has increased dramatically due to the desire to design economic interactions in the wireless network setup. In a typical mechanism design setup, a center attempts to perform some task, or to arrive at some decision, based on information available to a set of agents. The major problem is that the agents might not supply the desired information to the center, and might cheat the center in order to increase their individual payoffs. The central problem is to design a mechanism that when the agents interact rationally through that

mechanism, the center will be able to obtain its objective. In order to do so, a mechanism has to be designed; typically, this mechanism determines actions to be taken by the center as a function of messages sent to its by the agents; the agents' strategies correspond to the messages they send to the center as a function of the private information they have. Such interaction is modeled by a Bayesian game. The aim is to design a mechanism that when the agents behave rationally (i.e., according to an equilibrium of the corresponding game), a desired behaviour is obtained.

A milestone in mechanism design is the VCG mechanism, which is a generalization of Vickrey's second price auction [89] proposed by Clark [20] and Groves [35]. The particular pricing policy of the VCG mechanism makes reporting true values the dominant strategy for all the players. Chapter 5 is an example where we model each user in a cognitive radio as a selfish player aiming to optimize his own utility and we try to find a mechanism which ensures efficient resource allocation within the network [42].

1.3 Main Contributions

This section presents a summary of methodologies, main structural results and learning algorithms developed in this thesis. As mentioned previously, the thesis consists of three parts. The first part of the thesis focuses on the Nash equilibrium solutions under switching control stochastic game setting. The second part of the thesis looks into the correlated equilibrium solutions of resource allocation policies in a static game environment. The last part of the thesis studies the mechanism design and its application in opportunistic scheduling. In each of the considered optimization problem, analytical

results and then numerical algorithms that exploit the analytical results are derived. An outline of the progression of ideas in the thesis is as follows.

Chapter 2 and Chapter 3 constitute the first main part of the thesis, where the transmission adaptation problem under a constrained switching control stochastic game framework is analyzed. Switching control game is a special type of dynamic game where the transition probability in any given state depends on only one player. It is known that the Nash equilibrium for such a game can be computed by solving a sequence of Markov decision processes.

Chapter 2 considers the rate adaptation problem in a Time Division Multiple Access (TDMA) cognitive radio system model. The interaction among secondary users is characterized as a competition for the spectrum hole and can naturally be formulated as a dynamic game. By modeling transmission channels as correlated Markovian sources, the transmission rate adaptation problem for each user can be formulated as a general-sum switching control Markovian dynamic game with a latency constraint. The structure of the Nash equilibrium transmission policy is studied and a stochastic algorithm is proposed for numerically solving the optimization problem. The main results in the chapter include the followings:

- We formulate the secondary user rate adaptation problem in a cognitive radio network as a constrained general-sum switching control Markovian dynamic game. The TDMA cognitive radio system has a pre-specified channel access rule that is typically a function of secondary user channel qualities and buffer occupancies. The Markovian

block fading channels are formulated as a finite state Markov chain, and each secondary user aims to optimize its own utility with a transmission delay constraint.

- It is shown that under reasonable assumptions, the Nash equilibrium policy of the transmission control game (formulated as a general-sum switching control game) is a randomized mixture of two pure policies. Each of the policies is monotone nondecreasing in the buffer occupancy state. We combine the latency constraint with the optimization objective using a Lagrange multiplier. The original constrained problem is then transformed into an unconstrained Markovian game. It is shown that the Nash equilibrium policy of such a game can be computed by using a value iteration algorithm.

- Based on the structural result on the Nash equilibrium policy, a stochastic approximation algorithm is proposed to compute the policy. The algorithm provides insight into the nature of the solution without brute force computation and is able to adapt the Nash equilibrium policy in real time to the non-stationary channel and user statistics. Numerical results of the stochastic approximation algorithm are shown to verify the performance of the algorithm.

Chapter 3 considers the transmission adaptation problem in an IEEE 802.11 WLAN which deploys the CSMA/CA. By exploiting the correlated time varying sample paths of the channels and buffers, the transmission rate adaptation problem can be formulated as a Markovian dynamic game. We also describe the details of the WLAN system under study as well as the

dynamic behaviour of video sources and the scalable rate-distortion models used to represent this behaviour. The main results presented in Chapter 3 are listed as follows.

- We consider a multiuser WLAN system where each user is equipped with a scalable video quality encoder delivering video bitstream that conforms with scalable video coding (SVC) which is the scalable extension of H.264/AVC. We address the scheduling, rate-control, and buffer management of such system, and formulate the problem as a constrained switching control stochastic dynamic game combined with rate-distortion modeling of the source. The video states and the block fading channel qualities of each user are formulated as a finite state Markovian chain, and each user aims to optimize its own utility under a transmission latency constraint when it is scheduled for transmission. We then formulate the rate adaptation in the WLAN multimedia system as a constrained switching control dynamic Markovian game.

- We propose a value iteration algorithm that obtains the Nash equilibrium transmission policy of the constrained general-sum switching control game. The Nash equilibrium policy of such constrained game is a randomization of two deterministic policies. The value iteration algorithm involves a sequence of dynamic programming problems with Lagrange multipliers. In each iteration of the value iteration algorithm, a Lagrange multiplier is used to combine the latency cost with the transmission reward for each user, the Lagrange multiplier is then updated based on the delay constraint. The algorithm converges to

the Nash equilibrium policy of the constrained general-sum switching control game.

- The main result of the chapter is the structural result on the Nash equilibrium policy. It is shown that under reasonable assumptions, the Nash equilibrium policy is monotone nonincreasing on the buffer state occupancy. This structural result on the transmission policy enables us to search for the Nash equilibrium policy in the policy space via a policy gradient algorithm. The structural result is especially useful in dealing with multiuser systems where each user has a large buffer size. Finally, numerical results of the proposed switching control Markovian dynamic game policy in transmitting scalable video are presented.

The second part of the thesis is presented in Chapter 4. The chapter considers the downlink spectrum allocation problem in a macrocell area containing multiple femtocells within an LTE system. By incorporating cognitive capabilities into femtocell base stations, the HeNBs can be formulated as cognitive base stations seeking to maximize the spectrum utility while minimizing the interference to primary base station (eNB). Given the resource occupancy of the eNB, the competition for the spectrum resources among HeNBs can be formulated in a game theoretic setting. However, instead of computing the Nash equilibrium policy of the formulated game, we seek to characterize and compute the correlated equilibrium. The main results in Chapter 4 include the following:

- We formulate the RB allocation among HeNBs in the downlink LTE system as a game. Given the RB usage of the eNB in an LTE macro-

cell, the cognitive HeNBs are modelled as selfish players competing to acquire the common frequency resources. This framework borrows the idea of cognitive radio systems [38, 67], where we formulate the cognitive HeNBs as secondary users and the eNB as the primary user in the shared spectrum system. A global utility function is defined to evaluate the overall LTE network performance. To achieve the optimal global utility value in a distributed set-up, we also define a local utility function for each cognitive HeNB. This local utility comprises of components that incorporate self-interest, fairness and power consumption of each HeNB.

- We define the correlated equilibrium of the formulated game. An RB access algorithm (Algorithm 5) is proposed which converges to the correlated equilibrium solution. This RB access algorithm is based on the regret matching algorithm [36, 51, 64] and it has a distributed nature as each cognitive HeNB does not require the information of other HeNBs. We also prove that this proposed algorithm will converge to the correlated equilibrium set of the formulated game. Numerical examples are shown to verify the results.

The third part of this thesis is presented in Chapter 5 where we applies the mechanism design to eliminate the malicious behaviour in cognitive radio networks with conventional opportunistic scheduling algorithm. In the conventional opportunistic scheduling algorithm, the central scheduler adopts opportunistic scheduling to schedule the users under an overall transmission power constraint. The scheduling is based on the reported states of

each user. When the secondary users belong to different independent selfish agents, the users may lie about their true state values in order to optimize their own utilities, sometimes in the price of reducing the overall system performance. The main results of Chapter 5 can be concluded as follows.

- We combine the mechanism with the opportunistic scheduling to eliminate the users from lying to ensure the optimality of the overall system performance. The pricing mechanism we use for each user is based on VCG mechanism and it maintains the same desirable economic properties as that of the VCG mechanism.

- A mechanism learning algorithm is proposed to implement the pricing mechanism.The convergence of this mechanism is shown numerically in the chapter.

1.4 Thesis Organization

The remainder of the thesis consists of five chapters. Chapter 2 and Chapter 3 constitute the first part of the thesis, which concerns transmission adaptation. Chapter 2 studies the transmission control among cognitive radios using switching control Markov game theoretical approach. Structural result on the Nash equilibrium transmission policy and stochastic approximation algorithm are presented. Chapter 3 considers scheduling, rate adaptation, and buffer management in a multiuser WLAN where each user transmits scalable video payload. Algorithms are proposed to compute the optimal transmission policy and simulation highlight the performance gain compared to the myopic policies. Chapter 4 is the second part of the thesis, it con-

cerns the correlated equilibrium solution is a game formulation. Chapter 4 investigates the resource allocation problem in an OFDMA LTE downlink system which consists of a macrocell base station and multiple femtocell base stations. The resource allocation problem is considered under static environment. The third part of the thesis (Chapter 5) presents the application of mechanism design to enforce the truth revealing feature in a conventional opportunistic scheduling algorithm under a cognitive radio network. Lastly, Chapter 6 contains the discussion of main results, conclusions and proposal of future research directions. A review of the work related to the specific optimization context is also given in each chapter.

Chapter 2

Transmission Control in Cognitive Radio as a Markovian Dynamic Game [1]

This chapter considers an uplink overlay Time Division Multiple Access (TDMA) cognitive radio network where multiple cognitive radios (secondary users) attempt to access a spectrum hole. We assume that each secondary user can access the channel according to a decentralized predefined access rule based on the channel quality and the transmission delay of each secondary user. By modeling secondary user block fading channel qualities as a finite state Markov chain, we formulate the transmission rate adaptation problem of each secondary user as a general-sum Markovian dynamic game

[1] This chapter is based on the following publication. J. W. Huang and V. Krishnamurthy, "Transmission Control in Cognitive Radio as a Markovian Dynamic Game - Structural Result on Randomized Threshold Policies," *IEEE Transactions on Communication*, vol. 58, no. 2, pp. 301-310, February 2010.

with a delay constraint. Conditions are given so that the Nash equilibrium transmission policy of each secondary user is a randomized mixture of pure threshold policies. Such threshold policies can be easily implemented. We then present a stochastic approximation algorithm that can adaptively estimate the Nash equilibrium policies and track such policies for non-stationary problems where the statistics of the channel and user parameters evolve with time.

2.1 Background

Resource management in cognitive radio networks involves user scheduling and transmission rate adaptation. Among the non-game theoretic approaches to such a problem, we discuss [62] and [95]. [62] deals with dynamic resource management based on quantized channel state information for multi-carrier cognitive radio networks. Using a stochastic dual approach, optimum dual prices are found to optimally allocate resources across users per channel realization without requiring knowledge of the channel distribution. Alternatively, [95] applies decentralized cognitive MAC protocols to address dynamic spectrum management for single-hop networks.

Most games considered in wireless communication systems to date are static games. [73] and [30] are two of such examples which apply static game-theoretic analyses to address the resource allocation problem in cognitive radio networks. Stochastic dynamic game theory is an essential tool for cognitive radio systems as it is able to exploit the correlated channels in the analysis of decentralized behaviours of cognitive radios. In this chapter, we are going to formulate the rate adaptation problem in a cognitive radio

network as a general-sum constrained switching control Markovian dynamic game.

This chapter uses a Time Division Multiple Access (TDMA) cognitive radio system model (as specified in the IEEE 802.16 standard [46]) that schedules one user per spectrum hole at each time slot according to a predefined decentralized scheduling policy. Therefore, the interaction among secondary users is characterized as a competition for the spectrum hole and can naturally be formulated as a dynamic game. In a TDMA cognitive radio system, the system state transition probability at each time slot only depends on the active user action. This feature fulfills the property of a special type of dynamic game which is called a switching control game [31, 68, 90], where the transition probability in any given state depends on only one player. It is known that the Nash equilibrium for such a game can be computed by solving a sequence of Markov decision processes. In this chapter, the transmission rate control problem is formulated as a *constrained* switching control Markovian game [3].

Lagrangian dynamic programming [48] has recently been applied to Markov decision processes in transmission scheduling. Here, we extend its application to dynamic games. In the problem formulation, we combine the latency constraint with the optimization objective using a Lagrange multiplier. The original constrained problem is then transformed into an unconstrained Markovian game. It is shown that the Nash equilibrium policy of such a game can be computed by using a value iteration algorithm.

One of the main goals of this thesis is to use mathematical tools (supermodularity, etc.) to characterize the nature of the optimal solution rather

than do brute force computation. To the best of our knowledge, the structural results proved in the thesis are some of the most general available in the literature. They hold under reasonable assumptions as described in Section 2.3.2.

2.2 Rate Adaptation Problem Formulation

This section describes the system model (Fig. 2.1). We consider an overlay TDMA cognitive radio system with K secondary users, each user tries to transmit delay-sensitive data with Quality of Service (QoS) requirement. At each time slot, only one user can access the channel according to a predefined decentralized access rule. The access rule will be described later in this section. By modelling the correlated block fading channel of each user as a Markov chain, the rate control problem can then be formulated as a constrained Markovian dynamic game. More specifically, under the predefined decentralized access rule, the problem presented is a special type of game, namely a switching control Markovian dynamic game. Note that the problem formulation of this chapter can be extended to a scenario with multiple spectrum holes. In which case, the rate adaptation problem at each spectrum hole follows the formulation of this chapter.

2.2.1 System Description and TDMA Access Rule

The channel quality state of user k at time n is denoted as h_k^n and it is assumed to belong to a finite set $\{0, 1, \ldots, Q_h\}$. The channel state can be obtained by quantizing a continuous valued channel model comprising of circularly symmetric complex Gaussian random variables that depend only

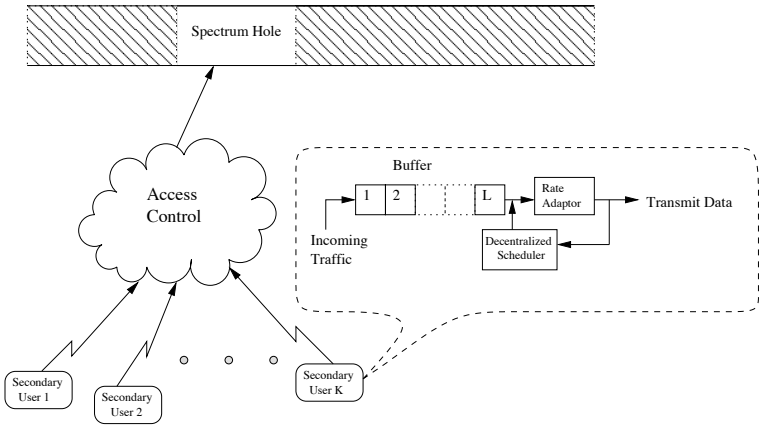

Figure 2.1: An overlay TDMA cognitive radio system where users access the spectrum hole following a predefined access rule. Each user is equipped with a size L buffer, a decentralized scheduler and a rate adaptor for transmission control.

on the previous time slot. The composition of channel states of all the K users can be written as $\mathbf{h}^n = \{h_1^n, \ldots, h_K^n\}$. Assuming that the channel state $\mathbf{h}^n \in \mathcal{H}, n = 1, 2, \ldots, N$ is block fading and each block length equals to one time slot, the channel state can be modeled using a finite states Markov chain model. The transition probability of the channel states from time n to $(n+1)$ can be denoted as $\mathbb{P}(\mathbf{h}^{n+1}|\mathbf{h}^n)$.

Let b_k^n denote the buffer occupancy state of user k at time n and it belongs to a finite set $b_k^n \in \{0, 1, \ldots, L\}$. The composition of the buffer states of all the K users can be denoted as $\mathbf{b}^n = \{b_1^n, \ldots, b_K^n\}$ and $\mathbf{b}^n$ is an element of the secondary user buffer state space $\mathcal{B}$.

New packets arrive at the buffer at each time slot and we denote the number of new incoming packets of the kth user at time n as f_k^n, $f_k^n \in \{0, 1, 2, \ldots, \infty\}$. The composition of the incoming traffic of all the K users

can be denoted as $\mathbf{f}^n = \{f_1^n, \ldots, f_K^n\}$, it is an element of the incoming traffic space $\mathcal{F}$. For simplicity, the incoming traffic is assumed to be independent and identically distributed (i.i.d.) in terms of time index n and user index k. The incoming traffic is not a part of the system state but it affects the buffer state evolution.

We use $\mathbf{s}_k^n = [h_k^n, b_k^n]$ to denote the state of user k at time n. The system state at time n can then be denoted as $\mathbf{s}^n = \{\mathbf{s}_1^n, \ldots, \mathbf{s}_K^n\}$. The finite system state space is denoted as $\mathcal{S}$, which comprises channel state $\mathcal{H}$ and secondary user buffer state $\mathcal{B}$. That is, $\mathcal{S} = \mathcal{H} \times \mathcal{B}$. Here $\times$ denotes a Cartesian product. Furthermore, $\mathcal{S}_k$ is used to indicate the state space where user k is scheduled for transmission. $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_K$ are disjoint subsets of $\mathcal{S}$ with the property of $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_K$.

This chapter adopts a TDMA cognitive radio system model (IEEE 802.16 [46]). A decentralized channel access algorithm can be constructed as follows: At the beginning of a time slot, user k attempts to access the channel after a certain time delay t_k^n. The time delay of user k can be specified via an opportunistic scheduling algorithm [28], such as

$$t_k^n = \frac{\gamma_k}{b_k^n h_k^n}. \tag{2.1}$$

Here γ_k is a user specified QoS parameter and $\gamma_k \in \{\gamma_p, \gamma_s\}$. If user k is a primary user, then, $\gamma_k = \gamma_p$, otherwise, $\gamma_k = \gamma_s$. By setting $\gamma_p << \gamma_s$, the network does not allow the transmission of secondary users with the presence of primary users. As soon as a user successfully access a channel, the remaining users detect the channel occupancy and stop their attempt to access. We use $k^{*(n)}$ to denote the index of the first user which successfully

accesses the spectrum hole. If there are multiple users with the same minimum waiting time, $k^{*(n)}$ is randomly chosen from these users with equal probability.

2.2.2 Action and Costs

If the kth user is scheduled for transmission at the nth time slot, its action $a_k^n \in \{0, 1, \ldots, A\}$ represents the bits/symbol rate of the transmission. Assuming the system uses an uncoded M-ary Quadrature Amplitude Modulation (QAM), bits/symbol rate determines the modulation scheme, that is, $M = 2^{a_k^n}$.

Transmission cost: When user k is scheduled for transmission at time n, that is, $\mathbf{s}^n \in \mathcal{S}_k$, the cost function of user k depends only on a_k^n, as all the other users are inactive. In this chapter, we choose the transmission cost of the kth user $c_k(\mathbf{s}^n, a_k^n)$ to be its transmission Bit Error Rate (BER). The costs of all the users in the system can be specified as:

$$c_k(\mathbf{s}^n, a_k^n) \geq 0, \quad c_{i,i\neq k}(\mathbf{s}^n, a_k^n) = 0. \tag{2.2}$$

For notational convenience, in the following sections we will drop the subscript k, by defining

$$c(\mathbf{s}^n, a_k^n) := c_k(\mathbf{s}^n, a_k^n).$$

Holding cost: Each user has an instantaneous QoS constraint denoted as $d_i(\mathbf{s}^n, a_k^n)$, $i = 1, \ldots, K$. In this chapter, we choose the QoS constraint to be the delay (latency constraint) and $d_i(\mathbf{s}^n, a_k^n)$ is a function of the buffer state

b_i^n.

2.2.3 Switching Control Game and Transition Probabilities

With the above setup, the decentralized transmission control problem in a Markovian block fading channel cognitive radio system can now be formulated as a switching control game. In such a game [31], the transition probabilities depend only on the action of the kth user when $\mathbf{s}^n \in \mathcal{S}_k$ (2.3). This feature enables us to solve such a game by a finite sequence of Markov decision processes.

$$\mathbb{P}(\mathbf{s}^{n+1}|\mathbf{s}^n, a_1, a_2, \ldots, a_K) = \begin{cases} \mathbb{P}(\mathbf{s}^{n+1}|\mathbf{s}^n, a_1), & \text{if} \quad \mathbf{s}^n \in \mathcal{S}_1 \\ \mathbb{P}(\mathbf{s}^{n+1}|\mathbf{s}^n, a_2), & \text{if} \quad \mathbf{s}^n \in \mathcal{S}_2 \\ \ldots & \\ \mathbb{P}(\mathbf{s}^{n+1}|\mathbf{s}^n, a_K), & \text{if} \quad \mathbf{s}^n \in \mathcal{S}_K \end{cases}. \tag{2.3}$$

If $\mathbf{s}^n \in \mathcal{S}_k$, the transition probability between the current system state $\mathbf{s}^n = [\mathbf{h}^n, \mathbf{b}^n]$ and the next state $\mathbf{s}^{n+1} = [\mathbf{h}^{n+1}, \mathbf{b}^{n+1}]$ can be specified as:

$$\mathbb{P}(\mathbf{s}^{n+1}|\mathbf{s}^n, a_k^n) = \prod_{i=1}^{K} \mathbb{P}(h_i^{n+1}|h_i^n) \cdot \prod_{i=1, i\neq k}^{K} \mathbb{P}(b_i^{n+1}|b_i^n) \cdot \mathbb{P}(b_k^{n+1}|b_k^n, a_k^n).$$

In state $\mathbf{s}^n \in \mathcal{S}_k$, the buffer occupancy of user k evolves according to Lindley's equation [24],

$$b_k^{n+1} = \min\left([b_k^n - a_k^n]^+ + f_k^n, L\right). \tag{2.4}$$

Here, $[\cdot]^+$ is defined as: $[x]^+ = x$ if $x \geq 0$, otherwise, $[x]^+ = 0$. The buffer state transition probability of user k is a function of its action and its

incoming traffic distribution, which can be specified as

$$\mathbb{P}(b_k^{n+1}|b_k^n, a_k^n) = \begin{cases} \mathbb{P}(f_k^n = b_k^{n+1} - [b_k^n - a_k^n]^+), & \text{if} \quad b_k^{n+1} < L \\ \sum_{x=L-[b_k^n - a_k^n]^+}^{\infty} \mathbb{P}(f_k^n = x), & \text{if} \quad b_k^{n+1} = L \end{cases}. \tag{2.5}$$

The buffer state of user i ($i \neq k$) evolves according to the following rule:

$$b_i^{n+1} = \min(b_i^n + f_i^n, L).$$

The buffer state transition probability of user i ($i \neq k$) is a function of its incoming traffic, which is

$$\mathbb{P}(b_i^{n+1}|b_i^n) = \begin{cases} \mathbb{P}(f_i^n = b_i^{n+1} - b_i^n), & \text{if} \quad b_i^{n+1} < L \\ \sum_{x=L-b_i^n}^{\infty} \mathbb{P}(f_i^n = x), & \text{if} \quad b_i^{n+1} = L \end{cases}. \tag{2.6}$$

2.2.4 Switching Controlled Markovian Game Formulation

We use π_i ($i = 1, 2, \ldots, K$) to denote the transmission policy vector of the ith user. With a slight abuse of notation, $\pi_i(\mathbf{s})$ is used to denote the transmission policy of user i in state $\mathbf{s}$ and it is a component of π_i. $\pi_i(\mathbf{s})$ lives in the same space as the action a_i of the ith user. Assume at time instant n user k is scheduled for transmission according to the system access rule (2.1), the infinite horizon expected total discounted cost[2] of any ith ($i = 1, 2, \ldots, K$)

[2]There are two cost criteria commonly used for MDPs: expected average cost criterion and expected total discounted reward criterion. We choose the discounted cost criterion because it is mathematically simpler. The average cost criterion has several technicalities involved with the existence of stationary optimal policies. In addition, if the stationary policy exists, as the discount factor $\beta \to 1$, the policy obtained under the discounted cost criterion is the same as that under the average cost criterion.

user under transmission policy π_i can be written as:

$$C_i(\pi_i) \quad = \quad \mathbb{E}_{\pi_i}\Big[\sum_{n=1}^{\infty}\beta^{n-1}\cdot c_i(\mathbf{s}^n, a_k^n)\Big], \tag{2.7}$$

with $0 \leq \beta < 1$ is a discount factor which models the fact that the future rewards worth less than the current reward. This discount factor can be justified as an inflation rate, as an otherwise unmodeled probability that the simulation ends at each time-step. The expectation of the above function is taken over the system state $\mathbf{s}^n$ which evolves over time. Denoting the holding cost of user i at the nth time slot as $d_i(\mathbf{s}^n, a_k^n)$, the infinite horizon expected total discounted latency constraint can be written as

$$D_i(\pi_i) = \mathbb{E}_{\pi_i}\Big[\sum_{n=1}^{\infty}\beta^{n-1}\cdot d_i(\mathbf{s}^n, a_k^n)\Big] \leq \widetilde{D}_i, \tag{2.8}$$

where $\widetilde{D}_i$ is a system specified parameter. In order to ensure the validity of the latency constraint, $\widetilde{D}_i$ is chosen so that the set of policies that satisfy such a constraint is non-empty. This is the feasibility assumption A 2.3.1 which will be discussed more specifically in Section 2.3.2.

Equations (2.3, 2.7, 2.8) define a constrained switching control Markovian game. It is assumed that all the users are transmitting delay sensitive traffic with QoS requirement. Our goal is to compute a Nash equilibrium[3] policy $\pi_i^*, i = 1, \ldots, K$ (which is not necessarily unique) that minimizes the discounted transmission cost (2.7) subject to the latency constraint (2.8). The following result shows that a Markovian switching control game can be solved using a sequence of Markov decision processes. For the proof, see

[3]A Nash equilibrium [31] is a set of policies, one for each player, such that no player has incentive to unilaterally change its action.

[31, 68].

Result: [31, Chapter 3.2] The constrained switching control Markovian game (2.7, 2.8) can be solved by a finite sequence of MDPs (as described in Algorithm 1). At each step, the algorithm iteratively updates the transmission policy π_i^n of user i given the transmission policies of the remaining users. The optimization problem at each iteration is:

$$\pi_i^{*(n)} = \{\pi_i^n : \min_{\pi_i} C_i^n(\pi_i) \text{ s.t. } D_i^n(\pi_i) \leq \widetilde{D}_i, i = 1, \ldots, K\}. \tag{2.9}$$

2.3 Randomized Threshold Nash Equilibrium for Markovian Dynamic Game

This section presents a value iteration algorithm to compute the Nash equilibrium policy of the general-sum Markovian dynamic switching control game. We then present a structural result on the Nash equilibrium policy. Finally, we propose a computationally efficient stochastic approximation algorithm to compute the Nash equilibrium.

2.3.1 Value Iteration Algorithm

A value iteration algorithm is presented in [31] to compute the Nash equilibrium of an unconstrained general-sum Markovian dynamic switching control game. By using the Lagrangian dynamic programming, we combine the optimization objective (2.7) with the latency constraint (2.8) via Lagrange multipliers λ_i, $i = 1, 2, \ldots, K$. We then compute the Nash equilibrium of the resulting unconstrained switching control Markovian dynamic game by

the value iteration algorithm described in Algorithm 1. In Algorithm 1, $\mathbf{V}_i$ denotes the value vector of user i, each of its components $v_i(\mathbf{s})$ denotes the discounted weighted sum of the transmission cost and the latency cost of user i in state $\mathbf{s}$ (the weighting factor is the Lagrange multiplier λ_i).

Algorithm 1 Value Iteration Algorithm

Step 1: Set $n = 0$; Initialize $\mathbf{V}_1^0, \ldots, \mathbf{V}_K^0$, $\lambda_1, \lambda_2, \ldots, \lambda_K$.
Step 2: Update the transmission policy and value vector of each user:
for $k = 1 : K$ **do**
for each $\mathbf{s} \in \mathcal{S}_k$,
$\pi_k^n(\mathbf{s}) = \arg\min_{a_k} \left\{ c(\mathbf{s}, a_k) + \lambda_k \cdot d_k(\mathbf{s}, a_k) + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}') \right\}$;
$v_k^{n+1}(\mathbf{s}) = c(\mathbf{s}, \pi_k^n(\mathbf{s})) + \lambda_k \cdot d_k(\mathbf{s}, \pi_k^n(\mathbf{s})) + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|s, \pi_k^n(\mathbf{s})) v_k^n(\mathbf{s}')$;
$v_i^{n+1}(\mathbf{s}) = \lambda_i \cdot d_i(\mathbf{s}, a_k) + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, \pi_k^n(\mathbf{s})) v_i^n(\mathbf{s}')$,
(where $i = 1, 2, \ldots, K,\ i \neq k$);
end for
Step 3: If $\mathbf{V}_k^{n+1} < \mathbf{V}_k^n, k = 1, \ldots, K$, set $n = n + 1$, and return to Step 2; otherwise, π_k^n for $k = 1, 2, \ldots, K$ is the optimal policy for user k.

With slight abuse of notation, we denote the iteration index as n. In Step 1 of Algorithm 1, we first set $n = 0$, then initialize the value vector and the Lagrange multiplier of each user $\mathbf{V}_k^0$, λ_k for $k = 1, 2, \ldots, K$. The Lagrange multipliers $\lambda_1, \lambda_2, \ldots, \lambda_K$ are chosen to be reasonable positive constants. Algorithm 1 is specified for fixed values of Lagrange multipliers. The optimal values of these Lagrange multipliers can be computed by applying the stochastic approximation algorithm which will be introduced in Section 2.3.3. Step 2 solves a kth user controlled game. It computes the optimal strategy for user k (if $\mathbf{s} \in \mathcal{S}_k$), with the strategies of the remaining users fixed. The value vectors of all the users are then computed. The algorithm terminates when the value vector of each user converges as $\mathbf{V}_k^{n+1} = \mathbf{V}_k^n$ $(k = 1, 2, \ldots, K)$; otherwise, it returns to Step 2.

Theorem 2.3.1 *Under the feasibility assumption A 2.3.1 (specified in Section 2.3.2) and given the Lagrange multipliers λ_k, $k = 1, 2, \ldots, K$, the value iteration algorithm (Algorithm 1) converges to the Nash equilibrium of the constrained switching control general-sum Markovian game.* □

Please refer to [31, Chapter 6.3] for the proof of Theorem 2.3.1. The intuition behind the proof is as follows: The value vector $\mathbf{V}_k^n$ $(k = 1, 2, \ldots, K)$ is nonincreasing on the iteration index n in the value iteration algorithm. There are only a finite number of strategies available for the optimal policy π_k^* for $k = 1, 2, \ldots, K$. It can be concluded that the algorithm converges in a finite number of iterations. Thus, the Nash equilibrium of the switching control dynamic game under the discounted reward criterion can be obtained by the value iteration algorithm.

Remark: The primary purpose of the value iteration algorithm is to prove the structural results on the Nash equilibrium policy. The value iteration algorithm (Algorithm 1) is not designed to be implement in a practical system because at each iteration of the algorithm, a user k is required to know the channel states of all the other users and the state transition probability matrix (2.3).

2.3.2 Structural Result on Randomized Threshold Policy

We aim to characterize the structure of the Nash equilibrium policy in this subsection. First, we list three assumptions. Based on these three assumptions, our main structural result is introduced in Theorem 2.3.2.

- *A 2.3.1:* The set of policies that satisfy the system delay constraint (2.8) is non-empty.

First, we define the holding cost of any user i ($i = 1, 2, \ldots, K$) when the state $\mathbf{s} \in \mathcal{S}_k$:

$$d_i(\mathbf{s}, a_k) = \frac{b_i}{\overline{f}}. \tag{2.10}$$

Here $\overline{f}$ is the average number of incoming packets, and is a system specified parameter. Assumption A 2.3.1 holds if there exists an action such that $a_k > \overline{f}$.

- *A 2.3.2:* Transmission cost $c(\mathbf{s}, a_k)$ and holding cost $d_k(\mathbf{s}, a_k)$ are submodular[4] functions of b_k, a_k given channel quality h_k of the current user, and are independent of the incoming traffic f_k. $c(\mathbf{s}, a_k)$ and $d_k(\mathbf{s}, a_k)$ are also nondecreasing functions of b_k for any h_k and a_k.

  The system transmission cost (2.2) is chosen to be the transmission BER. Assuming the channel states are quantized by system parameterized quantization thresholds $\Gamma(h)_1, \Gamma(h)_2, \ldots,$ the system transmission cost $BER(\gamma, a_k)$ is a function of the random channel gain $\gamma \in [\Gamma(h)_i, \Gamma(h)_i)$. The transmission cost is [19],

$$BER_k(\gamma, a_k) = 0.2 \times \exp\left[\frac{-1.6\gamma}{(2^{a_k} - 1)}\right], \tag{2.11}$$

$$BER_k^i(h_k, a_k) = \frac{\int_{\Gamma(h)_{i-1}}^{\Gamma(h)_i} BER(\gamma, a_k) g(\gamma) d\gamma}{\int_{\Gamma(h)_{i-1}}^{\Gamma(h)_i} g(\gamma) d\gamma}. \tag{2.12}$$

We assume an uncoded M-ary quadrature modulation (QAM). a_k is the number of bits per symbol and different values of a_k correspond to different modulation scheme. $g(\gamma)$ denotes the probability distri-

[4] A function $f : \mathcal{A} \times \mathcal{B} \times \mathcal{C} \to \mathcal{R}$ is said to be submodular in (a, b) for any fixed $c \in \mathcal{C}$ if $f(a', b'; c) - f(a, b'; c) \leq f(a', b; c) - f(a, b; c)$ holds for all $a' \geq a$ and $b' \geq b$.

bution of the Signal-to-Noise Ratio (SNR) γ. For a channel state h_k belongs to quantization region $[\Gamma(h)_{i-1}, \Gamma(h)_i)$, the expectation of $BER_k^i(h_k, a_k)$ is taken over γ for $\gamma \in [\Gamma(h)_{i-1}, \Gamma(h)_i)$.

From (2.11) and (2.12), it can be seen that the average transmission cost is independent of the buffer occupancy b_k. Furthermore, it is clear from (2.10) that the holding cost function $d_k(\mathbf{s}, a_k)$ is nondecreasing on b_k and independent of a_k. Thus, assumption A 2.3.2 holds.

- *A 2.3.3:* $\sum_{b_k'=l}^{L} \mathbb{P}(b_k'|b_k, a_k)$ is a submodular function of b_k, a_k and is nondecreasing on b_k for any l and a_k .

  As stated in (2.4), the buffer occupancy state evolves according to Lindley's recursion equation. Given the current state buffer occupancy and action, the transition probability depends on the incoming traffic distribution, which is shown in (2.5). Thus, the buffer state transition probability can be rewritten as $\mathbb{P}(b_k'|b_k, a_k) = \mathbb{P}(b_k'|(b_k - a_k))$. Assume the incoming traffic is evenly distributed in $0, 1, \ldots, L$, which can be mathematically written as: $\mathbb{P}(f_k < 0 \; or \; f_k > L) = 0$ and $\mathbb{P}(0 \leq f_k \leq L) = \frac{1}{L+1}$. Then the buffer state transition probability is

$$\mathbb{P}(b_k'|b_k, a_k) = \begin{cases} \frac{1}{L+1}, & \text{if} \quad b_k' < L \\ \frac{1+[b_k - a_k]^+}{L+1}, & \text{if} \quad b_k' = L \end{cases} . \tag{2.13}$$

  Therefore, the buffer occupancy state transition probability is independent of b_k and a_k when $b_k' < L$, and is first order stochastically nondecreasing on $b_k - a_k$ when $b_k' = L$. This result verifies $\sum_{b_k'=l}^{L} \mathbb{P}(b_k'|b_k, a_k)$ is nondecreasing on b_k in A 2.3.3. According to (2.13) we can see that $\mathbb{P}(b_k' = L|b_k, a_k)$ is submodular in b_k, a_k. This verifies A 2.3.3.

The following theorem is our main result. It states that the Nash equilibrium policy of a constrained dynamic switching control game is a randomized mixture of two pure monotone policies.

Theorem 2.3.2 *Consider the rate adaptation problem in a cognitive radio system as described in Section 2.2. Assume assumption A 2.3.1-A 2.3.3 hold. Then the Nash equilibrium policy of any user k ($k = 1, 2, \ldots, K$) π_k^* is a randomized mixture of two pure policies: π_k^1 and π_k^2. Each of these two pure policies is nondecreasing on the buffer occupancy state b_k.* □

The proof of Theorem 2.3.2 is given in Appendix A.

2.3.3 Stochastic Approximation Algorithm

The structural result stated in Theorem 2.3.2 can be exploited to compute the Nash equilibrium policy. We now present a *stochastic approximation algorithm* that exploits this structure.

Suppose the action set of each user contains two actions, that is $a_k \in \{0, 1\}$. The two actions can be chosen as *No Transmission* ($a_k = 0$) and *Transmission* ($a_k = 1$). Using the structural result on the Nash equilibrium policy as shown in Theorem 2.3.2, the Nash equilibrium $\pi_k^*(\mathbf{s})$ is a randomized policy parameterized by three parameters:

$$\pi_k^*(\mathbf{s}) = \begin{cases} 0, & \text{if} \quad 0 \le b_k < b_l(\mathbf{s}) \\ p, & \text{if} \quad b_l(\mathbf{s}) \le b_k < b_h(\mathbf{s}) \\ 1, & \text{if} \quad b_h(\mathbf{s}) \le b_k \end{cases} . \tag{2.14}$$

The parameter $p \in [0,1]$ is a randomization probability. The parameters $b_l(\mathbf{s})$ and $b_h(\mathbf{s})$ are the lower and higher buffer state thresholds, respectively. The search for the Nash equilibrium policy over the intractably large buffer space reduces to estimating the three parameters, $b_l(\mathbf{s})$, $b_h(\mathbf{s})$ and p. The above structural result can be extended to cognitive radio systems with larger action sets. For example, when the action set of each secondary user is $a_k = \{0,1,2\}$, the Nash equilibrium policy can be parameterized by 6 parameters (4 buffer state thresholds and 2 randomization factors). In the stochastic approximation algorithm, we compute the continuous optimal values of $b_l(\mathbf{s}) \in [0, L]$ and $b_h(\mathbf{s}) \in [b_l(\mathbf{s}), L]$ $(b_h(\mathbf{s}) \geq b_l(\mathbf{s}))$, then round them off to the nearest discrete values. This is a relaxation of the original discrete stochastic optimization problem as the buffer states in the problem setup are discrete. For convenience, let θ_k denote the vector of all the parameters that will be estimated for user k $(k = 1, 2, \ldots, K)$. The composition of the parameter vectors of all the K users can be denoted as $\Theta = \{\theta_1, \theta_2, \ldots, \theta_K\}$.

A primal and dual gradient projection method requires the Lagrangian to be locally convex at the optimum in order to ensure the convergence. In our problem, the optimization objective C_k (2.7) and latency constraint D_k (2.8) $(k = 1, 2, \ldots, K)$ are not convex over Θ since $\Theta = \{\theta_1, \theta_2, \ldots, \theta_K\}$ and θ_k is defined on the estimation parameters of the transmission policy of user k (e.g., $b_l(\mathbf{s})$, $b_h(\mathbf{s})$ and p). However, the augmented Lagrangian method can "convexify" the problem by adding a penalty term to the objective [53]. Based on which, the stochastic approximation algorithm is introduced in Algorithm 2. The gradients of C_k (2.7) and D_k (2.8) are estimated by using the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm [83]. The essential feature of SPSA is the underlying gradient ap-

proximation, which requires only two objective function measurements per iteration, regardless of the dimension of the optimization problem. These two measurements are made by simultaneously varying, in a properly random fashion, all of the variables in the problem.

Algorithm 2 Stochastic Approximation Algorithm

1: **Initialization**: $\Theta^{(0)}$, Λ^0; $n = 0$; $\rho = 4$;
2: Initialize constant perturbation step size μ and gradient step size α;
3: **Main Iteration**
4: **if** $\mathbf{s}^n \in \mathcal{S}_k$ **then**
5: $m_k = |\theta_k^n|$;
6: Generate $\Delta^n = [\Delta_1^n, \Delta_2^n, \ldots, \Delta_{m_k}^n]^T$; Δ_i^n are Bernoulli random variables with $p = \frac{1}{2}$.
7: $\theta_{k+}^n = \theta_k^n + \mu \times \Delta^n$;
8: $\theta_{k-}^n = \theta_k^n - \mu \times \Delta^n$;
9: ΔC_k^n
10: $= \frac{c(\mathbf{s}^n, \theta_{k+}^n) - c(\mathbf{s}^n, \theta_{k-}^n)}{2\mu} [(\Delta_1^n)^{-1}, (\Delta_2^n)^{-1}, \ldots, (\Delta_{m_k}^n)^{-1}]^T$;
11: ΔD_k^n
12: $= \frac{d(\mathbf{s}^n, \theta_{k+}^n) - d(\mathbf{s}^n, \theta_{k-}^n)}{2\mu} [(\Delta_1^n)^{-1}, (\Delta_2^n)^{-1}, \ldots, (\Delta_{m_k}^n)^{-1}]^T$;
13: $\theta_k^{n+1} = \theta_k^n - \alpha \times \left(\Delta C_k^n + \Delta D_k^n \cdot \max \left[0, \lambda_k^n + \rho \cdot \left(D(\mathbf{s}^n, \theta_k^n) - \widetilde{D_k} \right) \right] \right)$;
13: $\lambda_k^{n+1} = \max \left[(1 - \frac{\alpha}{\rho} \cdot \lambda_k^n), \lambda_k^n + \alpha \cdot \left(D(\mathbf{s}^n, \theta_k^n) - \widetilde{D_k} \right) \right]$;
14: **end if**
15: The parameters of other users remain unchanged;
16: $n = n + 1$;
17: The iteration terminates when the values of the parameters Θ^n converge; else return back to Step 3.

In Algorithm 2, μ and α are used to denote the constant perturbation step size and constant gradient step size, respectively. The composition of the Lagrange multipliers of all the users can be written as $\Lambda^n = \{\lambda_1^n, \lambda_2^n, \ldots, \lambda_K^n\}$. In the main body of the algorithm, the SPSA algorithm is used to estimate the gradients of C_k and D_k, then the parameters are updated iteratively. Specifically, when user k is scheduled for transmission at time slot n, pa-

rameters θ_k^n and the Lagrange multiplier λ_k^n can be updated after introducing a random perturbation vector Δ^n. Meanwhile, the parameters of the other users remain unchanged. ΔC_k^n and ΔD_k^n are used to denote the gradient estimators of C_k^n and D_k^n, respectively. More specifically, they can be written as:

$$\Delta C_k^n = g_{C_k}^n(\theta_k^n) + o_{C_k}^n, \quad g_{C_k}^n(\theta_k^n) = \frac{\partial C_k^n(\theta_k^n)}{\partial \theta_k^n}, \tag{2.15}$$

$$\Delta D_k^n = g_{D_k}^n(\theta_k^n) + o_{D_k}^n, \quad g_{D_k}^n(\theta_k^n) = \frac{\partial D_k^n(\theta_k^n)}{\partial \theta_k^n}, \tag{2.16}$$

with $o_{C_k}^n$ and $o_{D_k}^n$ denoting the noise of the gradient estimations.

Theorem 2.3.3 *Consider the sequence of estimates* $\{\Theta^n(\alpha,\mu), \Lambda^n(\alpha,\mu)\}$ *generated by the stochastic approximation algorithm. Define associated piecewise constant interpolated continuous-time processes as follows:*

$$\Theta^t(\alpha,\mu) = \Theta^n(\alpha,\mu), \quad \textit{for} \quad t \in [n\alpha, (n+1)\alpha), \tag{2.17}$$

$$\Lambda^t(\alpha,\mu) = \Lambda^n(\alpha,\mu) \quad \textit{for} \quad t \in [n\alpha, (n+1)\alpha). \tag{2.18}$$

Assume $o_{C_k}^n$ *and* $o_{D_k}^n$ *are uncorrelated noise processes with zero means and finite variances and the gradients* $g_{C_k}^n(\theta_k^n)$ *and* $g_{D_k}^n(\theta_k^n)$ *are uniformly bounded for sufficiently large* ρ. *Then, as* $\alpha \to 0$, $\frac{\alpha}{\mu^2} \to 0$ *and* $t \to \infty$, $\{\Theta^t(\alpha), \Lambda^t(\alpha)\}$ *converge in distribution to the Kuhn Tucker (KT) pair of (2.9) [54, Theorem 8.3.1], which is specified in (2.19).* □

According to [54, Theorem 8.3.1], Algorithm 2 converges in distribution under the following conditions.

- $\frac{\alpha}{\mu^2} \to 0$, $\alpha \to 0$. $g_{C_k}^n(\theta_k^n)$ and $g_{D_k}^n(\theta_k^n)$ are uniformly bounded.
- The uncorrelated noise processes $o_{C_k}^n$ and $o_{D_k}^n$ have zero means.

- ΔC_k^n and ΔD_k^n have finite variances. This condition holds because $o_{C_k}^n$ and $o_{D_k}^n$ have finite variances.

- There are functions $g_{C_k}^n(\cdot)$ and $g_{D_k}^n(\cdot)$ that are continuous uniformly in n and random variables $o_{C_k}^n$ and $o_{D_k}^n$ such that

$$\begin{aligned} \mathbb{E}_n[C_k^n(\theta_{k+}^n) - C_k^n(\theta_{k-}^n)]/2\mu &= g_{C_k}^n(\theta_k^n) + o_{C_k}^n, \\ \mathbb{E}_n[D_k^n(\theta_{k+}^n) - D_k^n(\theta_{k-}^n)]/2\mu &= g_{D_k}^n(\theta_k^n) + o_{D_k}^n. \end{aligned}$$

  According to the definition of ΔC_k^n (2.15) and ΔD_k^n (2.16), this condition holds.

The optimal policies of (2.9) satisfying the KT condition can be defined as follows. θ_i^* belongs to the KT set when

$$KT = \{\theta_i^* : \exists \lambda_i > 0, \text{ such that } \nabla_{\theta_i} C_i + \nabla_{\theta_i} \lambda_i (D_i - \widetilde{D}_i) = 0, i = 1, \ldots, K\}, \quad (2.19)$$

where C_i and D_i are the optimization objective (2.7) and discounted time delay (2.8), respectively. Moreover, θ_i^* satisfies the second order sufficiency conditions: $\nabla_{\theta_i}^2 C_i + \nabla_{\theta_i}^2 (D_i - \widetilde{D}_i) \geq 0$ is positive definite for all the i, and $(D_i - \widetilde{D}_i) = 0$, $\lambda_i > 0$, $i = 1, \ldots, K$.

Note that in the stochastic approximation algorithm, we first compute the continuous values of $b_l(\mathbf{s})$ and $b_h(\mathbf{s})$, then round them off to the nearest discrete values. This relaxation leads to the continuous value of θ_i during calculation, thus, (2.19) is differentiable on θ_i.

Remark: At each iteration of the stochastic approximation algorithm, user k is only required to know its own transmission cost and holding cost. The transmission cost and holding cost of user k are functions of its own

channel state and buffer state (Section 2.3.2). Thus, the stochastic approximation algorithm is distributed and implementable in a practical system.

2.4 Numerical Examples

This section presents numerical results on the Nash equilibrium transmission policy. For convenience, all the secondary users in the simulation setup are assumed to be identical. The simulation results shown here are of the transmission policies of the first user. In the system model, each user has a size 10 buffer, and the channel quality measurements are quantized into two different states, namely $\{1, 2\}$. In the system configuration, the transmission costs, the holding costs and buffer transition probability matrices are chosen to ensure assumption A 2.3.2-A 2.3.3 (specified in Section 2.3.2). The channel transition probability matrices are generated randomly.

In the system models used in Fig. 2.2 and Fig. 2.3, each user has three different action choices. The system uses an uncoded M-QAM modulation scheme, and different actions are followed by different modulation modes, leading to different transmission rates. The three different actions are $a_k = 1, 2, 3$ and consequently, the modulation schemes are 2-QAM, 4-QAM and 8-QAM. When the Nash equilibrium policy of a user equals to 0 in a state, it means that user is not scheduled for transmission in that state. The transmission policies shown in Fig. 2.2 and Fig. 2.3 are obtained by fixing the Lagrange multiplier to be $\lambda_1 = \lambda_2 = 1.3$ and then apply the value iteration algorithm (Algorithm 1).

a) General-Sum Constrained Game: Structure of the Nash: Fig. 2.2 shows the structured optimal transmission policy of a two-user general-sum

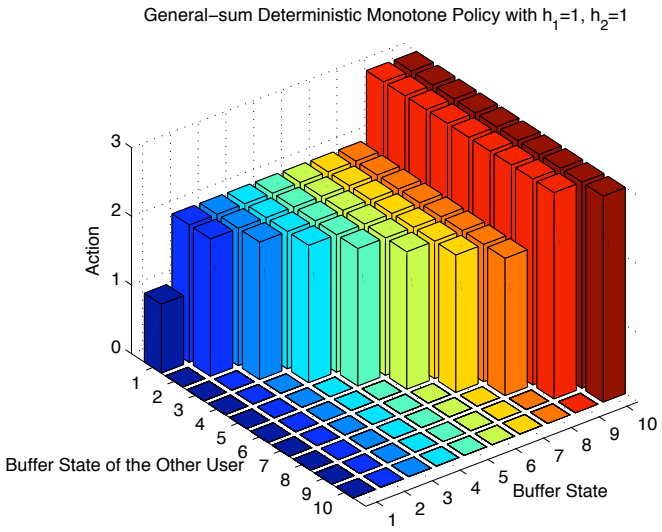

Figure 2.2: The transmission policy of user 1 in a switching controlled general-sum dynamic game system computed by the value iteration algorithm. The result is obtained when the channel states of user 1 and 2 are $h_1 = 1$ and $h_2 = 1$, respectively.

system model with $h_1 = 1$ and $h_2 = 1$. The figure is the Nash equilibrium policy of a switching control game with general-sum transmission costs and zero-sum holding costs. The transmission policy is monotone nondecreasing on the buffer occupancy state.

b) Zero-Sum Constrained Game: Shapley's Theorem says a discounted, zero-sum, two-person stochastic game possesses a value vector that is the unique solution of the game [31]. Fig. 2.3 considers a two-user zero-sum game model and shows the structural result on the optimal transmission policy. The first subfigure is of the result with $h_1 = 1$ and $h_2 = 1$, and the second subfigure is of that with $h_1 = 1$ and $h_2 = 2$. It can be seen from the figures that the policies are deterministic, and under a certain

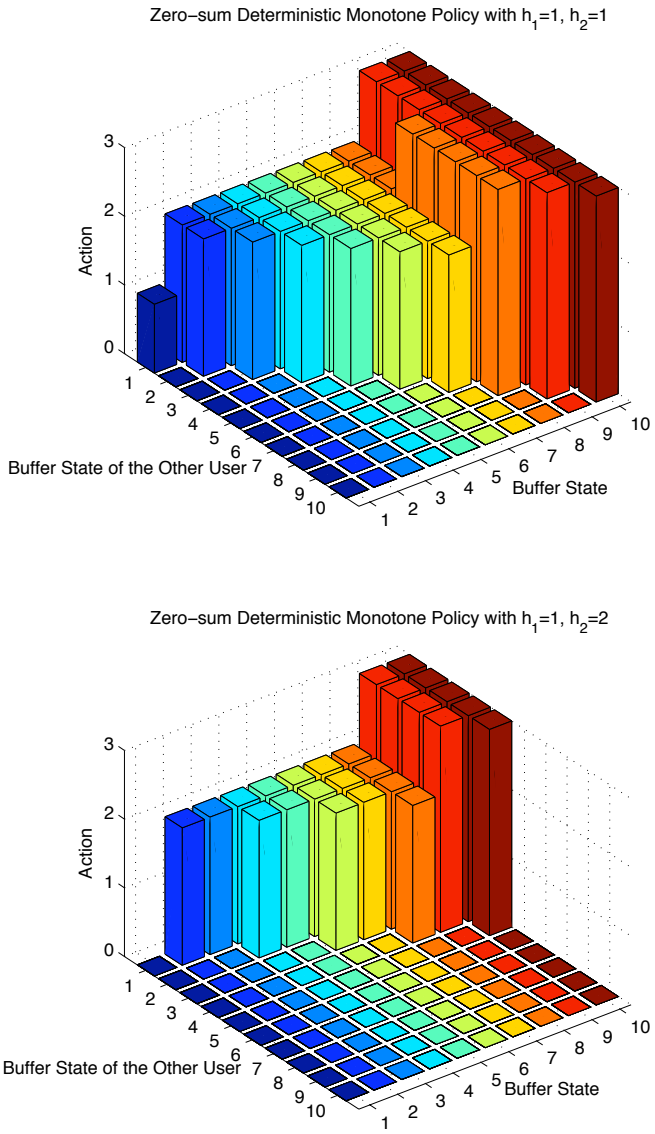

Figure 2.3: The transmission policy of a certain user in a switching controlled zero-sum dynamic game system obtained by value iteration algorithm. The first subfigure shows the result when the channel states of both users are $h_1 = 1$ and $h_2 = 1$, while the second subfigure shows the result when the channel states of user 1 and 2 are $h_1 = 1$ and $h_2 = 2$, respectively.

channel state value, the optimal action policy is monotone nondecreasing on the buffer occupancy state. By comparing subfigures 1 and 2 we see that as the channel state of the other user improves, the current user has lower probability to transmit and its transmission policy becomes more aggressive.

c) Stochastic Approximation Algorithm for Learning Nash Equilibrium Policy: Fig. 2.4 and Fig. 2.6 consider a two-user system with each user having two different action choices denoted as $\{1, 2\}$. As it is a constrained switching controlled Markovian game, the Nash equilibrium policy is a randomized mixture of two pure policies. The optimal transmit policy can be parameterized by lower threshold $b_l(\mathbf{s})$, upper threshold $b_h(\mathbf{s})$ and randomization factor p, as described in (2.14). The stochastic approximation algorithm is applied to estimate the Nash equilibrium policy. The simulation results of user 1 with $h_2 = 1, b_2 = 1$ are shown in Fig. 2.4. The figure shows that the optimal transmission policy is a randomized mixture of two pure policies, each of these policies is monotone nondecreasing on the buffer state.

In order to demonstrate the effectiveness of the proposed stochastic approximation algorithm, we compare its performance with that of the myopic policy. The myopic policy does not consider the effect of current actions on future states, it can be computed by choosing the discount factor $\beta = 0$ in equations (2.7,2.8). It is shown in Fig. 2.5 that under the same channel conditions, the proposed algorithm has better BER performance and buffer management compared to the myopic policy.

The stochastic approximation algorithm proposed in Section 2.3.3 is able to adapt its estimation of the Nash equilibrium policy as the system

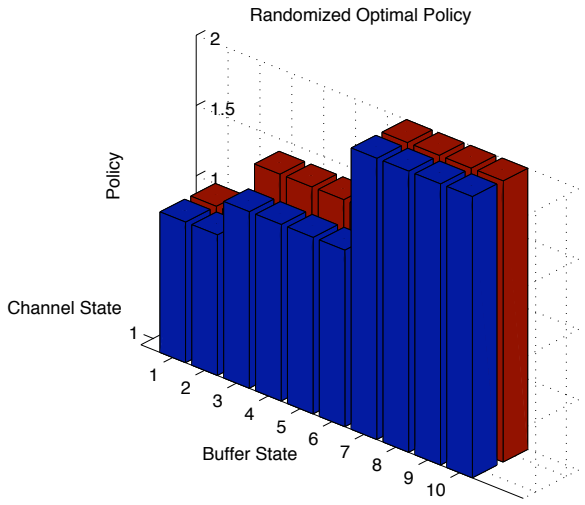

Figure 2.4: The Nash equilibrium transmission control policy computed via stochastic approximation algorithm. A 2-user system is considered, and each user has a size 10 buffer.

parameters evolve with time. Fig. 2.6 shows the time evolution of lower threshold $b_l(\mathbf{s})$, upper threshold $b_h(\mathbf{s})$ and randomization factor p for $h_1 = 1$, $h_2 = 1$ and $b_2 = 1$. The system configuration used from iteration 101 to 200 is different from that used from iteration 1 to 100. We can see from the simulation result that when the system parameters evolve with time, the stochastic approximation algorithm can adapt the transmission policy accordingly. The estimated three parameters of the policy at iteration 100 are $b_l(\mathbf{s}) = 2.4$, $b_h(\mathbf{s}) = 7.6$ and $p = 0.5$. After detecting the change of the system parameters, the new estimated parameters are updated to be $b_l(\mathbf{s}) = 3.8$, $b_h(\mathbf{s}) = 9.7$ and $p = 0.29$. In Fig. 2.6, the estimated upper and lower thresholds of the system computed by Algorithm 2 are shown

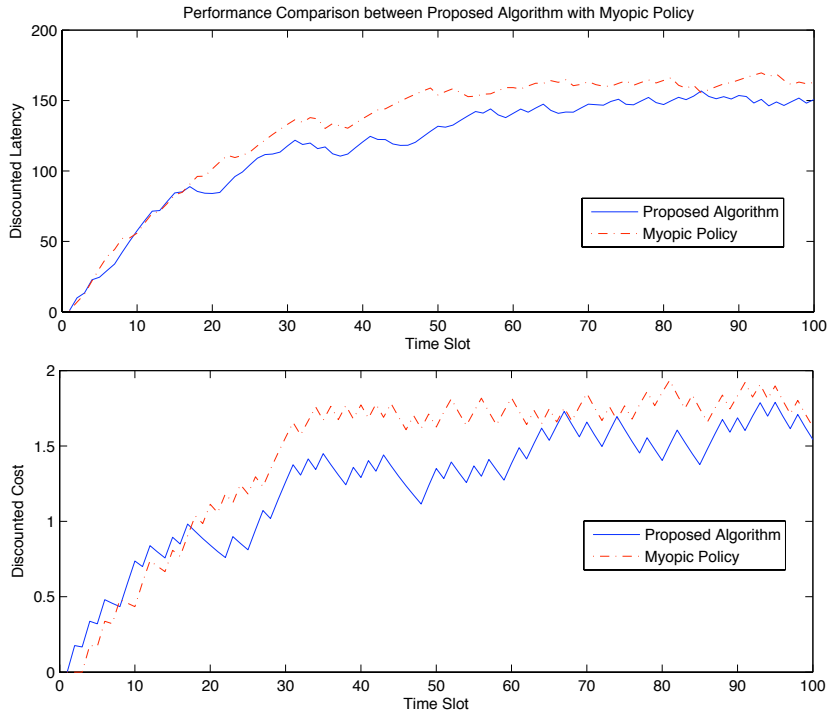

Figure 2.5: Performance comparison between the proposed stochastic approximation algorithm and myopic policy.

by solid lines. By rounding the estimated parameters off to the nearest discrete values, the discrete optimal parameters are obtained (shown by dashed lines).

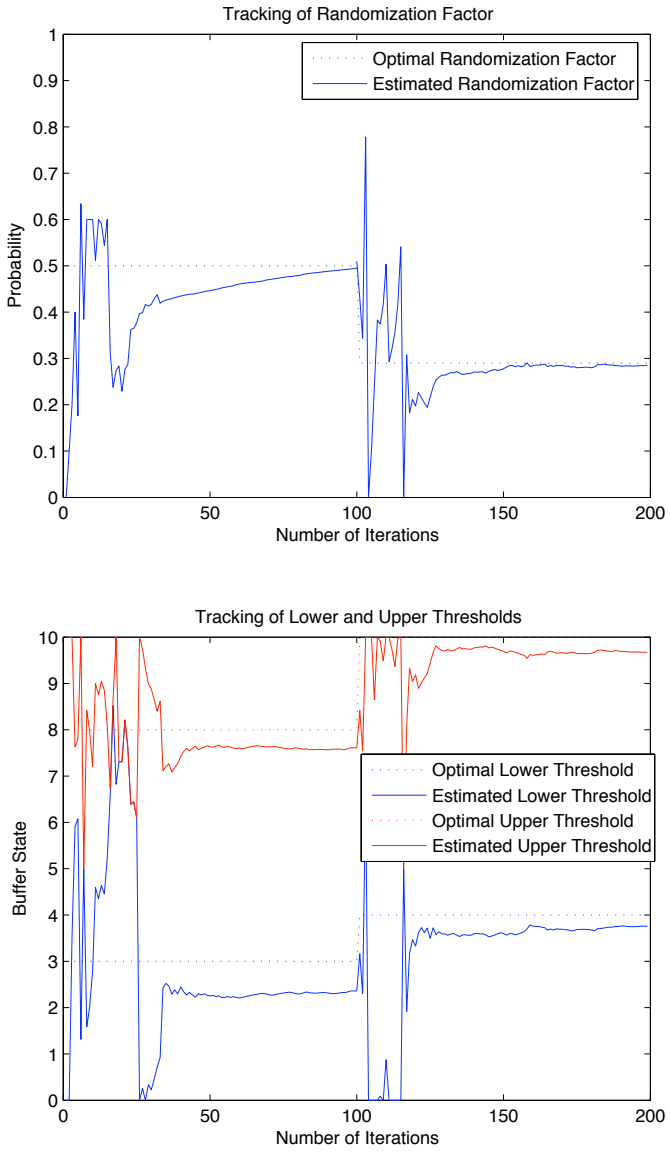

Figure 2.6: Tracking Nash equilibrium policy using the stochastic approximation algorithm (Algorithm 2). System parameters change at the 100th iteration, as specified in Section 2.3. These two figures compare the estimated randomization factor and buffer thresholds with the discrete optimal values.

2.5 Summary

By extending the structural results in [72] and [29],we formulated the rate adaptation problem of secondary users in cognitive radio systems as a constrained general-sum switching control Markovian game, where the block fading channels were modeled as a finite state Markov chain. Assuming perfect information on the primary user behaviour, each secondary user accessed the channel according to a decentralized access rule. The mechanism design ideas can be incorporated in the access rule as in [41]. We showed the Nash equilibria of the game can be calculated by the value iteration algorithm. Under reasonable assumptions, we proved that the Nash equilibrium policy was a randomized mixture of two pure policies, and each of them was monotone nondecreasing on the buffer occupancy state. We then exploited this structural result to devise a computationally efficient stochastic approximation algorithm. Numerical examples were provided to verify these results.

Chapter 3

A Dynamic Games Approach to Transmission Rate Adaptation in Multimedia WLAN [1]

This chapter considers the scheduling, rate adaptation, and buffer management in a multiuser WLAN where each user transmits scalable video payload. Based on opportunistic scheduling, users access the available medium (channel) in a decentralized manner. The rate adaptation problem of the WLAN multimedia networks is then formulated as a general-sum switching control dynamic Markovian game by modelling the video states and block

[1] This chapter is based on the following publication. J. W. Huang, H. Mansour, and V. Krishnamurthy, "A Dynamical Games Approach to Transmission Rate Adaptation in Multimedia WLAN," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3635-3646, July 2010.

fading channel qualities of each user as a finite states Markovian chain. A value iteration algorithm is proposed to compute the Nash equilibrium policy of such a game and the convergence of the algorithm is also proved. We also give assumptions on the system so that the Nash equilibrium transmission policy of each user is a randomization of two pure policies with each policy nonincreasing on the buffer state occupancy. Based on this structural result, we use policy gradient algorithm to compute the Nash equilibrium policy.

3.1 Background

The time-varying nature of wireless channels and video content motivate the need for scalable media which can adapt the video bit-rate without significantly sacrificing the decoded video quality. The Scalable Video Coding (SVC) project provides spatial, temporal, and signal-to-noise ratio (SNR) scalability which ensures a graceful degradation in video quality when faced with channel fluctuations [77].

In this chapter, we consider multiple users transmitting scalable video payload in a WLAN system. The aim is to adapt the video layers of each user to the fluctuations in channel quality and the transmission control polices, such that, the transmission buffer of each user does not saturate and the transmission delay constraint is not violated.

Rate allocation for multimedia transmission in wireless networks has been studied extensively. Previous works that are of interest include [55, 56, 96]. In [56], joint radio link buffer management and scheduling strategies are compared for wireless video streaming over High Speed Download

Package Access (HSDPA) networks. We use the results of [56] to choose an appropriate buffer management strategy for our proposed switching control stochastic dynamic game solution. In [96] and [55] video rate and distortion models are used to improve the transmitted video quality for low latency multimedia traffic. Rate and distortion modelling has become a keystone in model-based video rate and transmission control algorithms. The models are used to predict the coded video packet size and the distortion of the corresponding decoded picture prior to the actual encoding process in order to be used in a constrained optimization framework. However, these models cannot be used for characterization of scalable video data.

In this chapter, we consider an IEEE 802.11 [45] wireless local area network (WLAN) which deploys the CSMA/CA. We propose a modified CSMA/CA channel access mechanism (Section 3.3.3) which takes into account the dynamic behaviours of the video variation, as well as the channel quality and transmission delay of each user. Due to the fact that there is no central controller for resource allocation in a WLAN system, there is strong motivation to study the case where individual users seek to selfishly maximize their transmission rate while taking into account the access rule. This is akin to individuals minimizing their tax (throughput subject to latency constraint in our case) without violating the tax-law (the modified CSMA/CA channel access rule in our case). The interactions among system users can be naturally formulated using game theory. Furthermore, by exploiting the correlated time varying sample paths of the channels and buffers, the transmission rate adaptation problem can be formulated as a Markovian dynamic game. Section 3.2 describes the details of the WLAN system under study

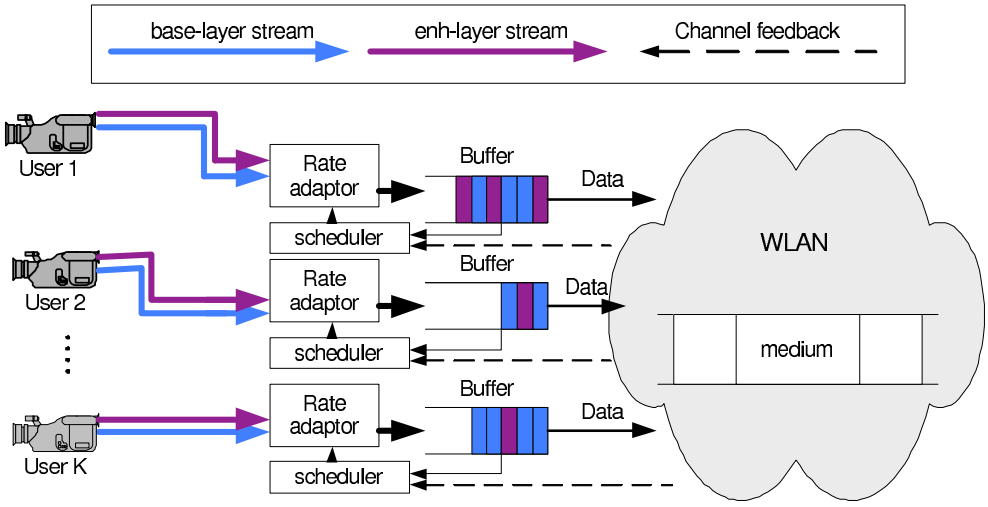

Figure 3.1: A WLAN system where each user is equipped with a size B buffer, a decentralized scheduler and a rate adaptor for transmission control. The users transmit a scalable video payload in which enhancement layers provide quality refinements over the base layer bitstream.

as well as the dynamic behaviour of video sources and the scalable rate-distortion models used to represent this behaviour.

One of the main goals of this thesis is to use mathematical tools to characterize the nature of the optimal solution rather than do brute force computation. To the best of our knowledge, the structural results proved in the thesis are some of the most general available in the literature. They hold under reasonable assumptions as described in Section 3.4.2.

3.2 System Description and the Video Rate-Distortion Model

This section introduces the time-slotted WLAN system model (Fig. 3.1) which consists of multiple users attempting to transmit their data over available channels (medium) [45]. Each user is equipped with a buffer, a decentralized scheduler and a rate adaptor to transmit a scalable video payload which consists of a base layer and multiple enhancement layers. We use game theory to formulate the decentralized behaviours of all the users access the common spectrum resource.

3.2.1 System Description

We consider a K user WLAN system where only one user can access the channel at each time slot according to a modified CSMA/CA mechanism (Section 3.3). The system model also assumes there is no hidden nodes in the network. Thus, we do not need to consider packet collision in our formulation. The rate control problem of each user can be formulated as a constrained dynamic Markovian game by modelling the correlated block fading channel as a Markov chain. Under the predefined decentralized access rule (Section 3.3.3), the problem presented is a special type of game namely a switching control Markovian dynamic game.

We assume that K users (indexed by $k = 1, \ldots, K$) are equipped with scalable video encoders delivering video bitstreams that conform with SVC [80]. In SVC, quality scalability is achieved using Medium-Grained Scalability (MGS) or Coarse-Grained Scalability (CGS) where scalable enhancement packets deliver quality refinements to a preceding layer representation by

re-quantizing the residual signal using a smaller quantization step size and encoding only the quantization refinements in the enhancement layer packets [80]. Moreover, MGS/CGS enhancement layer packets are coded using the *key-picture concept* where the coding loop is closed at the highest scalability layer for key-pictures only, while the loop is closed at the base-layer quality for the remaining pictures. This approach achieves a tradeoff between the coding efficiency of enhancement layers and the drift at the decoder [80].

Assume that each user k encodes its video stream at fixed base and enhancement layer Quantization Parameter (QP) values $q_{l,k}$, where $l = 0$ corresponds to the base layer, $l \geq 1$ to every subsequent enhancement layer. This results in a fluctuation of the video bitrate and distortion as the scene complexity and level of motion varies. The video variation can be captured using the Mean Absolute Difference (MAD) of the prediction residual since it quantifies the mismatch between the original un-coded picture and the Intra/Inter prediction. Therefore, we resort to video rate and distortion models to predict the video distortion of the corresponding decoded picture prior to the actual encoding process. In [61], new rate and distortion models are proposed that capture the variation of MGS/CGS scalable coded video content as a function of two encoding parameters, namely the quantization parameter (QP) and the MAD of the residual signal. We will summarize the results in the next subsection.

3.2.2 Scalable Rate-Distortion Modelling

In this section, we present new rate and distortion estimation models for individual frame representation of MGS/CGS scalable coded video content.

Two real-time encoding parameters are commonly used in rate-distortion estimation models, namely the QP and the MAD of the residual signal.

Dropping the user subscript k, let $q_l, q_{l'} \in \{0, 1, ...51\}$ [47] be the QP values of two distinct video layers l and l', respectively. We have found a simple relationship that estimates the residual MAD $\tilde{m}_l$ of a specific frame given the residual MAD of the same frame at an initial $q_{l'}$ as shown below:

$$\tilde{m}_l = \tilde{m}_{l'} 2^{\varsigma(q_l - q_{l'})} \tag{3.1}$$

where ς is a model parameter typically valued around 0.07 for most sequences [47, 61]. In MGS/CGS scalability, enhancement layer packets contain refinements on the quantization of residual texture information [80]. Therefore, we use the expression in (3.1) to estimate the perceived MAD of each of the MGS/CGS enhancement layers.

Rate model

Here we present the rate model for base and CGS enhancement layer SVC coded video frames. The MGS/CGS [55] base and enhancement layer bitrate in SVC can be expressed as follows:

$$r(l, \tilde{m}_{l'}) = c_l (\tilde{m}_l)^u 2^{-q_l/6} = c_l \big(\tilde{m}_{l'} 2^{\varsigma(q_l - q_{l'})}\big)^u 2^{-q_l/6} \tag{3.2}$$

where c_l is a model parameter, and u is a power factor that depends on the frame type, such that, $u = 1$ for Inter-coded frames and $u \approx \frac{5}{6}$ for Intra-coded frames [47, 61].

Distortion model

The distortion model estimates the decoded picture Luminance Peak Signal-to-Noise Ratio (Y-PSNR). Let q_l be the QP value of layer l, and let $\tilde{m}_l$ be the prediction MAD estimate. The expression of the video Peak Signal-to-Noise Ratio (PSNR) of achieved by decoding all layers up to layer l is then expressed as follows:

$$\begin{aligned} \delta(l, \tilde{m}_{l'}) &= \nu_1 \log_{10}\left((\tilde{m}_l)^u + 1\right) \cdot q_l + \nu_2 \\ &= \nu_1 \log_{10}\left(\left(\tilde{m}_{l'} 2^{a(q_l - q_{l'})}\right)^u + 1\right) \cdot q_l + \nu_2, \end{aligned} \tag{3.3}$$

where u is the same power factor described in (3.2), ν_1 and ν_2 are sequence-dependent model parameters typically valued at 0.52 and 47, respectively [61]. The parameters ν_1 and ν_2 can be refined for each sequence during encoding. Fig. 3.2 illustrates the performance of the proposed distortion model compared to actual encoding of two reference video sequences: Foreman and Football. Fig. 3.2 (a) and (b) measure the Y-PSNR estimation results, while (c) and (d) measure the rate estimation results. The video streams used in the simulation have one base layer and two CGS enhancement layers. The accuracy of this rate distortion model (3.3) can be easily seen in Fig. 3.2.

Note that the parameters ς, ν_1, ν_2, and c_l are only calculated once and remain constant for the entire video sequence.

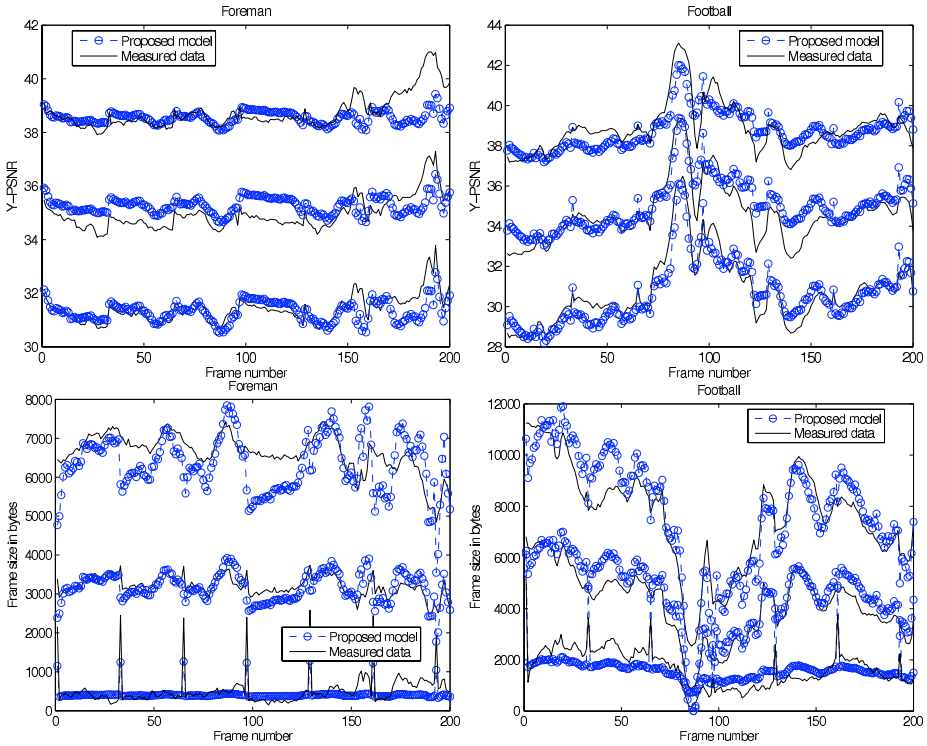

Figure 3.2: Illustration of the modeled PSNR estimates (a) and (b), and the modeled rate estimates (c) and (d) for the base layer and two CGS enhancement layers of the sequences Foreman and Football.

3.3 Uplink Rate Adaptation Problem Formulation

We assume every user has an uplink buffer which holds the incoming coded video frames. When a user is scheduled for transmission, the buffer output rate is chosen depending on the channel quality and buffer occupancy. Therefore, the rate adaptation problem is reformulated as a buffer control problem for every user. In this section, we give a Markovian dynamic game description of this problem.

We assume that the time slot coincides with the video frame duration (30-40 ms [80]). Each user has a fixed packet arrival rate and the the incoming packets are SVC bitstreams with both base and enhancement layers. We denote the incoming number of video frames of user k as $f_{in,k}$ $(k = 1, 2, \ldots, K)$.

Let b_k^n represent the buffer occupancy state of user k at time n and $b_k^n \in \{0, 1, \ldots, B\}$, where B is the maximum number of coded video frames that can be stored in the buffer. The composition of buffer states of all the K users is $\mathbf{b}^n = \{b_1^n, \ldots, b_K^n\}$ and $\mathbf{b}^n \in \mathcal{B}$, where $\mathcal{B}$ is used to denote the space of system buffer states. Furthermore, we use $\mathbf{b}_{-k}^n$ to denote the buffer state composition of all the users excluding user k.

The channel quality of user k at time n is denoted as h_k^n. The channel is characterized using circularly symmetric complex Gaussian random variables which depend only on the previous time slot. By quantizing the channel quality metric, we can denote the resulting discrete channel state space by $h_k^n \in \{1, 2, \ldots\}$. Let $\mathbf{h}^n = \{h_1^n, \ldots, h_K^n\}$ be the composition of channel states of all the K users. Assuming that $\mathbf{h}^n \in \mathcal{H}, n = 1, 2, \ldots, N$ is block fading, with block length equal to each time period, the channel states constitute a Markov process with transition probability from time n to $(n + 1)$ given by $\mathbb{P}(\mathbf{h}^{(n+1)}|\mathbf{h}^n)$.

Let m_k^n be the MAD of user k at time n. In [96], the MAD between two consecutive video frames is modeled as a stationary first-order Gauss Markov process. We quantize the range of m_k^n to achieve a video variability state space denoted by $m_k^n \in \mathcal{M} = \{0, 1, 2 \ldots, M\}$. The video states constitute a Markov process with transition probabilities given by $\mathbb{P}(m_k^{(n+1)}|m_k^n)$. The

system video state at time n is a composition of the video states of all the users and $\mathbf{m}^n = \{m_1^n, m_2^n, \ldots, m_K^n\}$.

The system state at time n is composed of the channel state, video state and buffer state, which is denoted as $\mathbf{s}^n = \{\mathbf{h}^n, \mathbf{m}^n, \mathbf{b}^n\}$. Let $\mathcal{S}$ denote the finite system state space, which comprises channel state space $\mathcal{H}$, video state space $\mathcal{M}$, and user buffer state space $\mathcal{B}$. That is, $\mathcal{S} = \mathcal{H} \times \mathcal{M} \times \mathcal{B}$. Here $\times$ denotes a Cartesian product. $\mathcal{S}_k$ is used to indicate the states where user k is scheduled for transmission. $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_K$ are disjoint subsets of $\mathcal{S}$ with the property of $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_K$.

3.3.1 Actions, Transmission Reward and Holding Cost

The action a_k^n by user k at time slot n is defined to be the number of coded video frames taken from the buffer and transmitted. The rate control algorithm considered in this chapter adjusts the output frame-rate a_k^n according to the channel quality and buffer occupancy. Fig. 3.3 illustrates the buffer control mechanism adopted in this chapter.

Actions: Let a_k^n denote the action of the kth user at the nth time slot. We express a_k^n as the number of coded video frames included in the outgoing traffic payload. If $\mathbf{s}^n \in \mathcal{S}_k$, then $a_k^n \in \mathcal{A} = \{a_{\min}, \ldots, a_{\max}\}$, otherwise, $a_k^n = 0$. $a_{\min}$ and $a_{\max}$ denote the minimum and maximum video frames that user k can output at a time n, respectively.

We assume that the number of video layers $l_{a_k^n}$ at the output window at time slot n is dictated by a channel dependent rule. Therefore, we express

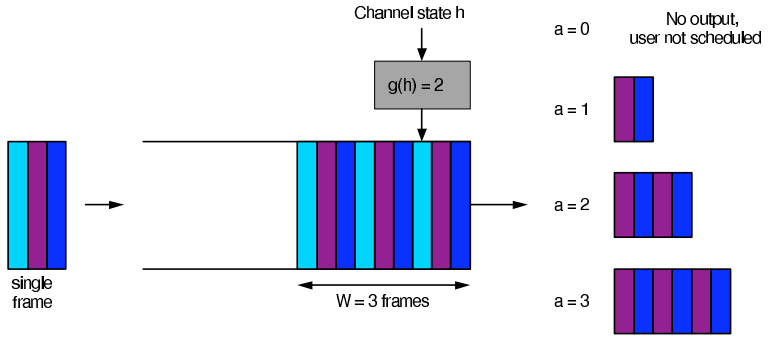

Figure 3.3: Example of the buffer control mechanism assumed in this chapter where $f_{in,k} = 1$ $(k = 1, 2, \ldots, K)$. At every time-slot, a new coded video frame enters the buffer. The buffer output depends on the scheduling algorithm involved, the buffer occupancy, and the channel quality. If a user is scheduled for transmission, then the action taken will extract a specific number l of video frame layers from up to N frames stored in the buffer.

the output number of layers of the video frames at time slot n as follows:

$$l_{a_k^n} = g(h_k^n), \tag{3.4}$$

where $g(.)$ is the channel dependent rule and it is an increasing function on the channel state. The remaining video layers $l > l_{a_k^n}$ are dropped. This assumption is to reduce the action space and allows us to derive the structural results as illustrated later in Section 3.4.2. Consequently, the action space constitutes the number of video frames to be transmitted and is given by $\mathcal{A} = \{1, 2, \cdots, W\}$.

Let L_k be the number of video layers admitted into the buffer of user k, each input packet of user k has L_k layers. The video layer index is denoted as $l \in \mathcal{L} = \{0, 1, \cdots L_k - 1\}$, where $l = 0$ corresponds to the base layer, $l \geq 1$

to every subsequent enhancement layer.

The buffer output packet a_k^n of user k at time n is determined by the modified CSMA/CA decentralized channel access algorithm. If user k is scheduled for transmission at time slot n, $a_k^n > 0$, otherwise, $a_k^n = 0$. However, the system selectively transmits some packet layers according to the channel quality. E.g., the system transmits all of the L_k layers of each packet if the channel quality h_k^n is good, while it only transmits the base layers of each packet if the channel quality h_k^n is poor.

Transmission reward: Let $c_k(\mathbf{s}^n, a_1^n, \ldots, a_K^n)$ denote the transmission reward of user k at time n. Specifically, $c_k(\mathbf{s}^n, a_1^n, \ldots, a_K^n)$ is chosen to be the expected video PSNR resulting from transmitting $l_{a_k^n}$ video layers from a_k^n video frames stored in the buffer of user k. Let $\delta(l, m_k^n)$ be the video PSNR achieved by decoding all packets up to layer l of user k at time n achieved with video state m_k^n, where m_k^n is the base-layer MAD. Let $p_l(h_k^n)$ be the Packet Error Rate (PER) encountered by user k at the lth layer during the transmission. The transmission reward can be expressed as [61]:

$$\begin{aligned} & c_k(\mathbf{s}^n, a_1^n, \ldots, a_K^n) \\ = \; & a_k^n \cdot \big((1 - p_0(h_k^n)) \cdot \delta(0, m_k^n) + \\ & \sum_{l=1}^{l_{a_k^n}-1} \prod_{j=0}^{l} (1 - p_j(h_k^n)) \cdot [\delta(l, m_k^n) - \delta(l-1, m_k^n)] \big). \end{aligned} \tag{3.5}$$

The bit errors are assumed to be independent identically distributed (i.i.d.), and $\delta(l, m_k^n)$ is given in (3.3). Note here that the video distortion measure $\delta(l, m_k^n)$ can either be considered as the picture mean-square-error (MSE) or the PSNR. In the case of PSNR, the objective will be to *maximize* $c_k(\mathbf{s}^n, a_k^n)$, otherwise it is minimized. In this chapter, we consider the PSNR as the video

quality metric.

Note in equation (3.5), we assume a target Packet Error Rate (PER) which allows the flexibility of modulation schemes and channel coding selections. It simplifies both the expression of the transmission reward function (3.5) and the derivation of the Nash equilibrium policy.

It can be seen from (3.5) that at time slot n if the kth user is scheduled for transmission the performance of user k depends solely on its own channel state h_k^n and not on the actions. Thus, for $\mathbf{s}^n \in \mathcal{S}_k$ the rewards of all the users in the system are

$$c_k(\mathbf{s}^n, a_1^n, \ldots, a_K^n) = c_k(\mathbf{s}^n, a_k^n) \geq 0 \tag{3.6}$$

$$c_i(\mathbf{s}^n, a_1^n, \ldots, a_K^n) = 0, \quad (i \neq k). \tag{3.7}$$

For notational convenience, in the following sections we will drop the subscript k by defining

$$c(\mathbf{s}^n, a_k^n) := c_k(\mathbf{s}^n, a_k^n).$$

Holding cost: Each user has an instantaneous Quality of Service (QoS) constraint denoted as $d_k(\mathbf{s}^n, a_1^n, \ldots, a_K^n)$ where $k = 1, \ldots, K$. If the QoS is chosen to be the delay (latency) then $d_k(\mathbf{s}^n, a_1^n, \ldots, a_K^n)$ is a function of the buffer state b_k^n of the current user k. The instantaneous holding costs will be subsequently included in an infinite horizon latency constraint.

We express the holding cost of user k with $\mathbf{s}^n \in \mathcal{S}_k$ as follows:

$$d_k(\mathbf{s}^n, a_k^n) = \frac{1}{\kappa} \cdot ([b_k^n - a_k^n + f_{in,k}]^+)^\tau, \quad \tau \geq 1 \tag{3.8}$$

where κ is the average output frame rate which is determined by the system. τ is a constant factor which is specified to be $\tau \geq 1$, this is due to the fact that the data overflow probability becomes higher with more data in the buffer. The latency cost at time slot n is evaluated by the buffer state after taking action a_k^n.

Since the transmission latency is independent of the actions of all the remaining users, it can be simplified as

$$d_k(\mathbf{s}^n, a_1^n, \ldots, a_K^n) = d_k(\mathbf{s}^n, a_k^n) \geq 0, \text{ where } \mathbf{s}^n \in \mathcal{S}_k. \tag{3.9}$$

For the remaining $K-1$ users who are not scheduled for transmission, their holding costs can be expressed as

$$d_{i,i\neq k}(\mathbf{s}^n, a_k^n) = \frac{1}{\kappa} \cdot \min(b_i^n + f_{in,i}, B)^\tau. \tag{3.10}$$

3.3.2 Markovian Game Formulation

The intuition behind the Markovian game formulation for the rate adaptation problem in a WLAN system is as follows: A WLAN system does not have a central authority for the resource allocation and the system access rule is typically a decentralized opportunistic scheme. Due to the selfish nature of the system users, each user aims to maximize its own payoff. The interaction among users is characterized as the competition for the common system resources, which can best be formulated using game theory. By modelling the transmission channel as correlated Markov process, the rate adaptation problem can further be formulated as a Markovian game.

At time instant n, assume user k is scheduled for transmission according to the system access rule specified in Section 3.3.3. We define $c(\mathbf{s}^n, a_k^n) \geq 0$ to be the instantaneous reward of user k when the system is in state $\mathbf{s}^n$. We use $\Phi_1, \Phi_2, \ldots, \Phi_K$ to represent the set of all the deterministic policies of each user, respectively. The infinite horizon expected total discounted reward of any user i given its transmission policy π_i ($\pi_i \in \Phi_i$), can be written as

$$C_i(\pi_i) \quad = \quad \mathbb{E}_{\mathbf{s}^n}\left[\sum_{n=1}^{\infty} \beta^{n-1} \cdot c_i(\mathbf{s}^n, a_k^n)\big|\pi_i\right] \tag{3.11}$$

where $0 \leq \beta < 1$ is the discount factor, the state $\mathbf{s}^n \in \mathcal{S}_k$ and the expectation is taken over action a_k^n as well as system state $\mathbf{s}^n$ evolution for $n = 1, 2, \ldots$. The holding cost of user i at that time is $d_i(\mathbf{s}^n, a_k^n)$ and the infinite horizon expected total discounted latency constraint can be written as

$$D_i(\pi_i) = \mathbb{E}_{\mathbf{s}^n}\left[\sum_{n=1}^{\infty} \beta^{n-1} \cdot d_i(\mathbf{s}^n, a_k^n)\big|\pi_i\right] \leq \widetilde{D}_i, \tag{3.12}$$

with $\widetilde{D}_i$ being a system parameter depending on the system requirement on user i.

Optimization Problem: Given the policies of the other users, the optimal transmission policy of user i, π_i^*, is chosen so as to maximize the overall expected discounted reward subject to its constraint, this can be mathematically written as:

$$\pi_i^* = \{\pi_i : \max_{\pi_i \in \Phi_i} C_i(\pi_i) \;\; s.t.\; D_i(\pi_i) \leq \widetilde{D}_i, i = 1, \ldots, K\}. \tag{3.13}$$

Not that the choice of the optimal rate transmission policy of ith user

is a function of the policies of other users. Each user aims to optimize his own discounted reward (3.11) under the latency constraint (3.12).

3.3.3 System Access Rule

This chapter adopts a WLAN system model (IEEE 802.11 [45]) with a modified CSMA/CA mechanism which will be implemented by a decentralized channel access rule. The decentralized channel access algorithm can be constructed as follows: At the beginning of a time slot, each user k attempts to access the channel after a certain time delay t_k^n. The time delay of user k can be specified via an opportunistic scheduling algorithm [28], such as

$$t_k^n = \frac{\gamma_k}{b_k^n h_k^n}. \tag{3.14}$$

Here γ_k is a user specified quality of service (QoS) parameter and $\gamma_k \in \{\gamma_p, \gamma_s\}$. As soon as a user successfully access the channel, the remaining users detect the channel occupancy and stop their attempt to access. Let k^{*n} denote the index of the first user which successfully accesses the channel. If there are multiple users with the same minimum waiting time, k^{*n} is randomly chosen from these users with equal probability.

3.3.4 Transition Probabilities and Switching Control Game Formulation

With the above setup, the system feature satisfies the assumptions of a special type of dynamic game called a switching control game [31, 68, 90], where the transition probabilities depend on only one player in each state. It is known that the Nash equilibrium for such a game can be computed by solving a sequence of Markov decision processes [31]. The decentralized

transmission control problem in a Makovian block fading channel WLAN system can now be formulated as a switching control game. In such a game [31], the transition probabilities depend only on the action of the kth user when the state $\mathbf{s} \in \mathcal{S}_k$, this enables us to solve such type of game by a finite sequence of Markov decision processes.

According to the property of the switching control game, when the kth user is scheduled for transmission, the transition probability between the current composite state $\mathbf{s} = \{\mathbf{h}, \mathbf{m}, \mathbf{b}\}$ and the next state $\mathbf{s}' = \{\mathbf{h}', \mathbf{m}', \mathbf{b}'\}$ depends only on the action of the kth user a_k. The transition probability function of our problem can now be mathematically expressed by the following equation.

$$\begin{aligned} & \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_1, a_2, \ldots, a_K) \\ = \; & \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) \\ = \; & \prod_{i=1}^{K} \mathbb{P}(h_i'|h_i) \cdot \prod_{i=1}^{K} \mathbb{P}(m_i'|m_i) \cdot \prod_{i=1, i\neq k}^{K} \mathbb{P}(b_i'|b_i) \cdot \mathbb{P}(b_k'|b_k, a_k). \end{aligned} \tag{3.15}$$

As each user is equipped with a size B buffer, the buffer occupancy of user k evolves according to Lindley's equation [24, 25]

$$b_k' = \min([b_k - a_k + f_{in,k}]^+, B). \tag{3.16}$$

The evolution of the buffer state of user $i = 1, 2, \ldots, K$ when $i \neq k$ follows the following rule:

$$b_i' = \min(b_i + f_{in,i}, B).$$

For user k, the buffer state transition probability depends on input and

output date rate. This can be expressed as follows:

$$\mathbb{P}(b'_k|b_k, a_k) = I_{\{b'_k=\min([b_k-a_k+f_{in,k}]^+,B)\}}, \tag{3.17}$$

where $I_{\{\cdot\}}$ is the indicator function.

For those users that are not scheduled for transmission, the buffer state transition probability depends only on the distribution of incoming traffic, which can be written as

$$\mathbb{P}(b'_i|b_i) = I_{\{b'_i=\min(b_i+f_{in,i},B)\}}. \tag{3.18}$$

Equations (3.11, 3.12, 3.15) define the constrained switching control Markovian game we consider. Our goal is to solve such games. That is, we seek to compute a Nash equilibrium[2] policy $\pi_i^*, i = 1, \ldots, K$ (which is not necessarily unique). However, if both the reward (3.11) and constraint (3.12) are zero-sum among all the users, then all the Nash equilibria have a unique value vector and are globally optimal [31].

The Markovian switching control game can be solved by constructing a sequence of MDP as described in Algorithm 3. We refer to [31, Chapter 3.2], [68] for the proof. So, the constrained switching controlled Markovian game (3.11, 3.12) can be solved by iteratively updating the transmission policy π_i^{*n} of user $i, i = 1, \ldots, K$ with the policy of remaining users fixed. Here n denotes the iteration index as mentioned in Algorithm 3. At each step, ith

[2]A Nash equilibrium [31] is a set of policies, one for each player, such that no player has incentive to unilaterally change their action. Players are in equilibrium if a change in policies by any one of them would lead that player to earn less than if it remained with its current policy.

user aims to maximize the overall expected discounted reward subject to its constraint as specified in (3.13).

3.4 Nash Equilibrium Solutions to the Markovian Dynamic Game

This section first presents a value iteration algorithm to compute the Nash equilibrium solution to the formulated general-sum dynamic Markovian switching control game and proves the convergence of such algorithm. Then, we introduce four assumptions on the transmission reward, holding cost and state transition probability functions which lead us to the structural result on the Nash equilibrium policy. This structural result enables us to search for the Nash equilibrium policy in the policy space using a policy gradient algorithm and it has reduced computational cost.

3.4.1 Value Iteration Algorithm

In [31], a value iteration algorithm was designed to calculate the Nash equilibrium for an unconstrained general-sum dynamic Markovian switching control game. Therefore, we convert the problem in (3.13) to an unconstrained one using Lagrangian relaxation and use the value iteration algorithm specified in Algorithm 3 to find a Nash equilibrium solution.

The algorithm can be summarized as follows. We use $\mathbf{V}^n_{k=1,2,\ldots,K}$ and $\lambda^m_{k=1,2,\ldots,K}$ to represent the value vector and Lagrange multiplier, respectively, of a user k ($k = 1, \ldots, K$) at the nth and mth time slot. The algorithm mainly consists of two parts: the outer loop and the inner loop. The outer loop updates the Lagrange multipliers of each user and the inner loop

Algorithm 3 Value Iteration Algorithm

Step 1:
Set $m = 0$; Initialize l.
Initialize $\{\mathbf{V}_1^0, \mathbf{V}_2^0, \ldots, \mathbf{V}_K^0\}$, $\{\lambda_1^0, \lambda_2^0, \ldots, \lambda_K^0\}$.
Step 2: Inner Loop: Set n=0;
Step 3: Inner Loop: Update Transmission Policies;
for $k = 1 : K$ **do**
for each $\mathbf{s} \in \mathcal{S}_k$,
$\pi_k^n(\mathbf{s}) = \arg\min_{\pi_k^n(\mathbf{s}) \in \mathcal{A}} \Big\{ - c(\mathbf{s}, a_k) + \lambda_k^m \cdot d_k(\mathbf{s}, a_k) + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}') \Big\}$;
$v_k^{n+1}(\mathbf{s}) = -c(\mathbf{s}, \pi_k^n(\mathbf{s})) + \lambda_k^m \cdot d_k(\mathbf{s}, \pi_k^n(\mathbf{s})) + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, \pi_k^n(s)) v_k^n(\mathbf{s}')$;
$v_{i=1:K, i\neq k}^{n+1}(\mathbf{s}) = \lambda_i^m \cdot d_i(\mathbf{s}, \pi_k^n(\mathbf{s})) + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, \pi_k^n(\mathbf{s})) v_i^n(\mathbf{s}')$;
end for
Step 4: If $\|\mathbf{V}_k^{n+1} - \mathbf{V}_k^n\| \leq \varepsilon, k = 1, \ldots, K$, set $n = n + 1$, and return to Step 3; Otherwise, go to Step 5.
Step 5: Update Lagrange Multipliers
for $k = 1 : K$ **do**
$\lambda_k^{m+1} = \lambda_k^m + \frac{1}{l}\Big[D_k(\pi_1^n, \pi_2^n, \ldots, \pi_K^n) - \widetilde{D}_k\Big]$
end for
Step 6: The algorithm stops when $\lambda_k^m, k = 1, 2, \ldots, K$ converge, otherwise, set $m = m + 1$ and return to Step 2.

optimize the transmission policy of each user under fixed Lagrange multipliers. The outer loop index and inner loop index are m and n, respectively. Note from Algorithm 3 the interaction between users is through the update of value vectors since $v_{i=1:K, i\neq k}^{n+1}(s)$ is a function of $\pi_k^n(s)$.

In Step 1, we set the outer loop index m to be 0 and initialize the step size l, the value vectors $\mathbf{V}_{k=1,2,\ldots,K}^0$ and Lagrange multipliers $\lambda_{k=1,2,\ldots,K}^0$. Step 3 is the inner loop where at each step we solve kth user controlled game and obtain the new optimal strategy for that user with the strategy of the remaining players fixed. Note here that the objective of the system is to

maximize the transmission reward and minimize the holding cost as shown in (3.13). Since $-c(\mathbf{s}, a_k)$ is used for each step the optimal transmission policy $\pi_k^n(\mathbf{s})$ is obtained by doing the minimization. Variable ε in step 4 is a small number chosen to ensure the convergence of $\mathbf{V}_k^n$. Step 5 updates the Lagrange multipliers based on the discounted delay value of each user given the transmission policies $\{\pi_1^n, \pi_1^n, \ldots, \pi_K^n\}$. $\frac{1}{l}$ is the step size which satisfies the conditions for convergence of the policy gradient algorithm. This sequence of Lagrange multipliers $\{\lambda_1^m, \ldots, \lambda_K^m\}$ with $m = 0, 1, 2, \ldots$ converges in probability to $\{\lambda_1^*, \ldots, \lambda_K^*\}$ which satisfies the constrained problem defined in (3.13) [50, 83]. The algorithm terminates when $\lambda_{k=1,2,\ldots,K}^m$ converge, otherwise, go to Step 2.

Since this is a constrained optimization problem, the optimal transmission policy is a randomization of two deterministic polices [24, 25]. We use $\lambda_{k=1,2,\ldots,K}^*$ to represent the Lagrange multipliers obtained with the above algorithm. The randomization policy of each user can be written as:

$$\pi_k^*(\mathbf{s}) = q_k \pi_k^*(\mathbf{s}, \lambda_{k,1}) + (1 - q_k)\pi_k^*(\mathbf{s}, \lambda_{k,2}), \tag{3.19}$$

where $0 \leq q_k \leq 1$ is the randomization factor and $\pi_k^*(\mathbf{s}, \lambda_{k,1})$, $\pi_k^*(\mathbf{s}, \lambda_{k,2})$ are the unconstrained optimal policies with Lagrange multipliers $\lambda_{k,1}$ and $\lambda_{k,2}$. Specifically, $\lambda_{k,1} = \lambda_k^* - \Delta$ and $\lambda_{k,2} = \lambda_k^* + \Delta$ for a perturbation parameter Δ. The randomization factor of the kth user q_k is calculated by:

$$q_k = \frac{\widetilde{D}_k - D_k(\lambda_{1,2}, \ldots, \lambda_{K,2})}{D_k(\lambda_{1,1}, \ldots, \lambda_{K,1}) - D_k(\lambda_{1,2}, \ldots, \lambda_{K,2})} \tag{3.20}$$

The convergence proof of the inner loop of Algorithm 3 is shown in Appendix B and this value iteration algorithm obtains a Nash equilibrium solution to the constrained switching control Markovian game with general sum reward and general sum constraint [31].

Remark: The primary purpose of the value iteration algorithm is to prove the structural results on the Nash equilibrium policy. The value iteration algorithm (Algorithm 3.4.1) is not designed to be implemented in a practical system because at each iteration of the algorithm, a user k is required to know the channel states of all the other users and the system state transition probability matrix.

3.4.2 Structural Result on Randomized Threshold Policy

In this subsection, we characterize the structure of the Nash equilibrium achieved by Algorithm 3. First, we list four assumptions. Based on these four assumptions, Theorem 3.4.1 is introduced.

- *A 3.4.1:* The set of policies that satisfy constraint (3.12) is non-empty, to ensure the delay constraint of the system is valid.

  A 3.4.1 is the feasible assumption on the system constraint, and it is assumed to be satisfied.

- *A 3.4.2:* Transmission reward $c(\mathbf{s}, a_k)$ is a supermodular[3] function of

[3]If a function $f : \mathcal{A} \times \mathcal{B} \times \mathcal{C} \rightarrow \mathcal{R}$ is supermodular in (a, b) for any $c \in \mathcal{C}$. Then for all $a' \geq a$ and $b' \geq b$, $f(a', b'; c) - f(a, b'; c) \geq f(a', b; c) - f(a, b; c)$ holds. If f is a supermodular function in (a, b), then $-f$ is a submodular in (a, b).

b_k, a_k for any channel quality h_k and MAD m_k of the current user. $c(\mathbf{s}, a_k)$ is also a integer concave function of $b_k - a_k$ for any h_k and m_k.

It can be seen from Section 3.3.1 that the transmission reward $c(\mathbf{s}, a_k)$ is linear in a_k (3.5) and independent of the buffer state b_k. Thus the assumption A 3.4.2 holds.

- *A 3.4.3:* Holding cost $d_k(\mathbf{s}, a_k)$ is a submodular function of b_k, a_k for any channel quality h_k and MAD m_k of the current user. $d_k(\mathbf{s}, a_k)$ is also integer convex in $b_k - a_k$ for any h_k and m_k.

  The holding cost of user k with $\mathbf{s} \in \mathcal{S}_k$ as follows:

$$d_k(\mathbf{s}, a_k) = \kappa \cdot (b_k - a_k + f_{in,k})^{\tau}. \tag{3.21}$$

  The submodularity property of the holding cost function $d_k(\mathbf{s}, a_k)$ can be easily verified. It is also straightforward to show $d_k(\mathbf{s}, a_k)$ is integer convex in $b_k - a_k$.

- *A 3.4.4:* $\mathbb{P}(b'_k | b_k, a_k)$ is second order stochastically increasing on $(b_k - a_k)$ for any b_k and a_k.

  $\mathbb{P}(b'_k | b_k, a_k)$ expression given in (3.17) shows it is first order stochastically increasing in $(b_k - a_k)$. First order stochastic dominance implies second order stochastic dominance [24, 25], thus assumption A 3.4.4 holds.

The following theorem is one of our main results. It shows that the Nash equilibrium of a constrained dynamic switching control game is a randomization of two pure monotone policies.

Theorem 3.4.1 *For each user k and a given channel state, the Nash equilibrium policy $\pi_k^*(\mathbf{s})$ is a randomization of two pure policies $\pi_k^1(\mathbf{s})$ and $\pi_k^2(\mathbf{s})$. Each of these two pure policies is a nonincreasing function with respect to buffer occupancy state b_k.* □

The detailed proof of Theorem 3.4.1 is given in Appendix C.

3.4.3 Learning Nash Equilibrium Policy via Policy Gradient Algorithm

The value iteration algorithm in Algorithm 3 is applicable to a general type of constrained switching control game. In the case that the system parameters satisfy assumptions A 3.4.1-A 3.4.4 as stated above, we can exploit the structural result (Theorem 3.4.1) to search for the Nash equilibrium policy in the policy space. This results in a significant reduction on the complexity of computing the Nash equilibrium policies.

The main idea of searching the Nash equilibrium policy in the policy space is as follows. If there are three actions available with action set $a_k =$

$\{1, 2, 3\}$, the Nash equilibrium policy $\pi_k^*(\mathbf{s})$ is a randomized mixture:

$$\pi_k^*(s) = \begin{cases} 1, & \text{if} \quad 0 \leq b_k < b_1(\mathbf{s}) \\ p_1, & \text{if} \quad b_1(\mathbf{s}) \leq b_k < b_2(\mathbf{s}) \\ 2, & \text{if} \quad b_2(\mathbf{s}) \leq b_k < b_3(\mathbf{s}) \\ p_2, & \text{if} \quad b_3(\mathbf{s}) \leq b_k < b_4(\mathbf{s}) \\ 3, & \text{if} \quad b_4(\mathbf{s}) \leq b_k \end{cases} . \tag{3.22}$$

Here, $\{p_1, p_2\} \in [0, 1]$ is the randomization factor, $b_1(\mathbf{s})$, $b_2(\mathbf{s})$, $b_3(\mathbf{s})$ and $b_4(\mathbf{s})$ are the buffer thresholds. The search for each Nash equilibrium policy problem is now converted to the estimation of these 6 parameters. Note here that the policy search applies to a system with any number of actions, here we consider a 3 action system as an example.

The simultaneous perturbation stochastic approximation (SPSA) method [83] is adopted to estimate the parameters. SPSA is especially efficient in high-dimensional problems in terms of providing a good solution for a relatively small number of measurements for the objective function. The essential feature of SPSA is the underlying gradient approximation, which requires only two objective function measurements per iteration regardless of the dimension of the optimization problem. These two measurements are made by simultaneously varying, in a properly random fashion all of the variables in the problem. The detailed algorithm is described in Algorithm 4.

The first part of the algorithm initializes system variables. $\boldsymbol{\theta}^n$ represents the union of all the parameters we search for at the nth time slot, and $\boldsymbol{\theta}^n = \{\boldsymbol{\theta}_1^n, \boldsymbol{\theta}_2^n, \ldots, \boldsymbol{\theta}_K^n\}$. $\boldsymbol{\theta}_k^n$ indicates parameter vector of the kth user. The Lagrange multipliers of all the K users are defined to be $\boldsymbol{\lambda}^n = \{\lambda_1^n, \lambda_2^n, \ldots, \lambda_K^n\}$.

Algorithm 4 Policy Gradient Algorithm

1: **Initialization**: $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\lambda}^0$; $n = 0$; $\rho = 4$;
2: Initialize constant perturbation step size β and gradient step size α;
3: **Main Iteration**
4: **for** $k = 1 : K$ **do**
5: $dim_k = 6 \times |h_k| \times |h_i| \times |b_i|$;
6: Generate $\boldsymbol{\Delta}^n = [\Delta_1^n, \Delta_2^n, \ldots, \Delta_{dim_k}^n]^T$; Δ_i^n are Bernoulli random variables with $p = \frac{1}{2}$.
7: $\boldsymbol{\theta}_{k+}^n = \boldsymbol{\theta}_k^n + \beta \times \boldsymbol{\Delta}^n$;
8: $\boldsymbol{\theta}_{k-}^n = \boldsymbol{\theta}_k^n - \beta \times \boldsymbol{\Delta}^n$;
9: ΔC_k^n
$= \frac{c_k(\mathbf{s}^n, \boldsymbol{\theta}_{k+}^n) - c_k(\mathbf{s}^n, \boldsymbol{\theta}_{k-}^n)}{2\beta}[(\Delta_1^n)^{-1}, (\Delta_2^n)^{-1}, \ldots, (\Delta_{m_k}^n)^{-1}]^T$;
10: ΔD_k^n
$= \frac{d_k(\mathbf{s}^n, \boldsymbol{\theta}_{k+}^n) - d_k(\mathbf{s}^n, \boldsymbol{\theta}_{k-}^n)}{2\beta}[(\Delta_1^n)^{-1}, (\Delta_2^n)^{-1}, \ldots, (\Delta_{m_k}^n)^{-1}]^T$;
11: $\boldsymbol{\theta}_k^{(n+1)} = \boldsymbol{\theta}_k^n - \alpha \times \left(\Delta C_k^n + \Delta D_k^n \cdot \max\left[0, \lambda_k^n + \rho \cdot \left(D(\mathbf{s}^n, \boldsymbol{\theta}_k^n) - \widetilde{D}_k\right)\right]\right)$;
11: $\lambda_k^{(n+1)} = \max\left[(1 - \frac{\alpha}{\rho} \cdot \lambda_k^n), \lambda_k^n + \alpha \cdot \left(D(\mathbf{s}^n, \boldsymbol{\theta}_k^n) - \widetilde{D}_k\right)\right]$;
12: **end for**
13: The parameters of other users remain unchanged;
14: $n = n + 1$;
15: The iteration terminates when the values of the parameters $\boldsymbol{\theta}^n$ converge; else return back to Step 3.

β and α denote the constant perturbation step size and constant gradient step size, respectively. When $\alpha \to 0$, $\frac{\alpha}{\beta^2} \to 0$ and $n \to \infty$, the policy gradient algorithm converges weakly [54, Theorem 8.3.1]. In the main part of the algorithm, SPSA algorithm is applied to iteratively update system parameters. When the kth user is scheduled to transmit at time slot n, parameters $\boldsymbol{\theta}_k^n$ and the Lagrange multiplier λ_k^n can be updated after introducing a random perturbation vector $\boldsymbol{\Delta}^n$. Meanwhile, the parameters of the other users remain unchanged. The algorithm terminates when $\boldsymbol{\theta}^n$ converge.

Remark: At each iteration of the policy gradient algorithm, user k is only required to know its own transmission reward and holding cost. The

transmission reward and holding cost of user k are functions of its own channel state and buffer state (Section 3.4.2). Thus, the policy gradient algorithm is distributed and implementable in a practical system.

3.5 Numerical Examples

In this section, we illustrate the performance improvement of the proposed dynamic switching control game policy over myopic policies. We also present numerical examples of the Nash equilibrium transmission policy. For our comparisons, we used the Joint Scalable Video Model (JSVM)-9-8 reference software [87] to encode two reference video sequences in Common Intermediate Format (CIF) resolution: Foreman and Football. Each sequence is composed of 100 frames encoded in SVC, with an H.264/AVC compatible base layer, and two CGS enhancement layers. The coded video streams are composed of I and P frames only, with only the first frame encoded as an I frame, a frame rate of 30 fps, and minimum QoS guarantee corresponding to the average base-layer quantization parameter value of $q_0 = 38$. The first and second CGS enhancement layers have average QP values of 32 and 26, respectively. The video MAD m_k is quantized into three states $m_k \in \mathcal{M} = \{0, 1, 2\}$, such that,

$$m_k = \begin{cases} 0, & \text{if } MAD_k < \mu_k - \frac{\sigma_k}{2}, \\ 1, & \text{if } \mu_k - \frac{\sigma_k}{2} \leq MAD_k \geq \mu_k + \frac{\sigma_k}{2}, \\ 2, & \text{if } MAD_k > \mu_k + \frac{\sigma_k}{2} \end{cases}$$

where μ_k and σ_k are the mean and standard deviation of the MAD of user k.

Table 3.1 lists the bit-rate in bits per second and average distortion in terms of the luminance peak signal-to-noise ratio (Y-PSNR) of each video

layer of the two sequences.

Table 3.1: Average incoming rate and distortion (Y-PSNR) characteristics of the two video users.

Sequence:	Foreman		Football	
Video Layer	$\bar{r}_{l,1}$ (kbps)	**Y-PSNR (dB)**	$\bar{r}_{l,2}$ (kbps)	**Y-PSNR (dB)**
Base $l = 0$	40.9	30.75	55.82	28.17
First CGS $l = 1$	103.36	34.59	139.03	32.47
Second CGS $l = 2$	242.64	38.87	252.97	37.29

In all the cases considered here, the channel quality measurements h_k of user k are quantized into two different states, $h_k \in \{1, 2\}$. In the models used, each user has a size 10 buffer with an incoming rate equivalent to increasing the buffer size by *one* during a transmission time slot ($f_{in,1} = f_{in,2} = \cdots = f_{in,K} = 1$). The transmission time slot duration is set equal to 10 ms thus allowing a maximum transmission delay of 100 ms. This system configuration ensures that the transmission rewards, holding costs and buffer transition probability matrices satisfy assumptions A 3.4.2-A 3.4.3 specified in Section 3.4.2. The channel is assumed to be Markovian and the transition probability matrices are generated randomly.

In the system models used, each user has 3 different action choices when it is scheduled for transmission. The 3 different actions are $a_k = 1, 2, 3$ and correspond to the user buffer output of one, two, and three frames, respectively. The action $a_k = 0$ corresponds to the *No-Transmit* state. Therefore, when the Nash equilibrium policy is 0, the current user is not scheduled for transmission and $a_k = 0$. However, the outgoing traffic is still one frame $f_{in,k} = 1$.

General-Sum Constrained Game: Randomized Monotone Transmission Policy. Fig. 3.4 considers the same two user system of the previous example with a 25 ms transmission delay constraint. As it is a constrained switching controlled Markovian game, the Nash equilibrium policy is a randomization of two pure policies. The figure shows that the optimal transmission policies are no longer deterministic but are a randomization of two pure policies, and that each Nash equilibrium policy is monotone nonincreasing on the buffer state. The results of Fig. 3.4 are obtained by applying the value iteration algorithm. Fig. 3.4 (a) is the optimal policy of user 1 when $h_2 = 1$ and $b_2 = 1$, while, Fig. 3.4 (b) shows the optimal policy of user 2 when $h_1 = 1$ and $b_1 = 1$.

Result by Policy Gradient Algorithm: The policy gradient algorithm (Algorithm 4) we propose uses the structural result on the optimal transmit policy and each policy can be determined by three parameters, namely, lower threshold $b_l(\mathbf{s})$, upper threshold $b_h(\mathbf{s})$ and randomization factor p as described in (3.22) (assume there are 2 actions). The simulation results of the policy of user 1 with $h_2 = 1, b_2 = 1$ shown in Fig. 3.5. In the system setup, each user has a buffer size of 10 and 3 actions and the system transmission delay constraint is 25ms. By comparing Fig. 3.5 to Fig. 3.4 (a) we can see that the results obtained by the value iteration algorithm and policy gradient algorithm are very close.

Transmission Buffer Management: We adopt a priority based packet drop similar to the Drop Priority Based (DPB) buffer management policy proposed in [56]. With this policy, low priority packets that have resided longest in the transmission buffer are dropped when the buffering delay ap-

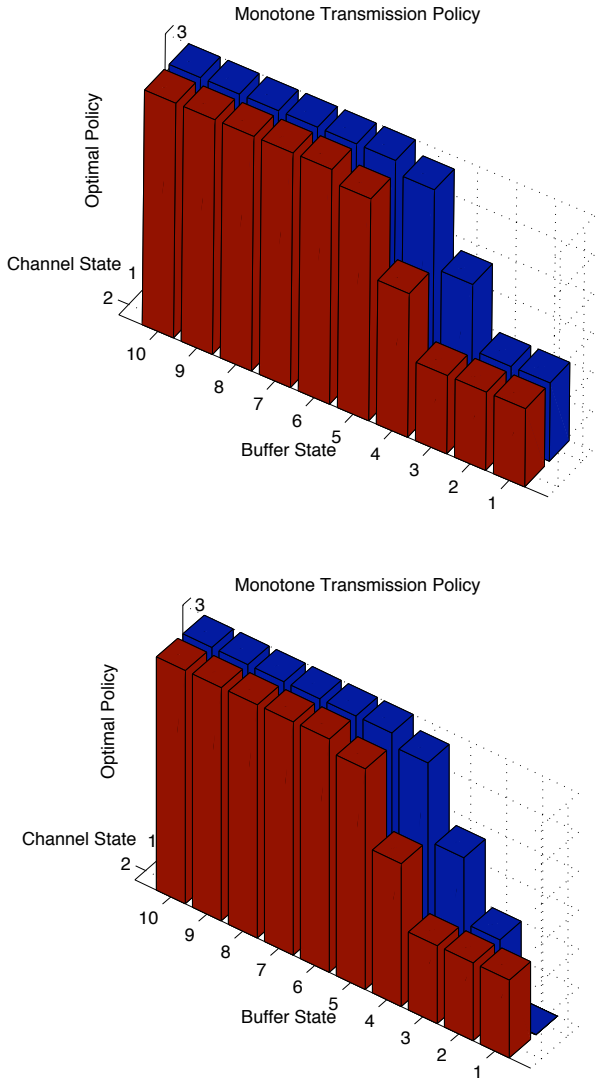

Figure 3.4: The Nash equilibrium transmission policy obtained via value iteration algorithm for user 1 (a) and user 2 (b). The result (a) is obtained when $h_2 = 1$ and $b_2 = 1$ and (b) is obtained when $h_1 = 1$ and $b_1 = 1$, respectively. The transmission delay constraint is specified to be 25 ms. Each Nash equilibrium transmission policy is a randomization of two pure monotone policies.

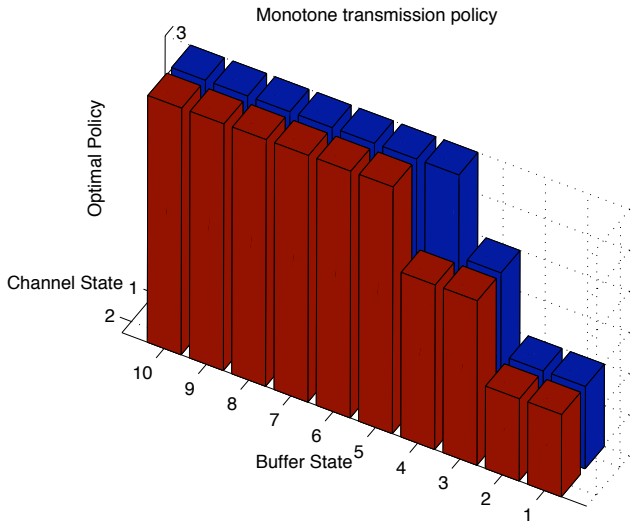

Figure 3.5: The Nash equilibrium transmission control policy obtained via policy gradient algorithm for user 1. The result is obtained with a 25 ms transmission delay constraint and the states of users 2 are $h_2 = 1$ and $b_2 = 1$. The transmission policy is monotone nonincreasing on its own buffer state.

proaches the delay constraint. In [56], the DPB policy was found to be second best to the Drop Dependency Based (DDB) which requires video dependency information that cannot be collected in real time. The DPB policy is attractive since it requires minimal priority information about the video payload which is readily available in the video Network Abstraction Layer Unit (NALU) header of a MGS/CGS coded video stream. We implement this policy by shifting the user action a_k to a lower state that satisfies the buffer delay constraint.

In order to demonstrate the effectiveness of the proposed dynamic switching control game algorithm, we compare its performance to that of a myopic

policy that selects at each time slot the user action that maximizes the video PSNR while satisfying the transmission buffer delay constraint. The time slot duration equals to the video frame capturing duration which is around 30ms. The myopic policy does not consider the effect of current actions on future states. We assume that the delay constraint corresponds to the packet play-out deadline, therefore, all buffered packets that exceed the delay constraint are assumed to be lost and are discarded form the user buffer. Figs. 3.6 and 3.7 show the results of simulating the transmission of two users Foreman and Football for both the proposed switching control game policy and the myopic policy. The figures show that under the same channel conditions, the proposed policy has better Y-PSNR performance and transmission buffer utilization for both users.

3.6 Summary

This chapter considers a time-slotted multiuser WLAN system where each user delivers SVC video bitstream. The modified CSMA/CA mechanism of such system is implemented through an opportunistic scheduling system access rule. By modelling video states and block fading channels as a finite state Markov chain, the rate adaptation problem among users can be formulated as a constrained general-sum switching control Markovian game. The Nash equilibrium transmission policy of such game can be computed through a value iteration algorithm. Given four assumptions on the transmission reward, holding cost and transition probability functions (Section 3.4.2), the Nash equilibrium policy is a randomized mixture of two pure policies with each policy monotone nonincreasing on the buffer state occupancy. This

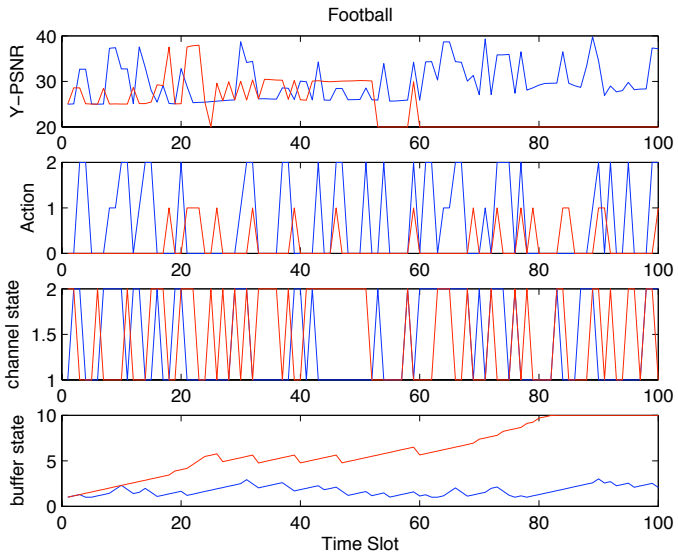

Figure 3.6: Result of the transmission of the Football sequence comparing the performance in terms of video PSNR and buffer utilization between the proposed switching control game policy and the myopic policy with 80 ms delay constraint. The result shows that the proposed switching control game policy performs better than the myopic policy.

structural result enables us to search for the Nash equilibrium policy in the policy space via a policy gradient algorithm.

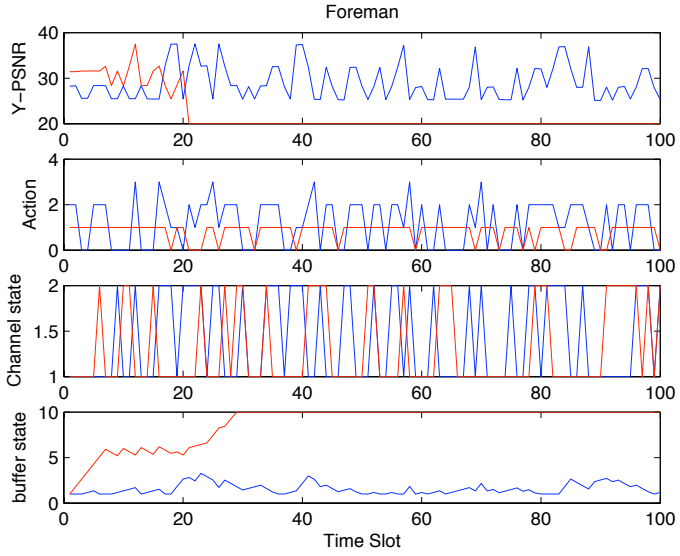

Figure 3.7: Result of the transmission of the Foreman sequence comparing the performance in terms of video PSNR and buffer utilization between the proposed switching control game policy and the myopic policy with 80 ms delay constraint. The result shows that the proposed switching control game policy performs better than the myopic policy.

Chapter 4

Cognitive Base Stations in 3GPP LTE Femtocells: A Correlated Equilibrium Game Theoretic Approach [1]

This chapter considers downlink spectrum allocation in an LTE system macrocell which contains multiple femtocells. By incorporating cognitive capabilities into femtocell base stations, the Home evolved Node Bs (HeNBs) can be formulated as secondary base stations seeking to maximize the spectrum utility while minimizing interference to primary base stations (evolved Node-Bs). The competition amongst cognitive HeNBs for spectrum re-

[1] This chapter is based on the following manuscript. J. W. Huang and V. Krishnamurthy, "Cognitive Base Stations in LTE/3GPP Femtocells: A Correlated Equilibrium Game Theoretic Approach," *IEEE Transactions on Communications*, September 2011.

sources is formulated as a non-cooperative game-theoretic learning problem where each agent (HeNB) seeks to adapt its strategy in real time. We formulate the resource block (RB) allocation among HeNBs in the downlink of an LTE system using a game-theoretic framework, where the correlated equilibrium solutions of the formulated game are being investigated. A distributed RB access algorithm is proposed to compute the correlated equilibrium RB allocation policy.

4.1 Background

An important feature of 3GPP LTE systems [1] is that it allows distributed implementation of femtocells to meet a variety of service requirements. The femtocell access points, denoted as or Home evolved Node-B (HeNB) in 3GPP, are low-cost, low-power, plug-and-play cellular base stations. In order to provide broadband connectivity, these HeNBs will need to possess *adaptive/cognitive* facilities. In the October 2010 release of 3GPP, HeNBs are described as self-optimized nodes in a Self-Organized Network (SON) which need to maintain quality of service (QoS) with minimal intervention from the service operator [86]. HeNBs are equipped with cognitive functionalities for load balancing, interference management, random access channel optimization, capacity and coverage optimization, and handover parameter optimization.

With the above motivation, this chapter considers spectrum resource allocation in an orthogonal frequency division multiple access (OFDMA) LTE downlink system which consists of a macrocell base station (eNB) and multiple femtocell base stations (HeNBs). By incorporating cognitive capabilities

into these self-optimized femtocell base stations, the cognitive HeNBs aim to maximize the spectrum utility by utilizing the unoccupied frequencies while minimizing interference to the eNB (primary base station) in a spectrum overlay LTE system. The unit of spectrum resource to be allocated in an LTE system is called a resource block (RB) and it is comprised of 12 subcarriers at a 15 kHz spacing.

Given the RB occupancy of the eNB, the competition for the spectrum resources among HeNBs can be formulated in a game-theoretic setting [91]. Instead of computing the Nash equilibrium policy of the formulated game, we seek to characterize and compute the *correlated equilibrium* policy set [7, 8]. The set of correlated equilibria is a convex polytope. It includes the set of Nash equilibria – indeed the Nash equilibria are isolated points at the extrema of this set [69, 70]. The set of correlated equilibria [8] is arguably the most natural attractive set for a decentralized adaptive algorithm such as the one considered here, and describes a condition of competitive optimality between agents (cognitive femtocell base stations). It is more preferable than Nash equilibria since it directly considers the ability of agents to coordinate their actions. This coordination can lead to higher performance than if each agent was required to act in isolation. Indeed, Hart and Mas-Colell observed in [36] that, for most simple adaptive procedures, "... there is a natural coordination device: the common history, observed by all players. It is thus not reasonable to expect that, at the end, independence among players will obtain." Since the set of correlated equilibria is convex, fairness between players can be addressed in this domain. Finally, decentralized, online adaptive procedures naturally converge to the correlated equilibria,

whereas the same is not true for Nash equilibria (the so-called law of conservation of coordination [37]).

There are several important issues being addressed in recent literature regarding the deployment of HeNB femtocells in an LTE system. One area of interest is how to mitigate the interferences among HeNBs so as to improve the system performance. In [59] and [58], Lopez-Perez *et al.* used Dynamic Frequency Planning (DFP), an interference avoidance technique, to decrease the inter-cell interference and increase the network capacity by dynamically adapting the radio frequency parameters to the specific scenario. They also verified the performance of the DFP technique in an OFDMA WiMAX macrocells and femtocells system. Choi *et al.* investigated in [18] how to minimize the interference caused by femtocells in an open access (spectrum underlay) network. They showed adaptive open access will maximize the value of a femtocell both to its owner and to the network using numerical results.

Attar *et al.* studied the benefits of developing cognitive base-stations in a UMTS LTE network [6]. Radio resource management protocols are not specified by standards, such as 3GPP's UMTS LTE. Thus, there is considerable flexibility in their design. The insufficiency of traditional coexistence solutions in LTE context is shown in [6]. It is argued that cognitive base-stations are crucial to an efficient and distributed radio resource management of LTE given its distributed architecture. One motivation for such argument is the lessons learnt from wide-spread deployment of wireless local area network (WLAN) access points. The simple plug-and-play nature of WLAN routers, along with the unlicensed nature of WLAN spectrum access, alleviated the

need for time and cost-intensive network plannings. This in turn helped a rapid proliferation of WLAN hotspots. However, as the number of coexisting WLAN networks increases, so does their mutual interfering effect, rendering such simple, selfish coexistence strategies problematic. By incorporating the main three cognitive radio capabilities into the LTE base-stations, which are

1. radio scene analysis;
2. online learning based on the feedback from RF environment;
3. and agile/dynamic resource access schemes;

a successful coexistence strategy among eNBs and HeNBs can be achieved.

4.2 Resource Allocation Among HeNBs: Problem Formulation

We consider a macrocell area with a number of femtocells randomly deployed by home and offices within an OFDMA LTE network (Fig. 4.1). By incorporating the cognitive capacities into femtocell base stations, the spectrum occupancy behaviour of the macrocell base station (eNB) can be formulated as a primary base station and that of the femtocell base stations (HeNBs) can be formulated as secondary base stations (cognitive base stations) in a spectrum overlay network[2]. Due to the selfish nature of each base station, the competition for the common spectrum resource among cognitive base stations can be formulated using a game-theoretic framework.

[2]As described in [6], it is also possible to study an *underlay* cognitive femtocell strategy in which RBs accessed by eNB and HeNBs are not orthogonal. However, as the objective of our analysis is modelling the competition among selfish HeNBs, the proposed solutions can be readily extended to the aforementioned spectrum underlay LTE systems.

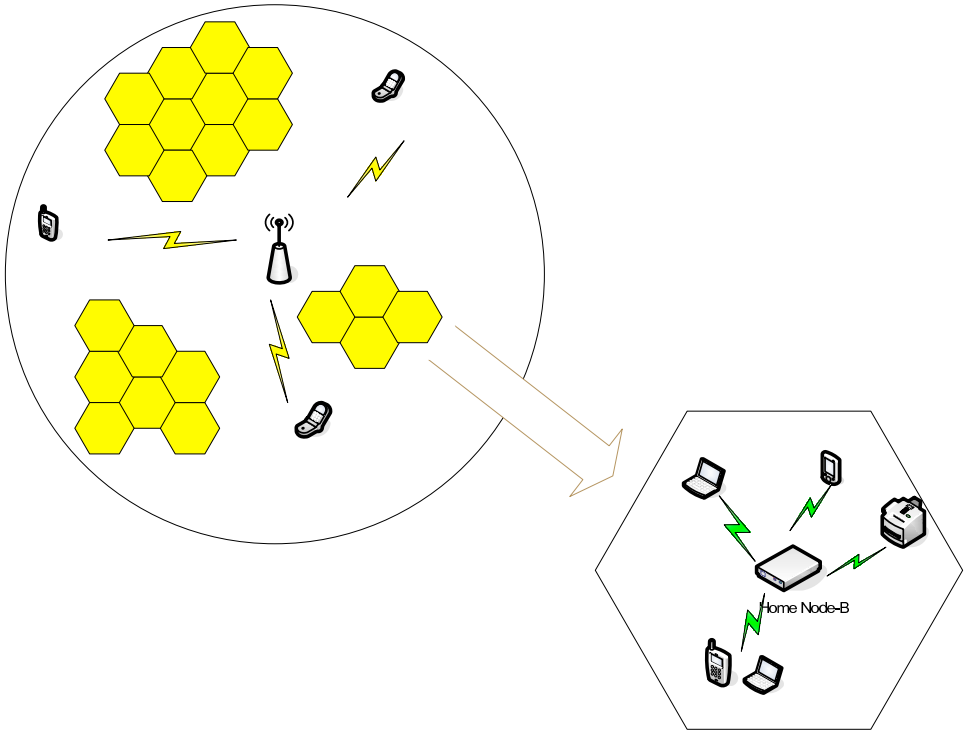

Figure 4.1: System schema of a single eNB macrocell which contains multiple HeNB femtocells in a 3GPP LTE network.

4.2.1 System Description

The resource allocation process in LTE networks follows a time slotted system model where each time slot length equals to that of an RB (0.5 ms), with RB being the smallest unit of resource that can be allocated. Let $t \in \{1, 2, \ldots, T\}$ denote the time slot index. T defines the time horizon of the formulated game. Let N denote the total number of available RBs in the system, N^t_{henb} denote the number of RBs occupied by HeNBs at time t and N^t_{eNB} denotes the number of RBs occupied by eNB at time t. As we consider a spectrum overlay system, $N^t_{henb} \leq \left(N - N^t_{eNB}\right)$.

Let K denote the total number of HeNBs coexisting in the network, $k \in \mathcal{K} = \{1, 2, \ldots, K\}$ denote the user index and $f \in \{1, 2, \ldots N^t_{henb}\}$ denote the RB index occupied by cognitive base stations (HeNBs). We use $p^t_k(f) \in \{0, 1\}$ to denote the action of the kth HeNB on the fth RB at time t, where 0 represents *not transmit* and 1 represents *transmit*.

Let $\mathbf{p}^t_k = \{p^t_k(1), \ldots, p^t_k(N^t_{henb})\} \in \mathcal{P}_k$ denote the action of the kth HeNB over all the available RBs to HeNBs at time t, with $\mathcal{P}_k$ denoting the action space of the kth HeNB. $\mathbf{p}^t = \{\mathbf{p}^t_1, \ldots, \mathbf{p}^t_K\} \in \mathcal{P}$ is used to denote the composition of the actions from all the HeNBs at time t. $\mathcal{P}$ is the joint action space of all HeNBs.

We use $s^t_k(f) \in \{1, 2, \ldots, Q_s\}$ to denote channel quality state of the kth HeNB on the fth RB at time t after quantization. For example, the channel quality can be obtained by quantizing a continuous valued channel model comprising of circularly symmetric complex Gaussian random variables into Q_s different states. Let $\mathbf{s}^t_k = \{s^t_k(1), \ldots, s^t_k(N^t_{henb})\}$ denote the channel state composition of the kth HeNB over all the available RBs and $\mathbf{s}^t = \{\mathbf{s}^t_1, \ldots, \mathbf{s}^t_K\}$ denote the channel state composition over all the HeNBs.

Let $I^t_k(f)$ denote the interference introduced to the kth HeNB at time t on the fth RB. The interference comprises two parts, namely, noise and Co-Channel Interference (CCI). The noise $n^t_k(f)$ is assumed to be Additive White Gaussian Noise (AWGN) with a noise covariance of $\sigma^2_k(f)$ and the CCI is introduced by having different HeNBs sharing the same RB. An interference matrix $\mathbf{w}^t(f)$ is introduced to model the CCI among all HeNBs on the fth RB at time t. The elements of this interference matrix, i.e., $w^t_{i,j}(f)$ where $i, j \in \mathcal{K}$, denote the cross channel quality between the ith and

jth HeNBs. Assuming channel reciprocity, $\mathbf{w}^t(f)$ is a symmetric matrix, i.e., $w^t_{i,j}(f) = w^t_{j,i}(f)$. It is clear that the interference matrix has zero-valued diagonal elements, i.e., $w^t_{i,i}(f) = 0$ for $\forall\ i \in \mathcal{K}$. The value of $\mathbf{w}^t(f)$ depends on the location of all the HeNBs.

We assume the kth ($k \in \mathcal{K}$) HeNB has the following information at the beginning of a time slot t.

1. N^t_{henb}: the number of available RBs for HeNBs,
2. $\mathbf{s}^t_k$, i.e., the channel state vector of the kth HeNB,
3. $I^t_k(f)$ which is the received interference at HeNB k on RB f at time t,
4. the current demand level of the kth HeNB which is denoted by d^t_k.

N^t_{henb} can be obtained through a fixed broadband access network (e.g., Digital Subscriber Line (DSL), Cable) as described in [85]. $\mathbf{s}^t_k$ and $I^t_k(f)$ can be obtained by channel sensing mechanism during the guard interval at the beginning of each time slot. d^t_k is used to denote the system resource demand level of HeNB k at time t, it is of the same unit as that of system capacity. d^t_k is determined by the user requirement with in a HeNB cell at time t. In the case that there are more devices transmit data using frequency bandwidth within HeNB k at time t, d^t_k is of higher value. An important characteristic of this model is that radio-specific quantity d^t_k needs only to be known to the kth HeNB cell, thus allowing decentralized resource allocation algorithms. Based on the above information, HeNB k chooses its action vector $\mathbf{p}^t_k$, selfishly, to maximize its local utility function. The definition of utility will be presented in Section 4.2.2.

Note that in the case that HeNBs are owned by different agents and they are so sophisticated to behave maliciously, they can opt not to reveal their true demand levels d_k^t $(k = 1, \ldots, K)$ to optimize their own utilities at the cost of reducing the overall system performance. It requires mechanism design theory in order to prevent this from happening. Similar problems has been studied in [42] where the authors applied pricing mechanism to ensure each rational selfish user maximize its own utility function, at the same time optimizing the overall system utility. Reputation based mechanism design is another area of research, where the system uses reputation as a tool to motivate cooperation between users and indicate a good behaviour within the network. If a user does not pay attention to its reputation and keep acting maliciously, it will be isolated and discarded. Such reputation based mechanism has been applied in ad hoc networks and sensor networks [16, 78]. However, this chapter assumes all the malicious behaviours have been eliminated in the system and each HeNB uses its true demand level d_k^t to compute its utility function.

The distributed decision making process amongst HeNBs defines the action set $\mathbf{p}^t$, which in turn leads to a different realization of interference level $I_k^t(f)$, $\forall\ k \in \mathcal{K}$ and $f = \{1, \ldots, N_{henb}^t\}$. Therefore, the action of one femtocell base station (HeNB) affects the utilities of other femtocells, which motivates the use of game-theoretic approaches to analyze and compute the RB allocation policies among all the HeNBs. In the following subsection, we define the global system utility function and the local utility functions for femtocell base stations.

4.2.2 Utility Function

The goal of this chapter is to optimize the global resource allocation problem using a decentralized approach. We should demonstrate a connection between global utility function and the local utility function that will guide the allocation decision of each HeNB. This connection is presented through the derivation of global and local performance measures. This subsection defines a global utility function to evaluate the overall system performance, based on which a local utility function is defined for each cognitive HeNB. Each HeNB selfishly maximizes its local utility function which does not guarantee the global system performance. We aim to design local utility functions which ensure global system performance quality at the correlated equilibrium of the formulated game.

Let C_k^t denote the capacity of HeNB k at time t. If a capacity-achieving code such as turbo code or low-density parity-check code (LDPC) code is applied for error correction, C_k^t can be expressed as follows using Shannon-Hartley's theorem [93].

$$C_k^t = \sum_{f=1}^{N_{henb}^t} \omega \cdot \log_2 \left[1 + \frac{p_k^t(f) \times s_k^t(f)}{I_k^t(f)} \right] \tag{4.1}$$

$$I_k^t(f) = \sigma_k^2(f) + \sum_{i=1}^{K} w_{k,i}^t(f) \times p_i^t(f), \tag{4.2}$$

where ω denotes the bandwidth of an RB. We assume that all the HeNBs treat the interferences as Gaussian noises. Note that $w_{k,k}^t = 0$ for $k = 1, \ldots, K$. (4.2) models the Inter User Interference (IUI) as a function of actions of all the other users in the system. However, in a practical system,

$I_k^t(f)$ is obtained by channel estimation through downlink common pilots. Thus, a user k is not required to have full knowledge of $\mathbf{p}^t$ in order to compute C_k^t.

Global Utility Function

Since all cognitive HeNBs have equal priority in accessing system resources, a global objective is chosen to maximize the performance of the worst-off HeNB such that the available resources are fairly allocated among HeNBs. Therefore, given the action vector $\mathbf{p}^t$ of all the HeNBs, the global utility function at time t is defined as follows.

$$U_G(\mathbf{p}^t) = \min_{k\in\mathcal{K}} \big[\min(\frac{C_k^t}{d_k^t}, 1)\big]. \tag{4.3}$$

Here, the term $\min(\frac{C_k^t}{d_k^t}, 1)$ represents the *satisfaction level* of the kth HeNB and it is a function of the instantaneous capacity of the kth HeNB at time t divided by its current demand level. Note that mathematically (4.3) is equivalent to $U_G(\mathbf{p}^t) = \min_{k\in\mathcal{K}} \frac{C_k^t}{d_k^t}$, operation $\min(\frac{C_k^t}{d_k^t}, 1)$ is applied because the satisfaction level is within the range of $[0, 1]$.

The action vector among all the HeNBs at time t $\mathbf{p}^t$ is chosen to maximize the minimum satisfaction level among all the K HeNBs. That is,

$$\mathbf{p}^t = \arg\max_{\mathbf{p}\in\mathcal{P}} U_G(\mathbf{p}) = \max_{\mathbf{p}\in\mathcal{P}} \min_{k\in\mathcal{K}} \big[\min(\frac{C_k^t}{d_k^t}, 1)\big]. \tag{4.4}$$

This global utility optimization problem (4.4) aims to maximize the minimum satisfaction level among all the K HeNBs. Note the global utility function can be of other forms, e.g., aiming to maximize the average sat-

isfaction level among all the K HeNBs, in which case the global utility function can be specified as follows (4.5). However, this chapter focuses on the scenario where the global utility is chosen to maximize the worst-off user (4.3).

$$U_G(\mathbf{p}^t) = \frac{1}{K} \sum_{k \in \mathcal{K}} \big[\min(\frac{C_k^t}{d_k^t}, 1) \big]. \tag{4.5}$$

The system objective is to find the action vector $\mathbf{p}^t$ so as to maximize the satisfaction level of the worst-off HeNB $U_G(\mathbf{p}^t)$. The proposed approach to achieve the correlated equilibrium policy in a decentralized way, is to allow each cognitive HeNB choosing its action $\mathbf{p}_k^t$ ($k \in \mathcal{K}$) based on the optimization of its local utility function. The question that remains to be answered is how to select a proper local utility function which also ensures a good overall system performance. In the next subsection, we will derive such a utility function and determine its relation to the global utility objective.

Local HeNB Utility Function

If each cognitive HeNB has a reasonable estimate of the global utility function, U_G, then the decentralized resource allocation policy can be directly realized by each cognitive HeNB choosing an action which maximizes its estimate of U_G. However, as the global utility function also depends on the private information of other players, i.e., the demand levels d_i^t and actions $\mathbf{p}_i^t$ ($i \in \mathcal{K}$ and $i \neq k$) of other HeNBs, such an assumption is not practical.

Below, we construct a local utility function U_k for HeNB k ($k \in \mathcal{K}$) consisting of three parts, where each part addresses a certain aspect of the

problem at hand.

The first part of the local utility function reflects the *self interest* component of a cognitive HeNB, given by,

$$U_k[1](\mathbf{p}_k^t) = \min\big(\frac{C_k^t}{d_k^t}, 1\big). \tag{4.6}$$

Maximizing $U_k[1](\mathbf{p}_k^t)$ is equivalent to maximizing HeNB k's portion of the global utility U_G (4.3). However, a game with (4.6) as the only component of the local utility function would resemble a congestion game, which may not be solvable in closed form. Moreover, (4.6) only shows the self interested part of the global objective (4.3) and it neglects the inter-relation of decisions of each player on the achieved utility of other players. Since each HeNB known only its own demand leven d_k^t and action $\mathbf{p}_k^t$, we induce such interaction through the following two principles.

1. At each time t, the capacity of the kth HeNB C_k^t should not exceed its demand level d_k^t, as it leaves less resource to other HeNBs.

2. Each HeNB should minimize its transmission power, as higher transmission power decreases the performance of other HeNBs by introducing higher interferences.

The first principle is satisfied by introducing the second component of the local utility function where a penalty is introduced if the choice of resources of a HeNB exceeds its demand level. The details of the second component

of the local utility function are shown as follows:

$$U_k[2](\mathbf{p}_k^t) = -\frac{1}{d_k^t}\left(C_k^t - d_k^t\right)^+, \tag{4.7}$$

where $(x)^+$ denotes the operation $\max(x, 0)$. The second local utility component suppresses greedy HeNB behaviour, and it brings C_k^k closer to its demand d_k^t. This local utility component also helps to maintain the fairness among all the HeNBs.

The third component of the local utility function is used to implement the second principle by considering power as part of the cost of a HeNB, thus, encouraging each HeNB to minimize its power consumption. The details of the third component are shown as follows:

$$U_k[3](\mathbf{p}_k^t) = -\sum_{f=1}^{N_{henb}^t} p_k^t(f). \tag{4.8}$$

Note that we assume unit power transmission of all the HeNBs on each of the RB in our system model. In this system model, each HeNB is assumed to be a selfish user aiming to maximize its own utility function with the minimum cost. By including the power consumption cost as part of the local utility function, it helps the local utility to represent the global utility.

Based on the above definitions, the local utility function of a HeNB k can be defined as follows.

$$U_k(\mathbf{p}_k^t) = U_k[1](\mathbf{p}_k^t) + \alpha_2 \cdot U_k[2](\mathbf{p}_k^t) + \alpha_3 \cdot U_k[3](\mathbf{p}_k^t), \tag{4.9}$$

where α_2 and α_3 are the weighting factors introduced to combine the three utility components assuming $U_k[1]$ has a unit weighting factor, i.e., $\alpha_1 = 1$. These weighting factors are necessary because the actual effect of each component is unknown. By carefully adjusting the values of (α_2, α_3), we can change the effect of each of the three utility components, which then enable the local utility to mimic the behaviour of the global utility (4.3) in the best way. Thus, the question remains to be how to choose (α_2, α_3) which lead to the best overall system performance (4.3). This chapter does not provide a closed form solution to this question, instead, we choose the weight factors according to the numerical studies in (4.4). The selected weighting factors (α_2, α_3) does not ensure the maximization of the instantaneous global utility function at time t, instead, it maximizes the expected system performance under different channel realizations, i.e., $\mathbb{E}_{s_1^t, s_2^t, \ldots s_K^t}\{U_G(\mathbf{P^t})\}$.

Recall that a cognitive HeNB k, $\forall k \in \mathcal{K}$, tries to maximize its utility function $U_k(\mathbf{p}_k^t)$ selfishly by choosing the action vector $\mathbf{p}_k^t$ at the beginning of each time slot. In the following sections, we will show the existence of the correlated equilibrium solution, given the distributed decision making process of HeNBs in a static environment.

4.3 Correlated Equilibrium Solutions with a Game-Theoretic Approach

This section uses game-theoretic approach to formulate the resource allocation among cognitive femtocell base stations (HeNBs) in a static environment, each of HeNB is formulated as a selfish game player aiming to maximize its local utility function (4.9). A static environment is where the

system parameters (e.g., the channel statistics $\mathbf{s}_k^t$ ($k \in \mathcal{K}$), the primary base station behaviour and the number of HeNBs K) are constants or slowly evolve with time. We investigate the correlated equilibrium solution of this static game, which can be obtained via a distributed RB access algorithm. The algorithm is an adaptive variant of the regret matching procedure of [36]. It dynamically adapts the behaviour of HeNBs to time varying environment conditions. We also prove the RB access algorithm converges to the correlated equilibrium set of the defined game.

4.3.1 Definition of Correlated Equilibrium

In a K-player (HeNB) game set-up, each HeNB k ($k \in \mathcal{K}$) is a selfish game player aiming to devise a rule for selecting an action vector $\mathbf{p}_k^t$ at each time slot to maximize (the expected value of) its utility function $U_k(\mathbf{p}_k^t)$. Since each player only has control over its own action $\mathbf{p}_k^t$, the optimal action policy depends on the rational consideration of the action policies from other users. We focus on the *correlated equilibrium* solution of the considered game [7, 8], this solution is an important generalization of the Nash equilibrium and is defined as follows.

Definition 4.3.1 *Define a joint policy $\boldsymbol{\pi}$ to be a probability distribution on the joint action space $\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2 \ldots \mathcal{P}_K$. Given actions of other players $\mathbf{p}_{-k}^t$, the policy $\boldsymbol{\pi}$ is a correlated equilibrium, if for any alternative policy $\widehat{\mathbf{p}}_k^t \in \mathcal{P}_k$, it holds that,*

$$\sum_{\mathbf{p}_{-k}^t \in \mathcal{P}_{-k}} \boldsymbol{\pi}(\mathbf{p}_{-k}^t, \mathbf{p}_k^t) U_k(\mathbf{p}_{-k}^t, \mathbf{p}_k^t) \geq \sum_{\mathbf{p}_{-k}^t \in \mathcal{P}_{-k}} \boldsymbol{\pi}(\mathbf{p}_{-k}^t, \mathbf{p}_k^t) U_k(\mathbf{p}_{-k}^t, \widehat{\mathbf{p}}_k^t). \quad (4.10)$$

Correlated equilibrium can be intuitively interpreted as if $\boldsymbol{\pi}$ provides the K players a strategy recommendation from the trusted third-party. The implicit assumption is that the $K-1$ other players follow this recommendation, and player k ask itself whether it is of its best interest to follow the recommendation as well. The equilibrium condition states that there is no deviation rule that could award player k a better expected utility than $\boldsymbol{\pi}$. Any Nash equilibrium can be represented as a correlated equilibrium when users can generate their recommendations independently. One of the advantages of using correlated equilibrium is that it permits coordination among users, generally through observation of a common signal, which leads to an improved performance over a Nash equilibrium [8].

4.3.2 Decentralized RB Access Algorithm

The RB access algorithm is an adaptive extension of the regret matching procedure [36], it enables HeNBs to adapt their policies to time varying system environment. Let $\mathbf{H}_k(\mathbf{p}^t)$ denote the regret matrix of the kth cognitive HeNB at time t, it is of size $|\mathcal{P}_k| \times |\mathcal{P}_k|$ with its $(|\mathbf{i}|, |\mathbf{j}|)$th entry $(\mathbf{i}, \mathbf{j} \in \mathcal{P}_k)$ specified as,

$$H_k^{(|\mathbf{i}|,|\mathbf{j}|)}(\mathbf{p}^t) = 1_{(\mathbf{p}_k^t = \mathbf{i})} \times [U_k(\mathbf{j}, \mathbf{p}_{-k}^t) - U_k(\mathbf{i}, \mathbf{p}_{-k}^t)], \tag{4.11}$$

where $1_{(\cdot)}$ is an indicator function. Furthermore, we define $\boldsymbol{\theta}_k^t$ to be the overall regret matrix of HeNB k and it is also of size $|\mathcal{P}_k| \times |\mathcal{P}_k|$. The regret value $\boldsymbol{\theta}_k^t(|\mathbf{j}|, |\mathbf{i}|)$ measures the average gain of user k at time t if k had chosen action $\mathbf{i}$ in the past (from time 0 to t) instead of $\mathbf{j}$. If the gain is positive, k is

more likely to switch to action $\mathbf{i}$ in the future, otherwise, k is more likely to stay with $\mathbf{j}$. Specifically, policy transition matrix is a function of the current regret matrix $\boldsymbol{\theta}_k^{t-1}$ as specified in (4.12). Note this regret based scheme requires users to know the reward for each action, even if that action is not taken. Users updates its policy based on the calculated regret matrix (4.12). An RB access algorithm is proposed to compute the correlated equilibrium resource allocation policy, the details of which are listed in Algorithm 5.

Algorithm 5 LTE Cognitive HeNB RB Access Algorithm

Step 1 Set $t = 0$; Initialize $\mathbf{p}^0$ and set $\boldsymbol{\theta}_k^0 = \mathbf{H}_k(\mathbf{p}^0)$ for $\forall k \in \mathcal{K}$.
Step 2
for $t = 1, 2, 3, \ldots$ **do**
Action Update: For $\forall k \in \mathcal{K}$, choose $\mathbf{p}_k^t = \mathbf{i}$ with probability

$$\mathbb{P}(\mathbf{p}_k^t = \mathbf{i} \mid \mathbf{p}_k^{t-1} = \mathbf{j}, \boldsymbol{\theta}_k^{t-1}) = \begin{cases} \frac{\max(\theta_k^{t-1}(|\mathbf{j}|,|\mathbf{i}|),0)}{\mu} & \text{if } \mathbf{i} \neq \mathbf{j} \\ 1 - \sum_{\mathbf{i} \neq \mathbf{j}} \frac{\max(\theta_k^{t-1}(|\mathbf{j}|,|\mathbf{i}|),0)}{\mu} & \text{if } \mathbf{i} = \mathbf{j} \end{cases} \quad (4.12)$$

Regret Matrices Update: Based on the new action, the overall regret matrices are updated according to the following stochastic approximation algorithm with step size ε^t.

$$\boldsymbol{\theta}_k^t = \boldsymbol{\theta}_k^{t-1} + \varepsilon^t \cdot (\mathbf{H}_k(\mathbf{p}^t) - \boldsymbol{\theta}_k^{t-1}), \quad \forall k \in \mathcal{K}. \quad (4.13)$$

end for

Algorithm 5 can be summarized as follows. Step 1 initializes the system by setting time index $t = 0$ and the initial values of $\mathbf{p}^0$ and $\boldsymbol{\theta}_k^0$. Step 2 is the main iteration of the algorithm which is composed of two parts: actions update and regret matrices update. A HeNB k chooses its action for time slot t according to the current action $\mathbf{p}_k^{t-1}$ and the regret matrix $\boldsymbol{\theta}_k^{t-1}$. In (4.12), μ is a constant parameter which is chosen to be $\mu \geq \sum_{\mathbf{i} \neq \mathbf{j}} \max(\theta_k^{t-1}(|\mathbf{j}|, |\mathbf{i}|), 0)$ to ensure the probabilities are non-negative. The choice of step size ε^t used in

(4.13), can be either a decreasing step size $\varepsilon^t = 1/t$ or a small constant step size $\varepsilon^t = \varepsilon$ $(0 < \varepsilon \ll 1)$. If system parameters do not evolve with time, using decreasing step size convergences the algorithm to the correlated equilibrium set with probability one. Using a constant step size enables the algorithm to track of the correlated equilibrium set if system parameters slowly evolve with time.

Depending on the selection of the step size ε^t, the regret matrix $\boldsymbol{\theta}_k^t$ can be rewritten as follows.

$$\begin{aligned} \boldsymbol{\theta}_k^t(|\mathbf{j}|,|\mathbf{i}|) &= \frac{1}{t}\sum_{l\leq t,\mathbf{p}_k^l=\mathbf{j}} \big(U_k(\mathbf{i},\mathbf{p}_{-k}^l) - U_k(\mathbf{j},\mathbf{p}_{-k}^l)\big), \text{ if } \varepsilon^t = \frac{1}{t}; \\ \boldsymbol{\theta}_k^t(|\mathbf{j}|,|\mathbf{i}|) &= \sum_{l\leq t,\mathbf{p}_k^l=\mathbf{j}} \varepsilon(1-\varepsilon)^{t-l}\big(U_k(\mathbf{i},\mathbf{p}_{-k}^l) - U_k(\mathbf{j},\mathbf{p}_{-k}^l)\big), \text{ if } \varepsilon^t = \varepsilon. \end{aligned} \quad (4.14)$$

The above RB access algorithm is a modification of the regret matching procedure in [36]. In the regret matching approach, the decisions of each HeNB is based on the average history of all past observed performance results. This choice, however, is not desirable in our scenario since the system parameters may vary over time. Instead, our algorithm adapts the regret matrices $\boldsymbol{\theta}_k^t$ $(k \in \mathcal{K})$ according to the updated system parameters which captures the time-varying nature of the system. The decentralized feature of this RB access algorithm permits its implementation among the distributed femtocells in a 3GPP LTE network.

Regret matrix $\boldsymbol{\theta}_k^t$ is one of the key parameters in Algorithm 1, based on which a cognitive HeNB adjusts its future action. The interpretation of the $(|\mathbf{i}|,|\mathbf{j}|)$th entry of $\boldsymbol{\theta}_k^t$ is that measures the average gain that a cognitive

HeNB k would have, had it chosen action $\mathbf{j}$ in the past (i.e, in the $(t-1)$th time slot) instead of $\mathbf{i}$.

Remark: RB access algorithm (Algorithm 5) is a distributed algorithm, the k user is only required to know its own action and interference at each iteration. IUI $I_k^t(f)$ is obtained by channel estimation through downlink common pilots. $I_k^t(f)$ in (4.2) provides the mathematical formulation of IUI.

4.3.3 Convergence of RB Access Algorithm

By using the result from [9], we introduce Theorem 4.3.2 which proves that the RB access algorithm (Algorithm 5) converges to the correlated equilibrium under certain conditions. Let $\boldsymbol{\theta}_k$ denote the regret matrix of the kth HeNB when $t \to \infty$. We use $\Gamma_\Omega(\boldsymbol{\theta}_k)$ to denote the projection of parameter $\boldsymbol{\theta}_k$ on Ω, where Ω is the closed negative orthant of $\mathbb{R}^{|\mathbf{p}_k|\times|\mathbf{p}_k|}$. Let $< x, y >$ denote the inner product of x and y.

Theorem 4.3.2 *The RB access algorithm (Algorithm 5) is ensured to converge to the correlated equilibrium set of the formulated game.*

Proof: By using Proposition 3.8 and Corollary 5.7 in [9], we know if every HeNB follows the strategy in Algorithm 5, it is enough to prove that the following inequality, given by (4.15), holds in order to prove that Algorithm 5 converges to the set of correlated equilibria of the defined game.

$$< \boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k^{'} - \Gamma_\Omega(\boldsymbol{\theta}_k) >\leq 0. \tag{4.15}$$

□

Condition (4.15) is originated from the Blackwell's sufficient condition for approachability [14]. Therefore, we only need to demonstrate that (4.15) holds in order to prove the convergence of RB access algorithm.

Note that the negative orthant Ω is a convex set. The left hand side of (4.15) can be expressed as,

$$\begin{aligned} & < \boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k^{'} - \Gamma_\Omega(\boldsymbol{\theta}_k) > \\ = \quad & < \boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k^{'} > - < \boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k), \Gamma_\Omega(\boldsymbol{\theta}_k) >, \end{aligned} \tag{4.16}$$

where $< \boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k), \Gamma_\Omega(\boldsymbol{\theta}_k) >= 0$ due to the definition of projection. Thus, in order to establish (4.15) we need to prove $< \boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k^{'} > \leq 0$.

Let us construct a Markov chain with the transition probability specified in (4.12) and use $\tau_{|\mathbf{i}|}$ ($\mathbf{i} \in \mathcal{P}_k$) to denote the stationary distribution of such a Markov chain. $\tau_{|\mathbf{i}|}$ can be specified as follows.

$$\tau_{|\mathbf{i}|} = \sum_{|\mathbf{j}| \neq |\mathbf{i}|} \tau_{|\mathbf{j}|} \frac{\theta_k^+(|\mathbf{j}|, |\mathbf{i}|)}{\mu} + \tau_{|\mathbf{i}|} \cdot \big(1 - \sum_{|\mathbf{i}| \neq |\mathbf{j}|} \frac{\theta_k^+(|\mathbf{i}|, |\mathbf{j}|)}{\mu}\big), \tag{4.17}$$

where μ is a constant chosen to be $\mu > \sum_{|\mathbf{j}| \neq |\mathbf{i}|} \theta_k^+(|\mathbf{j}|, |\mathbf{i}|)$. Then, (4.17) is equivalent to the following equation.

$$\sum_{|\mathbf{j}| \neq |\mathbf{i}|} \tau_{|\mathbf{j}|} \theta_k^+(|\mathbf{j}|, |\mathbf{i}|) = \tau_{|\mathbf{i}|} \sum_{|\mathbf{j}| \neq |\mathbf{i}|} \theta_k^+(|\mathbf{i}|, |\mathbf{j}|). \tag{4.18}$$

Since the projection is on the negative orthant Ω, $\boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k) = \boldsymbol{\theta}_k^+$. $< \boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k^{'} >$ can then be written as

$$
\begin{aligned}
& < \boldsymbol{\theta}_k - \Gamma_\Omega(\boldsymbol{\theta}_k), \boldsymbol{\theta}_k^{'} > = < \boldsymbol{\theta}_k^+, \boldsymbol{\theta}_k^{'} > \\
= & \sum_{\mathbf{j}} \sum_{\mathbf{i} \neq \mathbf{j}} \theta_k^+(|\mathbf{j}|, |\mathbf{i}|) \big[U_k(\mathbf{i}, \mathbf{p}_{-k}) - U_k(\mathbf{j}, \mathbf{p}_{-k}) \big] \tau_{|\mathbf{j}|} \\
= & \sum_{\mathbf{i} \neq \mathbf{j}} \sum_{\mathbf{j}} \theta_k^+(|\mathbf{i}|, |\mathbf{j}|) U_k(\mathbf{j}, \mathbf{p}_{-k}) \tau_{|\mathbf{i}|} - \sum_{\mathbf{j}} \sum_{\mathbf{i} \neq \mathbf{j}} \theta_k^+(|\mathbf{j}|, |\mathbf{i}|) U_k(\mathbf{j}, \mathbf{p}_{-k}) \tau_{|\mathbf{j}|} \\
= & \sum_{\mathbf{j}} \left[\sum_{\mathbf{i} \neq \mathbf{j}} \theta_k^+(|\mathbf{i}|, |\mathbf{j}|) \tau_{|\mathbf{i}|} - \sum_{\mathbf{i} \neq \mathbf{j}} \theta_k^+(|\mathbf{j}|, |\mathbf{i}|) \tau_{|\mathbf{j}|} \right] U_k(\mathbf{j}, \mathbf{p}_{-k}) \\
= & \ 0. \qquad (4.19)
\end{aligned}
$$

Therefore, the condition stated in (4.15) is proved to hold. This concludes the proof of Theorem 4.3.2.

4.3.4 Correlated Equilibrium under Dynamic Environments and Curse of Dimensionality

In the case that system contains large number of active mobile users, it causes high dynamic macro base station behaviour. In which scenario, the problem can be described as: resource allocation among femtocells (HeNBs) in an OFDMA LTE downlink system under a dynamic environment where the resource occupancy behaviour of the primary base station (system state) is varying quickly while other system parameters (e.g., number of HeNBs) are constants or evolving slowing.

By formulating the dynamic of system state as a Markov chain, it requires Markov game-theoretic approach to formulate the resource allocation problem among femtocell base stations (HeNBs) as a Markov game. Dif-

ferent from static game, system state and state transition probabilities are important elements in dynamic games as they abstract the time-varying nature of a dynamic environment. A reasonable choice of the system state is $[N^t_{henb}, \mathbf{s}^t_1, \ldots, \mathbf{s}^t_K]$ which is composed of the number of available RBs for HeNBs and the channel states of HeNBs. By defining the correlated equilibrium or Nash equilibrium of such a dynamic stochastic game, different optimization algorithms can be used to compute the equilibrium transmission polices. E.g., [43] proposed iterative value optimization algorithm and stochastic approximation algorithm to compute the Nash equilibrium policies in the formulated Markovian game.

Potential applications notwithstanding, there remains substantial hurdles in the application of dynamic stochastic games as a modelling tools in practice. Discrete-time stochastic games with finite number of states are central to the analysis of strategic interactions among selfish HeNBs in dynamic environment. The usefulness of discrete-time games, however, is limited by their computational burden; in particular, there is "curse of dimensionality". In a discrete-time dynamic stochastic game, each game player (HeNB) is distinguished by an individual state at each time slot. The system state is a vector encoding the number of players with each possible value of the individual state variable $[N^t_{henb}, \mathbf{s}^t_1, \ldots, \mathbf{s}^t_K]$. The system state is exponential to the number of the HeNBs in the system. How to efficiently reduce the state space is yet an issue to solve before the implementation of stochastic games in LTE systems. One direction of research is to consider continuous-time stochastic game models. E.g., [27] aims to reduce the dimensionality by exploring the alternative continuous-time stochastic

games with a finite number of states and show that the continuous time has substantial advantages.

4.4 Numerical Examples

Algorithm 5 is designed to compute the correlated equilibrium policies for cognitive HeNBs in the downlink of an OFDMA LTE system. This section illustrates the performance of the RB access algorithm (Algorithm 5) in a game set-up. The RB access algorithm is highly scalable and is able to handle large number of users. We consider $K = 6$ HeNBs in this section just for demonstration purpose. For the kth HeNB ($k \in \mathcal{K}$), its channel quality at any RB f ($f \in \mathcal{F}$) is $s_k(f) \in \{1, 2, 3\}$, its demand level belongs to the set $d_k \in \{10, 20, 30, 40\}$. The action of the kth HeNB at the fth RB is specified as $p_k(f) \in \{0, 1\}$, where 0 represents no transmission and 1 represents transmit. In our simulation set-up, we specify the noise covariance to be $\sigma_k^2(f) = 0.1$ ($\forall k \in \mathcal{K}$ and $\forall f \in \mathcal{F}$). In this simulation model, the number of available RB for HeNBs is fixed to be $N_{henb} = 10$.

In first example, we are going to investigate the impact of the pricing parameters (α_2, α_3) in the local utility function (4.9) on different global utilities (4.3) and (4.5). The pricing parameters is chosen to ensure the global system performance. This is an off-line calculation procedure. The simulation results in Fig. 4.2 and Fig. 4.3 are averaged over 1000 iterations, where 20 different scenarios with different noise $n_k(f)$, channel states $\mathbf{s}_k$ are being considered in each iteration. In both figures, x-axis and y-axis denote α_2 and α_3, respectively. Z-axis denotes the global system performance specified in (4.3) in Fig. 4.2, while it denotes the system average performance

specified in (4.5) in Fig. 4.3.

Based on Fig. 4.2, we notice the completely selfish behaviours from HeNBs ($\alpha_2 = 0, \alpha_3 = 0$) do not ensure the optimum of global system performance; while $\alpha_2 = 5$, $\alpha_3 = 0.25$ lead the least satisfaction level among HeNBs to 0.9, i.e., $U_G(\mathbf{p}^t) = 0.9$. We will specify $\alpha_2 = 5$ and $\alpha_3 = 0.25$ in the simulation for Fig. 4.4.

Fig. 4.3 shows the system average performance (4.5) is less sensitive to the change of pricing parameters (α_2, α_3). System average performance is of similar level when the pricing parameters are of the range $0 \leq \alpha_2 \leq 7$, $0.05 \leq \alpha_3 \leq 0.5$. Thus, in the case we choose (4.5) as the global utility function, the selection of (α_2, α_3) is not unique, they can be of any values within the above range. It can also be noticed from Fig. 4.3 that the system average performance decrease dramatically when $\alpha_2 > 7$. This can be explained as follows: if the power consumption cost is greater than a certain threshold, a HeNB will choose not to transmit as the payoff is less than the cost. Thus, a very high power consumption cost weighting factor (α_3) can have a negative effect on HeNB performances.

The next example (Fig. 4.4) compares the performance of the proposed RB access algorithm with the existing "Best Response" algorithm with a global utility function specified in (4.3). In the simulation, we use small constant step size and it is specified as $\varepsilon^t = \varepsilon = 0.05$. The results are averaged over 50 scenarios and there are 2000 iterations.

Best Response is a simple case where each HeNB chooses its action $\mathbf{p}_k^t$ at each time slot solely maximizing its local utility (4.9) and each HeNB assumes that the actions of other HeNBs are fixed. Thus, Best Response is

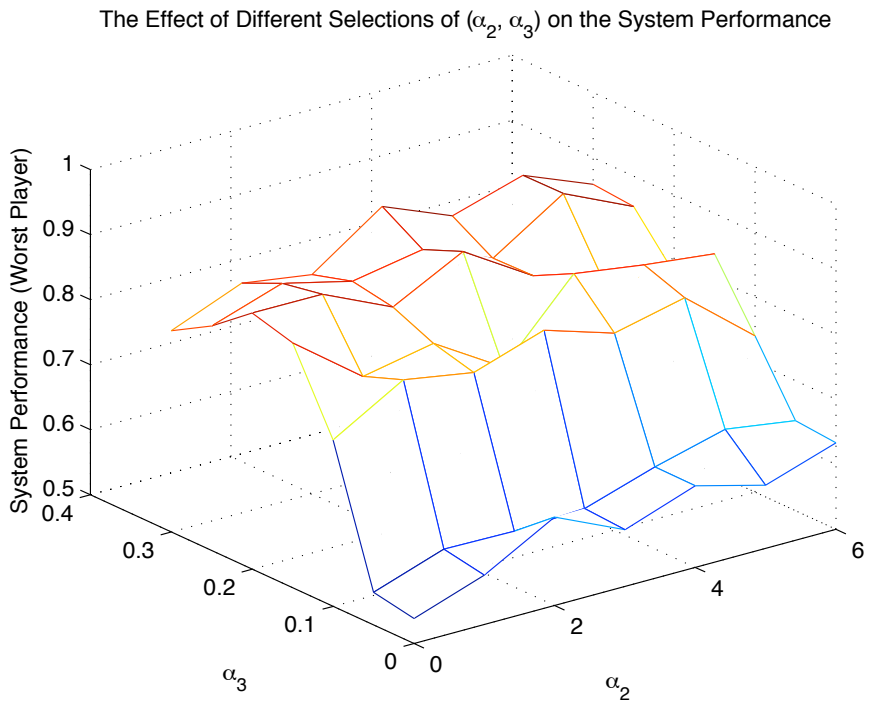

Figure 4.2: The effect of different values of (α_2, α_3) defined in (4.9) on the global system performance specified in (4.3) .

a special case of the proposed RB access algorithm with the step size chosen to be $\varepsilon^t = 1$. The action update in the Best Response is not a function of the previous regret $\boldsymbol{\theta}_k^{t-1}$ but only a function of the current instantaneous regret matrix $\mathbf{H}_k(\mathbf{p}^t)$. From Fig. 4.4 we can see that the system performance using the RB access algorithm reaches 0.9 after 1400 iterations and the result from the "Best Response" algorithm stays at around 0.6. It can be seen that the RB access algorithm (Algorithm 5) improves the system performance greatly compared to the "Best Response" algorithm. We can also observe from Fig. 4.4 that both RB access algorithm and "Best Response" algorithm

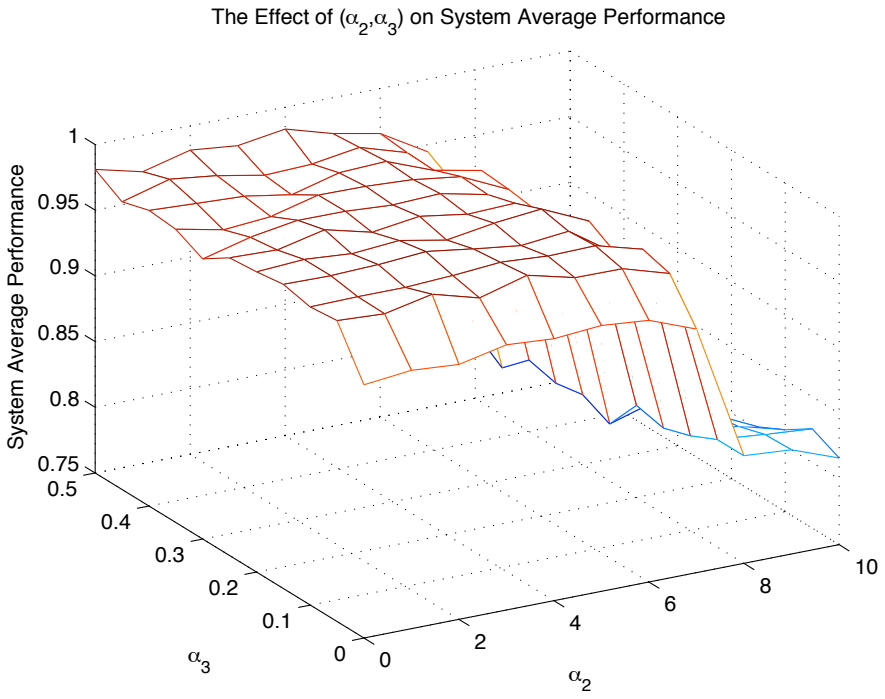

Figure 4.3: The effect of different values of (α_2, α_3) defined in (4.9) on the global system average performance specified in (4.5).

do not converge to constant values, it is because transmission policy $\mathbf{p}^t$ converges to a correlated equilibrium set when $t \to \infty$ and the correlated equilibrium set has more than one correlated equilibrium policy.

4.5 Summary

We have proposed implementation of cognitive femtocell base stations for resource block allocation in the downlink of a eNB macrocell 3GPP LTE system. By considering the eNB as the primary base station, the HeNBs are formulated as multiple secondary base stations competing for spectrum

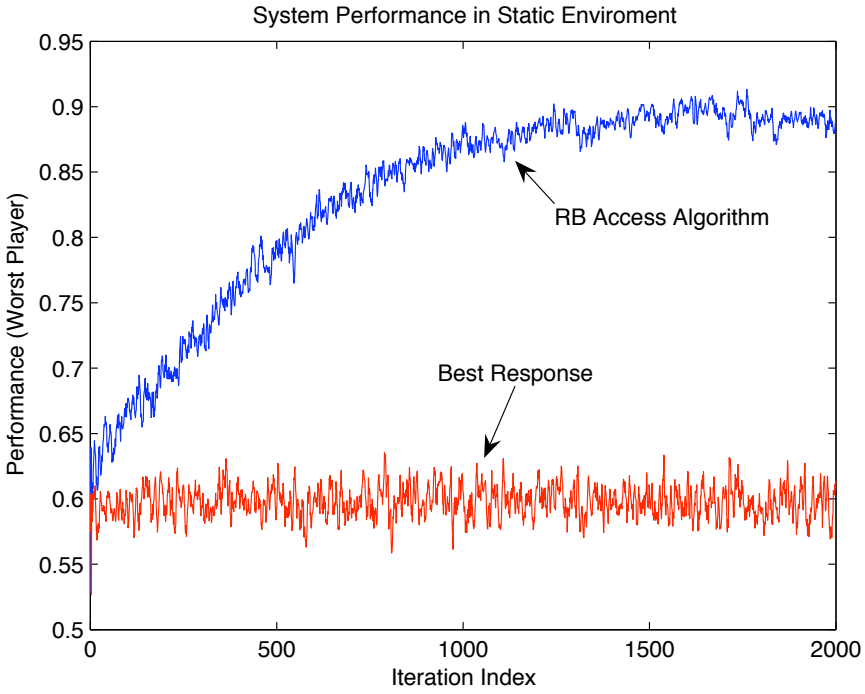

Figure 4.4: Performance comparison between RB access algorithm (Algorithm 5) and the "Best Response" algorithm.

resources. The RB allocation problem is formulated in a static environments, using static game framework. An RB access algorithm is proposed to compute the correlated equilibrium policy in such a environment. We also prove that the RB access algorithm converges to the correlated equilibrium set of the formulated game. Numerical examples are used to verify the performances of the proposed algorithm.

Chapter 5

Application of Mechanism Design in Opportunistic Scheduling under Cognitive Radio Systems [1]

The conventional opportunistic scheduling algorithm in cognitive radio networks does the scheduling among the secondary users based on the reported state values. However, such opportunistic scheduling algorithm can be challenged in a system where each secondary user belongs to a different independent agent and the users work in competitive way. In order to optimize his own utility, a selfish user can choose not to reveal his true information to the

[1] This chapter is based on the following publication. J. W. Huang and V. Krishnamurthy, "Game Theoretical Issues in Cognitive Radio Systems," *Journal of Communications*, vol. 4, no. 10, pp. 790-802, November 2009. (**Invited Paper**)

central scheduler. In this chapter, we proposed a pricing mechanism which combines the mechanism design with the opportunistic scheduling algorithm and ensures that each rational selfish user maximizes his own utility function, at the same time optimizing the overall system utility. The proposed pricing mechanism is based on the classic VCG mechanism and has several desirable economic properties. A mechanism learning algorithm is proposed for users to learn the mechanism and to obtain the Nash equilibrium policy. A numerical example shows the Nash equilibrium of the proposed algorithm achieves system optimality.

5.1 Background

An efficient spectrum assignment technology is essential to a cognitive radio system, which allows secondary users to opportunistically utilize the unoccupied spectrum holes based on agreements and constraints. These secondary users have to coordinate with each other in order to maintain the order and result in maximum spectrum efficiency. This motivates the development of spectrum access approaches in cognitive radio systems. An opportunistic scheduling in cognitive networks which assumes that the scheduling is aware of primary user transmissions is being considered in [75] and [92]. [17, 88, 95] consider a system where only partial information about the primary user activities is available.

However, all the existing opportunistic scheduling approaches overlook the fact that the secondary users may be owned by different agents and they may work in competitive rather than cooperative manners. These selfish users can become so sophisticated that they lie about their states to

optimize their own utility at the cost of reducing the overall system performance. It requires mechanism design theory in order to prevent this from happening. Mechanism design is the study of designing rules for strategic, autonomous and rational agents to achieve predictable global outcome [63]. A milestone in mechanism design is the VCG mechanism, which is a generalization of Vickrey's second price auction [89] proposed by Clark [20] and Groves [35]. The particular pricing policy of the VCG mechanism makes the true reporting values the dominant strategy of the buyers. Our goal is to model each user as a selfish agent whose aim is to optimize his own utility function and we try to find a pricing mechanism which can efficiently allocate the resources within the network.

In this chapter, we consider a cognitive radio network with multiple secondary users. The central scheduler adopts opportunistic scheduling to schedule the users under an overall transmission power constraint. The scheduling is based on the reported states of each user. When the secondary users belong to different independent selfish agents, the users may lie about their true state values in order to optimize their own utilities, sometimes at the cost of reducing the overall system performance.

5.2 Opportunistic Scheduling in Cognitive Radio Systems

This section describes the conventional opportunistic scheduling algorithm in the uplink of a K secondary users cognitive radio system (Fig. 5.1). One or more secondary users can be scheduled for transmission at each scheduling. We assume the system has perfect information about the primary user

activities and the algorithm is applicable when the primary user is absent. As soon as any primary user appears on the desired spectrum band, secondary users will release the spectrum resources.

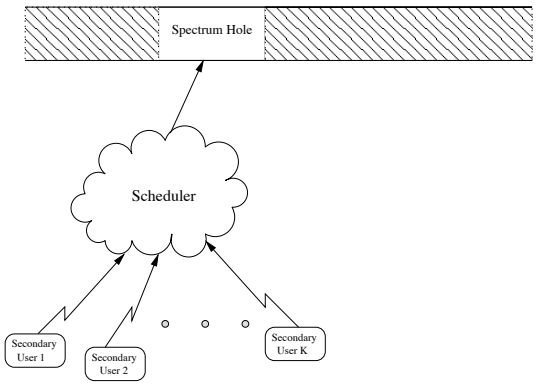

Figure 5.1: A K secondary users cognitive radio where the central scheduler does the scheduling according to the opportunistic accessing algorithm.

5.2.1 System States Description

b_k denotes the level of date in the buffer of user k and $b_k \in$ with L representing the size of the buffer. The composition of buffer state of all the users is $b = \{b_1, b_2, \ldots, b_K\} \in \mathcal{B}$ which $\mathcal{B}$ is the buffer state space.

We use $\hat{b}_k$ to represent the buffer state that the kth user reports to the central scheduler and $\hat{b}_k \in \{0, 1, \ldots, L\}$. Note here that in a truth telling system each user reports the true state value and we have $\hat{b}_k = b_k$, $k = 1, 2, \ldots, K$.

Assume each user has a normalized transmission power, the signal to noise ratio (SNR) of user k is denoted as c_k. The channel status of a user is assumed to be circularly symmetric complex Gaussian random variables

and it is independent from the channel states of other users. After applying quantization, the SNR state is $c_k \in \{1, 2, \dots Q_{c_k}\}$ where Q_{c_k} denotes the maximum quantization level. The composition of the SNR states of all the K secondary users is denoted as $c = \{c_1, c_2, \dots, c_K\} \in \mathcal{C}$ where $\mathcal{C}$ is SNR state space.

Let the symbols per second transmission rate of user k be w_k and its bits per symbol rate be m_k (different values of m_k lead to different modulation schemes), its instantaneous throughput can be expressed as

$$\rho_k = w_k m_k \big(1 - p_e(c_k, m_k)\big)^{s_k}. \tag{5.1}$$

Here, s_k is the average packet size in bits. $p_e(c_k)$ is the transmission bit error rate (BER), it is a function of the SNR and modulation mode of user k. In the case of an uncoded M-ary quadrature modulation (QAM), the BER $p_e(c_k)$ can be approximated as [24]:

$$p_e(c_k, m_k) = 0.2 \times \exp\big[\frac{-1.6 c_k}{2^{m_k} - 1}\big]. \tag{5.2}$$

After applying quantization to the instantaneous throughput ρ_k, we have $\rho_k \in \{0, 1, 2, \dots, Q_\rho\}$ for $k = 1, 2, \dots, K$ where Q_ρ indicating the maximum quantization level of the throughput. In the case that the primary users appear on the system, the throughput state of all cognitive radio user switch to 0, $\rho_k = 0$, and no secondary user is allowed to transmit.

Similarly, $\hat{\rho_k}$ is used to indicate the instantaneous throughput state that user k reports to the central scheduler and $\hat{\rho}_k \in \{0, 1, 2, \dots, Q_\rho\}$.

For notation convenience, in this chapter we use $\theta_k = \{\rho_k, b_k\}$ to rep-

resent the true states of user k and $\hat{\theta}_k$ to represent the reported states of user k. θ_{-k} is used to indicate the true states of all the remaining $K-1$ users, excluding user k, while $\hat{\theta}_{-k}$ is the corresponding reported values. We use $\Theta = \{\theta_k, \theta_{-k}\}$ to denote the combined states of all the K users and $\hat{\Theta} = \{\hat{\theta}_k, \hat{\theta}_{-k}\}$ to denote the combined reported states of all the users.

5.2.2 Conventional Opportunistic Accessing Scheme

Opportunistic accessing scheme is commonly used by central scheduler to assign certain users for transmission with the aim to maximize the total system utility subject to the system constraint. This section simply uses the results of the opportunistic scheduling algorithm but does not describe the details of the algorithm. The decentralized channel access algorithm described in Chapter 2.2.1 is an example of the implementation of the opportunistic scheduling algorithm.

The information that central scheduler has about each user includes the reported buffer state and reported throughput state, which is $\{\hat{\rho}_k, \hat{b}_k\}$ for $k = 1, 2, \ldots, K$. The central scheduler opportunistic accessing scheme is based on the reported buffer and throughput states. Define A as a *feasible* set, it is a subset of the union of all the users, and $A \subseteq \{1, 2, \ldots, K\}$. A feasible set is a set which satisfies the system constraint. There are several choices of the system constraint, however, in this chapter we choose it to be the overall transmission power in the system. The constraint is specified as: the overall transmission power in the system should be equal or less than the system power limit P. As we specified in Section 5.2.1, each user uses a unit transmission power to transmit. Thus, the transmission power constraint on the system can be converted to the constraint on the

total number of users who are transmitting, that is the number of users who are transmitting simultaneously should be less or equal to the system power limit P. The conventional opportunistic accessing scheme with transmission power constraint defines the optimal feasible set of chosen users A^* in the following way:

$$A^* = \arg\max \sum_{k \in A} U_k(\hat{\rho}_k, \hat{b}_k), \tag{5.3}$$

$$s.t. \qquad |A| \leq P. \tag{5.4}$$

$|A|$ is used to denote the number of users in set A and $U_k(\hat{\rho}_k, \hat{b}_k)$ is the corresponding utility of user k with throughput $\hat{\rho}_k$ and buffer state $\hat{b}_k$. One typical choice of the utility function is $U_k(\hat{\rho}_k, \hat{b}_k) = \hat{\rho}_k \hat{b}_k$ [57]. Central scheduler calculate A^* which optimizes the system performance subjects to the system constraint. Note if both set A_1 and A_2 lead to the same optimal result, A^* will be chosen randomly between them.

The conventional opportunistic accessing scheme works when the central scheduler has the true information of each user, which means the $\{\hat{\rho}_k, \hat{b}_k\}$ that the users report to the central scheduler are the true values, $\hat{\rho}_k = \rho_k$ and $\hat{b}_k = b_k$. However, the conventional opportunistic algorithm may be challenged when the secondary users become so sophisticated and are able to reconfigure themselves to make efficient use of the local resources (e.g., manage their own reporting data and have the most efficient data transmission). It is important to design a mechanism to optimize the overall system performance while ensuring the profit of each secondary user.

5.3 The Pricing Mechanism

This section applies the VCG pricing mechanism to the opportunistic scheduling algorithm, the new mechanism enforces the truth revealing property of each user. A mechanism learning algorithm is also proposed to implement the mechanism. It is shown that the Nash equilibrium of the proposed algorithm ensures each user reports its values truthfully.

5.3.1 The Pricing Mechanism

Different from the centralized conventional opportunistic scheduling algorithm, the proposed pricing mechanism is a distributed algorithm where each user tries to maximize his own utility function by choosing the right state values to report to the central scheduler.

The buffer and throughput that user k chooses to report to the central scheduler is a solution of the following optimization problem:

$$\begin{aligned}\{\hat{\rho}_k, \hat{b}_k\} &:= \max_{\hat{\theta}_k} v_k(\theta_k, \hat{\theta}_k, \hat{\theta}_{-k}) \\ &= \max_{\hat{\rho}_k, \hat{b}_k} \alpha^{\rho_k b_k} \times \frac{\prod_{j\in A^*, j\neq k} \alpha^{\hat{\rho}_j \hat{b}_j}}{\prod_{j\in A'} \alpha^{\hat{\rho}_j \hat{b}_j}} I_{k\in A^*} + I_{k\notin A^*}, \end{aligned} \tag{5.5}$$

with the set A^* and $A^{'}$ defined in the following ways:

$$\begin{aligned} A^* &:= \arg\max \sum_{j\in A} \hat{\rho}_j \hat{b}_j, \quad s.t. \quad |A| \leq P; \\ A^{'} &:= \arg\max_{\hat{\rho}_k = 0} \sum_{j\in A, j\neq k} \hat{\rho}_j \hat{b}_j, \quad s.t. \quad |A|_{k\notin A} \leq P. \end{aligned}$$

In the above optimization problem, α is a system fixed constant and it is $\alpha > 1$, $I_{\{.\}}$ is an indication function whose value is 1 when the condition is

true, otherwise, it is 0.

$v_k(\theta_k, \hat{\theta}_k, \hat{\theta}_{-k})$ is the utility function of user k, which is the function of the true states itself, the reported states of user k and the reported states of all the remaining users. If a user is not scheduled for transmission, his utility function equals to 1. If a user is scheduled for transmission, his utility function equals to the first part of (5.5), which is:

$$v_k(\theta_k, \hat{\theta}_k, \hat{\theta}_{-k}) = \alpha^{\rho_k b_k} \times \frac{\prod_{j \in A^*, j \neq k} \alpha^{\hat{\rho}_j \hat{b}_j}}{\prod_{j \in A'} \alpha^{\hat{\rho}_j \hat{b}_j}} \text{ if } k \in A^*; \quad (5.6)$$

$$v_k(\theta_k, \hat{\theta}_k, \hat{\theta}_{-k}) = 1 \text{ if } k \notin A^*. \quad (5.7)$$

The first part of (5.6) $\alpha^{\rho_k b_k}$ can be interpreted as the gain of user k per unit of subcarrier with throughput state ρ_k and buffer state b_k. The second term $\frac{\prod_{j \in A^*, j \neq k} \alpha^{\hat{\rho}_j \hat{b}_j}}{\prod_{j \in A'} \alpha^{\hat{\rho}_j \hat{b}_j}}$ can be interpreted as the number of unit of subcarrier that user k will be allocated and it is a function of the state of the remaining users in the system. In other words, the inverse of the second term could be interpreted as the price that user k has to pay to the system if it is scheduled for transmission. Each user select $\{\hat{\rho}_k, \hat{b}_k\}$ to report to the central scheduler in the aim to maximize its own utility.

There is one condition necessary in order to achieve an efficient allocation policy among selfish agents [21, 66]: if a user k ($k = 1, 2, \ldots, K$) reports a false state values $\hat{\theta}_k \neq \theta_k$ results in the same value of the utility function as that of if it reports the true value, which is $v_k(\theta_k, \hat{\theta}_k, \hat{\theta}_{-k}) = v_k(\theta_k, \theta_k, \hat{\theta}_{-k})$, $\hat{\theta}_k \neq \theta_k$, then the user will choose to report the true values. We name this as the *truth preferred rule* in this chapter. The interpretation of this rule is that when lying about the states does not bring any benefit

to a user, a user would prefer telling the truth.

The pricing mechanism we propose above is based on the VCG mechanism, where we modified the conventional summation form of the utility function into a product form. Such pricing mechanism can be easily related to a practical cognitive radio system and interpret the utility function in terms of physical parameters in a practical system.

5.3.2 Economic Properties of the Pricing Mechanism

The pricing mechanism we proposed above still maintains the same desirable economic properties as that of VCG mechanism, these properties are specified as follows [22, 23]:

1. The mechanism is incentive-compatible in ex-post Nash equilibrium. The best response strategy is to reveal the true state information $\hat{\theta}_k = \theta_k$ even after they have complete information about other users θ_{-k}.

2. The mechanism is individually rational. A selfish agent will join the mechanism rather than choosing not to, because the value of the utility function is non-negative.

3. The mechanism is efficient. Since all the users will truthfully reveal their state information, the opportunistic scheduling algorithm carried out by the central scheduler will maximize the system performance.

The detailed proof of these economic properties is shown as follows.

Proposition 5.3.1 *The mechanism is incentive-compatible in ex-post Nash equilibrium.*

A mechanism is *incentive-compatible* in ex-post Nash equilibrium if the best strategy for a user is to report truthfully even with complete state information of other users.

Proof: Suppose all users except user k report their true state values, $\hat{\theta}_{-k} = \theta_{-k}$. Assume the optimal allocations set when the kth user reports truthfully is A_0^*, we are going to analyze the utility function of user k by reporting $\hat{\theta}_k \neq \theta_k$.

If the kth user reports $\hat{\theta}_k \neq \theta_k$ but the optimal allocation is the same as that when he reports the true values, $A^* = A_0^*$, this will then result in equal value of the utility function, which is $v_k(\theta_k, \hat{\theta}_k, \hat{\theta}_{-k}) = v_k(\theta_k, \theta_k, \hat{\theta}_{-k})$. According to the *truth preferred rule*, user k will choose to report the true values.

Consider the case that if user k reports the true values θ_k, $k \notin A_0^*$, but if user k reports $\hat{\theta}_k \neq \theta_k$, $k \in A^*$. The relation between the reported values and the true values is: $\hat{\rho}_k \hat{b}_k > \rho_k b_k$. With reporting value $\hat{\theta}_k$, the utility function is given as (5.6). Since $k \notin A_0^*$, this implies that $\rho_k b_k \leq \min_{j, j \in A'}(\hat{\rho}_j \hat{b}_j)$, and it will in turn result in $v_k(\theta_k, \hat{\theta}_k, \hat{\theta}_{-k}) \leq 1$. Thus, user k will choose to report truthfully θ_k in this case.

Now we are going to consider the case that if user k reports the true values θ_k, $k \in A_0^*$, but if user k reports $\hat{\theta}_k \neq \theta_k$, $k \notin A^*$. The relation between the reported values and the true values is: $\hat{\rho}_k \hat{b}_k < \rho_k b_k$. The value of the utility function of user k with reporting values $\hat{\theta}_k$ equals to 1 according to the expression (5.7). The utility function if user k reports truthfully, $v_k(\theta_k, \theta_k, \hat{\theta}_{-k})$, is given according to that in (5.6). $k \in A_0^*$ implies that $\rho_k b_k \geq \hat{\rho}_j \hat{b}_j$ with $j \in A^{'}, j \notin A^*$, so we have $v_k(\theta_k, \theta_k, \hat{\theta}_{-k}) \geq 1$. Thus,

user k will choose to report its true values as well.

Proposition 5.3.2 *The mechanism is individually rational.*

A mechanism is *individually rational* if there is an incentive for users to join the system rather than opting out of it. Assume that the utility a user achieves by not joining the mechanism is 0, then we only have to prove the utility a user gains in this pricing mechanism is always ≥ 0.

Proof: The expression of the utility function of any user k is given in (5.5). If a user is not scheduled for transmission the utility function equals to 1. If it is scheduled, the utility function is always greater or equal to 0 as the utility function is an exponentiation function of a constant α, $\alpha > 1$.

Proposition 5.3.3 *The mechanism is efficient.*

This implies that the opportunistic scheduling algorithm carries out by the central scheduler schedules according to:

$$A^* = \arg\max \sum_{k \in A} U_k(\rho_k, b_k), \tag{5.8}$$

$$s.t. \quad |A| \leq P. \tag{5.9}$$

Proof: Using the result from Proposition 5.3.1, each user reports truthfully under this pricing mechanism, the opportunistic scheduling algorithm is able to achieve system efficiency.

5.3.3 Mechanism Learning Algorithm

In the case that the users in the cognitive radio are not fully acknowledged with the pricing mechanism, we propose a mechanism learning algorithm

for each user to learn the mechanism and obtain the Nash equilibrium state. The mechanism learning algorithm is shown in Algorithm 6.

Algorithm 6 Mechanism Learning Algorithm

1: **Initialization**: $\hat{\rho}_k^0$ and $\hat{d}_k^0$ for k=1, 2, ..., K;
2: **Main Iteration**
3: **for** $k = 1 : K$ **do**
4: $\{\hat{\rho}_k^{n+1}, \hat{b}_k^{n+1}\} :=$
5: $\max_{\hat{\rho}_k, \hat{b}_k} \alpha^{\rho_k b_k} \times \frac{\prod_{j \in A^*, j \neq k} \alpha^{\hat{\rho}_j^n \hat{b}_j^n}}{\prod_{j \in A'} \alpha^{\hat{\rho}_j^n \hat{b}_j^n}} I_{k \in A^*} + I_{k \notin A^*}$.
6: With A^* and $A^{'}$ defined by:
7: $A^* := \arg\max_A \sum_{j \in A} \hat{\rho}_j^n \hat{b}_j^n$,
8: $s.t.\ \ |A| \leq P$;
9: $A^{'} := \arg\max_{A, \rho_k = 0} \sum_{j \in A, j \neq k} \hat{\rho}_j^n \hat{b}_j^n$,
10: $s.t.\ \ |A|_{k \notin A} \leq P$;
11: **end for**
12: $n = n + 1$;
13: The iteration terminates when the parameters $\hat{d}_k^n, \hat{\rho}_k^n$ converge; else return back to Step 3.

The mechanism learning algorithm can be summarized as follows. First, each user randomly choose a initial reporting state $\{\hat{\rho}_k^0, \hat{d}_k^0\}$. The main iteration is from Step 2 to Step 9 where each user optimizes the reporting state $\{\hat{\rho}_k^{n+1}, \hat{b}_k^{n+1}\}$ iteratively with the states of other users fixed. The reporting state of each user is chosen based on maximizing its own utility function. Note here that at each step, each user follows the *truth preferred rule*. The algorithm terminates when the reporting states of all the users are converged.

5.4 Numerical Results

In this section, we are going to provide a simulation result to illustrate the performance of the mechanism learning algorithm proposed in Section 5.3.3. We simulation a 30 users cognitive radio system with each user has 5 buffer states and 10 throughput states. The transmission power constraint on the system is $P = 3$ which is equivalent to that the maximum number of users transmit simultaneously is 3. The constant α is set to be 2 during simulation. $\hat{\rho}_k^0$ and $\hat{b}_k^0$ are generated randomly during initialization.

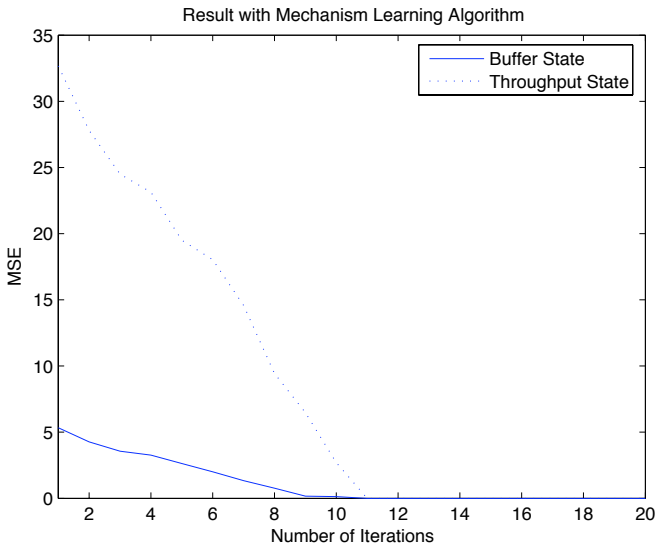

Figure 5.2: The MSE of the reported buffer states and throughput states using the mechanism learning algorithm. The result is of a 30 users system with $L = 5$, $Q_\rho = 10$ and $P = 3$.

In Fig. 5.2, the x-axis represents the number of iterations and y-axis represents the Mean Squared Error (MSE) of the reported buffer states and throughput states. With a slight abuse of notations, we let $\hat{\rho}^n =$

$\{\hat{\rho}_1^n, \ldots, \hat{\rho}_K^n\}$ and $\hat{b}^n = \{\hat{b}_1^n, \ldots, \hat{b}_K^n\}$, the MSE of the reported buffer states and throughput states can be written mathematically as:

$$MSE(\hat{\rho}^n) = \frac{1}{K} \times \sum_{k=1}^{K} (\hat{\rho}_k^n - \rho_k^n)^2; \tag{5.10}$$

$$MSE(\hat{b}^n) = \frac{1}{K} \times \sum_{k=1}^{K} (\hat{b}_k^n - b_k^n)^2. \tag{5.11}$$

The result is obtained by applying the mechanism learning algorithm proposed in Section 5.3.3. In the figure, the solid curve and the dash curve show the mean squared error (MSE) of the reported buffer states and throughput states, respectively. We can see from the figure the MSE converge to 0 after 11 iterations, at which state, all the users in the system report truthfully and $\hat{\Theta} = \Theta$. This simulation result shows numerically that the mechanism learning algorithm is able to converge the reported states to the true states.

5.5 Summary

In this chapter, we combine the mechanism design into the opportunistic scheduling used in cognitive radio networks so that each secondary user optimizes his own utility while the optimality of the overall system performance is ensured. The pricing mechanism we impose on each secondary user is based on the VCG mechanism and we interpret the price with system parameters in a practical cognitive radio system. We also prove the several desirable economic properties that the mechanism has. A mechanism learning algorithm is used to implement the pricing mechanism and the convergence of such algorithm is shown numerically.

Chapter 6

Discussion and Conclusions

6.1 Overview of Thesis

The thesis considered the topics of transmission adaptation, resource allocation, and mechanism design in the cross layers of wireless networks. In many ways, the thesis has highlighted how the game theoretic approach can be applied to optimize the performances in wireless communication systems. Specifically, the transmission adaptation problems in cognitive radio systems and WLANs are formulated as switching control games, the resource allocation problems in LTE systems are formulated as a static game, and a pricing mechanism design is proposed to design a truth revealing opportunistic scheduling algorithm. In some cases, the Nash equilibrium transmission policy is proved to have a specific structure, which makes the estimation and implementation of the optimal solutions highly efficient.

Due to code division multiple access schemes, and multi-antenna communications, as well as other advances in signal processing and communica-

tions, users can share or compete for a wireless medium. Then the two very important issues arising in wireless network optimization are how to allocate the resource among users so an equilibrium system state is obtained and how to balance between multiuser interference and multiple access efficiency in medium access control. Chapter 2 and Chapter 3 considered these issues by computing the Nash equilibrium solutions which optimizes each user's utility with its latency constraint. Chapter 4 considered these issues by studying the correlated equilibrium resource allocation solutions in static game formulation. Chapter 5 considered these issues from the mechanism design perspective.

The thesis applies several mathematical tools to solve various issues in wireless communication systems. For example, switching control game is used in Chapter 2 and Chapter 3 to compute the Nash equilibrium transmission policies in cognitive radio systems and WLANs. The concept of supermodularity has been used together with dynamic programming to prove the monotone structural results on the Nash equilibrium transmission policy. Several learning algorithms, including gradient-based stochastic approximation algorithm, regret matching algorithm, and Q-learning algorithm have been deployed extensively in the thesis for estimating/learning transmission and resource allocation policies.

One of the main focus of this thesis is optimization methodologies, specifically, we focus on its application in wireless communication networks. However, some of the structural results and learning algorithms can find application in resource allocation problems in other areas, such as operation research. In the remaining of the chapter, various aspects of results and

algorithms will be discussed and an outline of possible directions for future research will be presented.

6.2 Discussion on Results and Algorithms

The thesis contains several structural results and learning algorithms, which are discussed below. The first part of the thesis, which consists of Chapter 2 and Chapter 3, was on distributed transmission adaptation is both cognitive radio systems and WLANs. The analysis of such a problem is done using dynamic game theoretic approach. In particular, we formulate the transmission adaptation problem as a switching control game and the Nash equilibrium solutions can be obtained by solving a sequence of Markov decision processes. Such approach offers robust, distributed solutions that can be estimated in real time via very efficient adaptive learning algorithms.

The major results in the first part of the thesis include the structure of the Nash equilibrium transmission policies. Especially, it is proved in Chapter 2 and Chapter 3 that under certain conditions, the Nash equilibrium policy of each user is a randomization of two pure policies and each of the policies is a threshold function with respect to its buffer state. Optimality of the monotone transmission policy converts the problem of estimating the Nash equilibrium policy into estimating few parameters, which can be solved via gradient-based stochastic approximation algorithm with tracking capability. It is shown via simulation that as users adapt their policies using the proposed algorithm, the system converges to the Nash equilibrium.

The second part of the thesis, which consists of Chapter 4, concerns the application of correlated equilibrium concept in wireless communication

systems. Specifically, it formulates the resource allocation problem among cognitive radio base stations in an LTE system under a static environments as a static game. In the chapter, an RB access algorithm is proposed to compute the correlated equilibrium solutions under the static game formulation. It is proved theoretically that the RB access algorithm is ensured to converge to the correlated equilibrium set of the game.

Chapter 5 is the third part of the thesis. It combines the mechanism design into the conventional opportunistic scheduling in cognitive radio networks so that each secondary user optimizes its own utility while optimizing the overall system performance. The pricing mechanism proposed is based on VCG mechanism. It is improved that the proposed mechanism has several desirable economic properties. Furthermore, a mechanism learning algorithm is used to implement such pricing mechanism and the convergence of the algorithm is shown numerically.

6.3 Summary

The thesis has provided structural results for cross-layer transmission adaptation in wireless networks in several contexts, using many powerful mathematical tools that have many application in electrical computer engineering. The problems considered include studying switching control game theoretic approach toward transmission adaptation in cognitive radio system and a WLAN; investigating the correlated equilibrium resource allocation solutions in an LTE system; and designing a pricing mechanism which eliminates the malicious behaviours in the conventional opportunistic access algorithm. The analytical tools utilized in the thesis include static game, stochastic

game, switching control game, the Lagrange multiplier method, and mechanism design. Furthermore, most of the learning algorithms in the thesis are gradient-based stochastic approximation algorithms with tracking capability. Wireless networks application of the work in the thesis range from cognitive radio network, to WLANs, to LTE systems.

6.4 Future Research

This thesis has developed some systematic methods for deriving cross-layer transmission adaptation and scheduling algorithms. In most cases, the results are also applicable to more generale resource allocation problems in wireless networks. The theme of this thesis has been to derive structural results, efficient learning algorithms, and mechanism algorithms for relatively general models. In other words, the analysis in this thesis has assumed slightly abstract frameworks. For future research, the work in this thesis can be extended in several directions. Either the theoretical results can be strengthened/generalized or further analysis can be carried out for more specific network models. In what follows, we propose a few research problems related to the work in this thesis.

6.4.1 How to Avoid the Curse of Dimensionality in Stochastic Games?

Potential applications notwithstanding, there remain substantial hurdles in the application of dynamic stochastic games as a modeling tool in practice. Discrete-time stochastic games with a finite number of states are central to the analysis of strategic interactions among selfish users in dynamic en-

vironments. The usefulness of discrete-time games, however, is limited by their computational burden; in particular, there is a "curse of dimensionality". In a discrete-time dynamic stochastic game, each game player is distinguished by and individual state at each time slot. The system state is a vector encoding the number of players with each possible value of the individual state variable. E.g., the system state in Chapter 2 comprises the channel and buffer state information from all the players. At each time slot, a given player selects an action to maximize its expected discounted payoffs; its subsequent state is determined by its current individual state, its chosen action, and a random shock. The selection will depend on player's individual state and system state. The computation entailed in selecting such an action quickly becomes infeasible as the number of players and individual states grows. How to efficiently reduce the state space is yet an issue to solve before the implementation of stochastic games in practical wireless communication systems. [26] aims to reduce the dimensionality by exploring the alternative of continuous-time stochastic games with a finite number of states and show that the continuous time has substantial advantages.

6.4.2 Transmission Scheduling with Partially Observable States

This thesis has made the assumption that perfect quantized channel state information is available to each user. In many situations, channel states can only be observed in noise and perfect quantized channel information is difficult to obtain. Under a Markov decision process formulation, in the case when the channel state can not be obtained perfectly, the transmis-

sion scheduling problem can be formulated as a partially observable MDP (POMDP). In Chapter 2 and Chapter 3, the transmission adaptation problem is formulated as a switching control game. Switching control game is a special type of dynamic game where the transition probability in any given state depends on only one player. It is known that the Nash equilibrium for such a game can be computed by solving a sequence of Markov decision processes. Whether or not we can extend the POMDP formulation to solve a switching control game with imperfect channel state information is one direction of the future research.

6.4.3 More Analytical Results on Correlated Equilibrium in Stochastic Games

A correlated equilibrium is an outstanding solution concept in game theory and it is a generalization of the Nash equilibrium concept. Correlated equilibrium arises from situations where users select their strategy according to some distribution with inherent coordination. If all users have no incentive to unilaterally deviate, the distribution comprises a correlated equilibrium. In Chapter 4 of this thesis, we have looked into the correlated equilibrium resource allocation solutions in an LTE system under a static environment where static game theoretic framework has been applied. A decentralized RB access algorithm is proposed to compute the correlated equilibrium solutions in a static game formulation. We proved theoretically that the RB access algorithm always converges to the correlated equilibrium set. The study of more analytical result on the correlated equilibrium solutions in stochastic games, e.g., structural results on the correlated equilibrium poli-

cies, prove the convergence of the algorithms deployed therein theoretically, is yet an area to explore.

6.4.4 Other Types of Mechanism Design

The pricing mechanism used in Chapter 5 is based on the well-known VCG mechanism. The development of other types pricing mechanisms or reputation based mechanism can be one direction of the future work. In a reputation based mechanism, system use reputation as a tool to motivate cooperation between users and indicate a good behaviour within the network. To be more precise, a user can be assigned a reputation value determined by its neighbors. Based on how "good" this value, a user can be used or not in a given service provision. In addition, if a user does not pay attention to its reputation and keep acting maliciously, it will be isolated and discarded. Such reputation based mechanism has been applied in ad hoc networks and sensor networks [16, 78]. VCG pricing mechanism is the most common method for motivating users report truthfully, however, the study of other alternative pricing mechanism is a potential research area.

Bibliography

[1] 3rd Generation Partnership Project. Technical Specification and Technical Reports for a UTRAN-based 3GPP system. *TR21.101 v0.0.8*, 2009.

[2] 3rd Generation Partnership Project. Architecture Aspects for Home NodeB and Home eNodeB. *TR 23.830 V0.3.1*, March 2009.

[3] E. Altman. *Constrained Markov Decision Processes*. Chapman and Hall, London, 1999.

[4] E. Altman, T. Basar, and R. Srikant. Nash equilibria for combined flow control and routing in networks: Asymptotic behavior for a large number of users. *IEEE Transactions on Automatic Control*, 47(6):917–930, June 2002.

[5] E. Altman, B. Gaujal, and A. Hordijk. *Discrete-Event Control of Stochastic Networks: Multimodularity and Regularity*. Springer, Berlin, 2003.

[6] A. Attar, V. Krishnamurthy, and O. Namvar. Interference management

using cognitive base-stations for UMTS LTE. *To appear in: IEEE Communications Magazine*, 2011.

[7] R. J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1:67–96, March 1974.

[8] R. J. Aumann. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55(1):1–18, 1987.

[9] M. Benam, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions, Part II: Applications. *Mathematics of Operation Research*, 31(4):673–695, November 2006.

[10] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, MA, 1989.

[11] F. J. Beutler and K. W. Ross. Optimal policies for controlled Markov chains with a constraint. *Journal of Mathematical Analysis and Applications*, 112:236–252, November 1985.

[12] V. K. Bhargava and E. Hossain. *Cognitive Wireless Communication Networks*. Springer-Verlag, New York, 2007.

[13] K. G. Binmore. *Fun and Games: A Text on Game Theory*. D. C. Heath, Lexington, Mass, 1992.

[14] D. Blackwell. An analog of the minmax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[15] C. Boutilier, Y. Shoham, and M. P. Wellman. Economic principles of multiagent systems (editorial). *Artificial Intelligence*, 94(1):1–6, 1997.

[16] H. Chen. Task-based trust management for wireless sensor networks. *International Journal of Security and Its Applications*, 3(2):21–26, April 2009.

[17] Y. Chen, Q. Zhao, and A. Swami. Joint design and seperation principle for opportunistic spectrum access. In *Proceedings of IEEE ACSSC*, pages 696–700, October/November 2006.

[18] D. Choi, P. Monajemi, S. Kang, and J. Villasenor. Dealing with loud neighbors: The benefits and tradeoffs of adaptive femtocell access. In *Proceedings of IEEE GLOBECOM*, pages 1–5, 2008.

[19] S. T. Chung and A. J. Goldsmith. Degrees of freedom in adaptive modulation: A unified view. *IEEE Transactions on Communications*, 49(9):1561–1571, September 2001.

[20] E. H. Clarke. Multipart pricing of public goods. *Public Choice*, 2:19–33, 1971.

[21] P. Dasgupta and E. Maskin. Efficient auctions. *Quarterly Journal of Economics*, 115:341–388, 2000.

[22] R. K. Dash, D. C. Parkes, and N. R. Jennings. Computational mechanism design: A call to arms. *IEEE Intelligent Systems*, 18(6):40–47, November/December 2003.

[23] R. K. Dash, A. Rogers, N. R. Jennings, S. Reece, and S. Roberts. Constrained bandwidth allocation in multi-sensor information fusion:

A mechanism design approach. In *Proceedings of IEEE Information Fusion Conference*, pages 1185–1192, July 2005.

[24] D. Djonin and V. Krishnamurthy. MIMO transmission control in fading channels - A constrained Markov decision process formulation with monotone randomized policies. *IEEE Transactions on Signal Processing*, 55(10):5069–5083, October 2007.

[25] D. V. Djonin and V. Krishnamurthy. Q-Learning algorithms for constrained Markov decision processes with randomized monotone policies: Application to MIMO transmission control. *IEEE Transactions on Signal Processing*, 55(5):2170–2181, May 2007.

[26] U. Doraszelski and K. L. Judd. Avoiding the curse of dimensionality in dynamic stochastic games. *Rand Journal of Economics*, 2008.

[27] U. Doraszelski and K. L. Judd. Avoiding the curse of dimensionality in dynamic stochastic games. *Quantitative Economics*, 2011.

[28] A. Farrokh and V. Krishnamurthy. Opportunistic scheduling for streaming users in HSDPA multimedia systems. *IEEE Transactions on Multimedia Systems*, 8(4):844–855, August 2006.

[29] A. Farrokh, V. Krishnamurthy, and R. Schober. Optimal adaptive modulation and coding with switching costs. *IEEE Transactions on Communications*, 57(3):697–706, March 2009.

[30] M. Felegyhazi, M. Cagalj, S. S. Bidokhti, and J. Hubaux. Non-

cooperative multi-radio channel allocation in wireless networks. In *Proceedings of IEEE INFOCOM*, pages 1442–1450, May 2007.

[31] J. A. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer-Verlag, New York, 1997.

[32] D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*. The MIT Press, 1998.

[33] D. Fudenberg and J. Tirole. *Game Theory*. The MIT Press, Cambridge, MA, 1996.

[34] D. Grosu and A. T. Chronopoulos. Algorithmic mechanism design for load balancing in distributed systems. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 34(1):77–84, 2004.

[35] T. Groves. Incentives in Team. *Econometrica*, 41(4):617–631, 1973.

[36] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150, September 2000.

[37] S. Hart and A. Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, 93(5):1830–1836, December 2003.

[38] S. Haykin. Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2):201–220, February 2005.

[39] T. Heikkinen. Distributed scheduling via pricing in a communication network. *Wireless Networks*, 10(3):233–244, May 2004.

[40] T. Heikkinen. A potential game approach to distributed power control and scheduling. *Computer Networks*, 50(13):2295–2311, 2006.

[41] J. W. Huang and V. Krishnamurthy. Game theoretic issues in cognitive radio systems. *Journal of Communications*, 4(10):790–802, November 2009. Invited paper.

[42] J. W. Huang and V. Krishnamurthy. Truth revealing opportunistic scheduling in cognitive radio systems. In *Proceedings of the 10th IEEE International Workshop in Signal Processing Advances in Wireless Communications*, June 2009.

[43] J. W. Huang and V. Krishnamurthy. Transmission control in cognitive radio as a Markovian dynamic game - Structural result on randomized threshold policies. *IEEE Transactions on Communications*, 58(2):301–310, February 2010.

[44] J. W. Huang, H. Mansour, and V. Krishnamurthy. A dynamical games approach to transmission rate adaptation in multimedia WLAN. *IEEE Transactions on Signal Processing*, 58(7):3635 – 3646, July 2010.

[45] *IEEE Standard for Local and Metropolitan Area Networks, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. IEEE std 802.11-2007, 2007.

[46] *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems*. IEEE std 802.16-2004, 2004.

[47] *Advanced Video Coding for Generic Audiovisual Services*. ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, June 2005.

[48] K. Kar and L. Tassiulas. Layered multicast rate control based on Lagrangian relaxation and dynamic programming. *IEEE Journal on Selected Areas in Communications*, 24(8):1464–1474, August 2006.

[49] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annuals of Mathematical Statistics*, 23(3): 462–466, September 1952.

[50] V. Krishnamurthy and G. G. Yin. Recursive algorithms for estimation of hidden Markov models and autoregressive models with Markov regime. *IEEE Transactions on Information Theory*, 48(2):458–476, February 2002.

[51] V. Krishnamurthy, M. Maskery, and G. Yin. Decentralized adaptive filtering algorithms for sensor activation in an unattended ground sensor network: A correlated equilibrium Game Theoretic Analysis. *IEEE Transactions on Signal Processing*, 56(12):6086–6101, December 2008.

[52] V. Krishnamurthy, M. Maskery, and G. Yin. Decentralized adaptive filtering algorithm for sensor activation in an unattended ground sensor network. *IEEE Transactions on Signal Processing*, 56(12):6086–6101, December 2008.

[53] H. J. Kushner and D. S. Clark. *Stochastic Approximation Methods for*

Constrained and Unconstrained Systems. Springer-Verlag, New York, 1978.

[54] H. J. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer-Verlag, New York, 2003.

[55] D.-K. Kwon, M.-Y. Shen, and C. C. J. Kuo. Rate control for H.264 video with enhanced rate and distortion models. *IEEE Trans. Circuits Syst. Video Techn.*, 17(5):517–529, 2007.

[56] G. Liebl, H. Jenkac, T. Stockhammer, and C. Buchner. Radio link buffer management and scheduling for wireless video streaming. *Springer Telecommunication Systems*, 30(1-3):255–277, 2005.

[57] X. Liu, E. K. P. Chong, and N. B. Shroff. Optimal opportunistic scheduling in wireless networks. In *Proceedings of IEEE Vehicular Technology Conference*, pages 1417–1421, October 2003.

[58] D. Lopez-Perez, A. Juttner, and J. Zhang. Optimisation methods for dynamic frequency planning in OFDMA networks. In *Proceedings of Telecommunications Network Strategy and Planning Symposium*, pages 1579–1584, September 2008.

[59] D. Lopez-Perez, G. d. l. Roche, A. Valcarce, A. Juttner, and J. Zhang. Interference avoidance and dynamic frequency planning for wimax femtocells network. In *Proceedings of IEEE ICCS*, pages 1579–1584, November 2008.

[60] A. B. MacKenzie and S. B. Wicker. Game theory and the design of self-

configuring, adaptive wireless networks. *IEEE Communication Magazine*, 39(11):126–131, November 2001.

[61] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy. Real-time joint rate and protection allocation for multi-user scalable video streaming. In *Proceedings of IEEE Personal, Indoor, and Mobile Radio Communications (PIMRC)*, September 2008.

[62] A. G. Marques, W. Xin, and G. B. Giannakis. Optimal stochastic dual resource allocation for cognitive radios based on quantized CSI. In *Proceedings of IEEE ICASSP*, pages 2801–2804, April 2008.

[63] A. MasColell, M. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, Oxford, 1995.

[64] M. Maskery, V. Krishnamurthy, and Q. Zhao. Decentralized dynamic spectrum access for cognitive radios: Cooperative design of a non-cooperative game. *IEEE Transactions on Communications*, 57(2):459–469, February 2009.

[65] F. Meshkati, M. Chiang, H. V. Poor, and S. C. Schwartz. A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems. *IEEE Journal on Selected Areas in Communications*, 24(6):1115–1129, 2006.

[66] R. Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38:175–208, 1971.

[67] J. Mitola III. Cognitive radio for flexible mobile multimedia commu-

nications. In *Proceedings of IEEE International Workshop on Mobile Multimedia Communications*, pages 3–10, November 1999.

[68] S. R. Mohan and T. E. S. Raghavan. An algorithm for discounted switching control stochastic games. *OR Spektrum*, 55(10):5069–5083, October 2007.

[69] O. Morgenstern and J. v. Neumann. *The Theory of Games and Economic Behavior*. Princeton University Press, 1947.

[70] J. Nash. Equilibrium points in n-person games. In *Proceedings of the National Academy of Sciences*, pages 48–49, 1950.

[71] J. V. Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[72] M. Ngo and V. Krishnamurthy. Monotonicity of constrained optimal transmission policies in correlated fading channels with ARQ. *IEEE Transactions on Signal Processing*, 58(1):438–451, January 2010.

[73] N. Nie and C. Comaniciu. Adaptive channel allocation spectrum etiquette for cognitive radio networks. In *Proceedings of IEEE International Symposium on DySPAN*, pages 269–278, November 2005.

[74] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press, 1994.

[75] C. Peng, H. Zheng, and B. Y. Zhao. Utilization and fairness in spectrum assignment for opportunistic spectrum access. *Mobile Networks and Applications*, 11(4):555–576, August 2006.

[76] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, New York, 1994.

[77] J. Reichel, H. Schwarz, and M. Wien. Joint Scalable Video Model JSVM-9. Technical Report N 8751, ISO/IEC JTC 1/SC 29/WG 11, Marrakech, Morocco, January 2007.

[78] K. Ren, T. Li, Z. Wan, R. H. Deng, and K. Kim. Highly reliable trust establishment scheme in ad hoc networks. *Journal of Computer and Telecommunications Netwroking*, 45(6):687–699, August 2004.

[79] H. Robbins and S. Monro. A stochastic approximation method. *The Annuals of Mathematical Statistics*, 22(3):400–407, September 1951.

[80] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Trans. Circuits Syst. Video Techn.*, 17(9):1103–1120, 2007.

[81] G. Scutari, D. Palomar, and S. Barbarossa. Asynchronous iterative water-filling for Gaussian frequency-selective interference channels. *IEEE Transactions on Information Theory*, 54(7):2868–2878, July 2008.

[82] J. E. Smith and K. F. McCardle. Structural properties of stochastic dynamic programs. *Operations Research*, 50(5):796–809, September-October 2002.

[83] J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control*. Wiley-Interscience, 2003.

[84] V. Srivastava and M. Motani. Cross-layer design: A survey and the road ahead. *IEEE Communications Magzine*, 43(12):112–119, 2005.

[85] 3GPP. Overview of 3GPP. Release 9 V0.1.1, September 2010. URL http://www.3gpp.org/ftp/Information.

[86] 3GPP. 3GPP work items on Self-Organizing Networks, October 2010. URL http://www.3gpp.org/ftp/Information.

[87] ISO/IEC JTC 1/SC 29/WG 11 N8964. JSVM-10 software, 2007. URL http://wg11.sc29.org/mpeg/docs/_listwg11_80.htm.

[88] R. Urgaonkar and M. J. Neely. Opportunistic scheduling with reliability guarantees in cognitive radio networks. In *Proceedings of IEEE INFOCOM*, pages 1301–1309, April 2008.

[89] W. Vickrey. Counterspeculation auctions and competitive sealed tenders. *Journal of Finance*, 16(1):8–37, 1961.

[90] O. J. Vrieze, S. H. Tijs, T. E. S. Raghavan, and J. A. Filar. A finite algorithm for the switching control stochastic game. *OR Spektrum*, 5 (1):15–24, March 1983.

[91] B. Wang, Y. Wu, and K. J. R. Liu. Game theory for cognitive radio networks: An overview. *The International Journal of Computer and Telecommunications Networking*, 54(14):2537–2574, October 2010.

[92] W. Wang, X. Liu, and H. Xiao. Exploring opportunistic spectrum availability in wireless communication networks. In *Proceedings of IEEE VTC*, September 2005.

[93] I. C. Wong and B. L. Evans. Optimal downlink of OFDMA resource allocation with linear complexity to maximize ergodic rates. *IEEE Transactions on Wireless Communications*, 7(3):962–971, March 2008.

[94] W. Yu, G. Ginis, and J. M. Cioffi. Distributed multiuser power control for digital subscriber lines. *IEEE Journal on Selected Areas in Communications*, 20(5):1105–1115, June 2002.

[95] Q. Zhao, L. Tong, A. Swami, and Y. Chen. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework. *IEEE Journal on Selected Areas in Communications*, 25(3): 589–600, April 2007.

[96] X. Zhu and B. Girod. Analysis of multi-user congestion control for video streaming over wireless networks. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, July 2006.

Appendix A

Proof of Theorem 2.3.2

A.1 Proof of Theorem 2.3.2

Consider the case where the delay constraint parameter $\widetilde{D}_k$ in (2.8) is chosen so that a Nash equilibrium policy of the optimization problem exists. When the Nash equilibrium is obtained, the delay constraint (2.8) will hold with equality. This feature implies that by introducing the Lagrange multiplier λ_k, we can get the Nash equilibrium policy when the objective function (2.7) is minimized and the equality of the delay constraint (2.8) is obtained.

According to Algorithm 1, the transmission policy and value matrix of

the kth user are updated by the following steps:

$$\begin{aligned}
\pi_k^*(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)^n &= \arg \min_{\pi_k(\mathbf{s})^n} \Big\{ c(\mathbf{s}, a_k) + \lambda_k \cdot d_k(\mathbf{s}, a_k) + \\
&\qquad \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{(n-1)}(\mathbf{s}') \Big\} \qquad \text{(A.1)} \\
v_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k) &= \min_{\pi_k(\mathbf{s})^n} \Big\{ c(\mathbf{s}, a_k) + \lambda_k \cdot d_k(\mathbf{s}, a_k) + \\
&\qquad \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{(n-1)}(\mathbf{s}') \Big\}. \qquad \text{(A.2)}
\end{aligned}$$

In order to prove that the optimal transmit action policy $\pi_k^*(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)^n$ is monotone nondecreasing on the buffer state b_k, we have to show that the right-hand side of (A.1) is a submodular function of (b_k, a_k). According to assumption A 2.3.2, $c(\mathbf{s}, a_k) + \lambda_k \cdot d_k(\mathbf{s}, a_k)$ is a submodular function of (b_k, a_k), thus, we only need to demonstrate that $\sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{(n-1)}(\mathbf{s}')$ is also submodular in the pair (b_k, a_k).

As mentioned before, the overall state space $\mathcal{S}$ is the union of all the K sub-spaces: $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cdots \cup \mathcal{S}_K$. Thus, we can write $\sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{(n-1)}(\mathbf{s}')$ as a summation of K terms according to the state space partition. In order to show the submodularity of $\sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{(n-1)}(\mathbf{s}')$, we only need to prove the submodularity property of each term. The decomposition is given as follows:

$$\sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{(n-1)}(\mathbf{s}') = \sum_{i=1}^{K} \sum_{\mathbf{s}', \mathbf{s}' \in \mathcal{S}_i} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{(n-1)}(\mathbf{s}'). \qquad \text{(A.3)}$$

If we consider the ith term of the right-hand side of (A.3), using the state

transition probability expression from (2.4), we have

$$\begin{aligned} & \sum_{\mathbf{s}',\mathbf{s}'\in\mathcal{S}_i} \mathbb{P}(\mathbf{s}'|\mathbf{s},a) v_k^{(n-1)}(\mathbf{s}') \\ = & \prod_{l=1}^{K} \mathbb{P}(h_l'|h_l) \cdot \prod_{l=1,l\neq k}^{K} \mathbb{P}(b_l'|b_l) \cdot \mathbb{P}(b_k'|b_k,a_k) \cdot v_k^{(n-1)}(\mathbf{s}'). \end{aligned}$$

The proof of Theorem 2.3.2 requires the result from Lemma A.1.1, whose proof will be given afterwards.

Lemma A.1.1 *Under the assumptions of Theorem 2.3.2, $v_k(\mathbf{s})$ is nondecreasing on b_k when the state lives in state space $\mathcal{S}_i$, that is $\mathbf{s} \in \mathcal{S}_i, i = 1, 2, \ldots, K$.*

Continuing with the proof of the theorem, the assumption says that $\sum_{b_k'=l}^{L} \mathbb{P}(b_k'|b_k,a_k)$ is submodular in (b_k,a_k) for any l. According to the definition of submodularity, this assumption can be mathematically written in the following way, for all the l and $\mathbf{s}' \in \mathcal{S}_i$:

$$\begin{aligned} & \sum_{b_k'=l}^{L} \mathbb{P}(b_k'|b_k^-,a_k^-) + \sum_{b_k'=l}^{L} \mathbb{P}(b_k'|b_k^+,a_k^+) \\ \leq & \sum_{b_k'=l}^{L} \mathbb{P}(b_k'|b_k^-,a_k^+) + \sum_{b_k'=l}^{L} \mathbb{P}(b_k'|b_k^+,a_k^-). \end{aligned} \tag{A.4}$$

Based on the result from Lemma A.1.1, we can apply Lemma 4.7.2 from [76] to our model, which yields

$$\begin{aligned} & \sum_{b_k'=0}^{L} \mathbb{P}(b_k'|b_k^-,a_k^-) \cdot v_k(\mathbf{s}') + \sum_{b_k'=0}^{L} \mathbb{P}(b_k'|b_k^+,a_k^+) \cdot v_k(\mathbf{s}') \leq \\ & \sum_{b_k'=0}^{L} \mathbb{P}(b_k'|b_k^-,a_k^+) \cdot v_k(\mathbf{s}') + \sum_{b_k'=0}^{L} \mathbb{P}(b_k'|b_k^+,a_k^-) \cdot v_k(\mathbf{s}') \end{aligned} \tag{A.5}$$

for $\mathbf{s}' \in \mathcal{S}_i$.

The summation of (A.5) over $\mathbf{h}'$ for $[\mathbf{h}', \mathbf{b}'] \in \mathcal{S}_i$ yields

$$\sum_{\mathbf{s}', \mathbf{s}' \in \mathcal{S}_i} \left[\mathbb{P}(\mathbf{s}'|\mathbf{s}, b_k^-, a_k^-) v_k^{n-1}(\mathbf{s}') + \mathbb{P}(\mathbf{s}'|\mathbf{s}, b_k^+, a_k^+) v_k^{n-1}(\mathbf{s}')\right] \leq$$
$$\sum_{\mathbf{s}', \mathbf{s}' \in \mathcal{S}_i} \left[\mathbb{P}(\mathbf{s}'|\mathbf{s}, b_k^-, a_k^+) v_k^{n-1}(\mathbf{s}') + \mathbb{P}(\mathbf{s}'|\mathbf{s}, b_k^+, , a_k^-) v_k^{n-1}(\mathbf{s}')\right].$$

This is the definition of submodularity of $\sum_{\mathbf{s}', \mathbf{s}' \in \mathcal{S}_i} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{(n-1)}(\mathbf{s}')$ in (b_k, a_k). Furthermore, the positive weighted sum of submodular functions is also submodular, which establishes the submodularity of the right-hand side of function (A.1). Thus, the optimal transmit action policy $\pi_k^*(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)^n$ is monotone nondecreasing on the buffer state b_k for a fixed Lagrange multiplier λ_k.

According to [11], the constrained optimal transmission scheduling policy is a randomized mixture of two pure policies, which can be computed with two different Lagrange multipliers. As discussed above, it has been shown that each of these two pure policies is nondecreasing on the buffer occupancy state. Thus, the mixed policy is also nondecreasing on the buffer state, which concludes the proof.

A.2 Proof of Lemma A.1.1

We use backward induction to prove that $v_k(\mathbf{s})$ is nondecreasing on b_k with state $\mathbf{s} \in \mathcal{S}_i$ for $i = 1, 2, \ldots, K$.

First we assume that $v_k^{(n-1)}(\mathbf{s})$ is nondecreasing on b_k for $\mathbf{s} \in \mathcal{S}_i, i = 1, 2, \ldots, K$ at the time instant $(n-1)$. The value matrix at time instant n is updated according to (A.2). Denoting the optimal action policy as π_k^*, the

update of value matrix can be written more explicitly as:

$$v_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k) = c(\mathbf{s}, \pi_k^*(\mathbf{s})) + \lambda_k \cdot d_k(\mathbf{s}, \pi_k^*(\mathbf{s})) + \beta \sum_{i=1}^{K} \sum_{\mathbf{s}', \mathbf{s}' \in \mathcal{S}_i} \mathbb{P}(\mathbf{s}'|\mathbf{s}, \pi_k^*(\mathbf{s})) v_k^{(n-1)}(\mathbf{s}'). \quad \text{(A.6)}$$

Because of assumption A4, we know that $c(\mathbf{s}, \pi_k^*(\mathbf{s})) + \lambda_k \cdot d_k(\mathbf{s}, \pi_k^*(\mathbf{s}))$ is nondecreasing on b_k for any $\mathbf{h}$, $\mathbf{f}$ and a^*. By applying the result from Lemma 4.7.2 in [76], we can prove that $\sum_{\mathbf{s}', \mathbf{s}' \in \mathcal{S}_i} \mathbb{P}(\mathbf{s}'|\mathbf{s}, \pi_k^*(\mathbf{s})) v_k^{(n-1)}(\mathbf{s}')$ is nondecreasing on b_k for any $\mathbf{h}$, $\mathbf{f}$ and $\pi_k^*(\mathbf{s})$. Thus, $v_k^n(\mathbf{s})$ is nondecreasing on b_k for $\mathbf{s} \in \mathcal{S}_k, k = 1, 2, \ldots, K$ at time instant n.

For any $\mathbf{s} \in \mathcal{S}_i, i \neq k$, the value vector is updated according to (A.7). In this case, the instantaneous cost is independent of b_k, thus the nondecreasing property of $v_k^n(\mathbf{s})$ on b_k when $\mathbf{s} \in \mathcal{S}_i, i \neq k$ is preserved from $v_k^{(n-1)}(\mathbf{s})$.

$$v_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_i, i \neq k) = c(\mathbf{s}, \pi_i^*(\mathbf{s})) + \lambda_k \cdot d_k(\mathbf{s}, \pi_i^*(\mathbf{s}) + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, \pi_i^*(\mathbf{s})) v_k^{(n-1)}(\mathbf{s}'). \quad \text{(A.7)}$$

When the time horizon of the Markovian problem is very large, that is $n \to \infty$, the initial value matrix $v_k^{(0)}(\mathbf{s})$ no longer affects $v_k^n(\mathbf{s})$. The chosen value of $v_k^{(0)}(\mathbf{s})$ is arbitrary. In our case, we can initialize $v_k^{(0)}(\mathbf{s})$ to satisfy the condition stated in Lemma A.1.1, which concludes the proof.

Appendix B

Proof of The Convergence of Algorithm 3

1. By following the definition of $\pi_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)$ from Algorithm 3 we can deduce that

$$\mathbf{V}_k^{(n+1)} \leq \mathbf{V}_k^n. \tag{B.1}$$

2. When $\pi_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k) = \pi_k^{(n+\Delta)}(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)$, we have $\mathbf{V}_k^n = \mathbf{V}_k^{(n+\Delta)}$. In view of (B.1), if $\mathbf{V}_k^n \neq \mathbf{V}_k^{(n+\Delta)}$, then $\pi_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k) \neq \pi_k^{(n+\Delta)}(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)$ for any $\Delta = 1, 2, 3, \ldots$.

3. The payoff function of the matrix game with $\mathbf{s} \in \mathcal{S}_1$ is of the form $\big[- c(\mathbf{s}, a_k) + \lambda_k^m \cdot d_k(\mathbf{s}, a_k) \big] + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}')$. Please note the first term of this payoff function $\big[- c(\mathbf{s}, a_k) + \lambda_k^m \cdot d_k(\mathbf{s}, a_k) \big]$ is independent of $\mathbf{V}_k^n$. By using the result from [31] Lemma 6.3.2, $\pi_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)$ equals to some extreme optimal action in a submatrix game with payoff function

$\big[- c(\mathbf{s}, a_k) + \lambda_k^m \cdot d_k(\mathbf{s}, a_k)\big]$. As there are only finite number of extreme optimal action candidates for $\pi_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)$, there exist n and $\Delta \geq 1$ such that $\pi_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k) = \pi_k^{(n+\Delta)}(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)$, which in turn implies $\mathbf{V}_k^n = \mathbf{V}_k^{(n+\Delta)}$.

4. If $\mathbf{V}_k^n = \mathbf{V}_k^{(n-1)}$, $\pi_k^n(\mathbf{s}, \mathbf{s} \in \mathcal{S}_k)$ is the optimal action policy for user k ($k \in \{1, \ldots, K\}$) of the game with a fixed Lagrange multiplier λ_k^m.

Based on the $1 - 4$ observations above, we conclude that Algorithm 3 will converge in a finite number of iterations with fixed Lagrange multipliers $\lambda_{k=1,2,\ldots,K}^m$.

Appendix C

Proof of Theorem 3.4.1

C.1 Proof of Theorem 3.4.1

We choose delay constraint parameters $\widetilde{D}_{k,k=1,2,...,K}$ in the optimization problem (3.13) carefully to ensure the existence of an optimal policy. The optimal policy can be obtained when the objective function (3.11) is maximized and the equality of the delay constraint (3.12) is held.

According to Algorithm 3, for $\forall s \in \mathcal{S}$, the optimal policy and value matrix of kth user are updated by the following steps:

$$\begin{aligned} \pi_k^n(\mathbf{s}) &= \arg\min_{\pi_k^n(\mathbf{s})} \Bigg\{ -c(\mathbf{s}, a_k) + \lambda_k^m \cdot d_k(\mathbf{s}, a_k) \\ &\quad +\beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}') \Bigg\}; \qquad \text{(C.1)} \\ v_k^{(n+1)}(\mathbf{s}) &= -c(\mathbf{s}, \pi_k^n(\mathbf{s})) + \lambda_k^m \cdot d_k(\mathbf{s}, \pi_k^n(\mathbf{s})) \\ &\quad +\beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, \pi_k^n(\mathbf{s})) v_k^n(\mathbf{s}'). \qquad \text{(C.2)} \end{aligned}$$

A sufficient condition for the optimal transmission action policy $\pi_k^n(\mathbf{s})$ to be monotone nondecreasing in the buffer state b_k is that the right hand side of (C.1) is a submodular function of (b_k, a_k). According to assumptions A 3.4.2 and A 3.4.3, $-c(\mathbf{s}, a_k) + \lambda_k^m \cdot d_k(\mathbf{s}, a_k)$ is a submodular function of (b_k, a_k), thus, we only need to demonstrate $\sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}')$ is also submodular in the pair (b_k, a_k).

Recall the overall state space $\mathcal{S}$ is the union of all the K sub-spaces, $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cdots \cup \mathcal{S}_K$. Thus, we can write $\sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}')$ as a summation of K terms according to the partition of the state space, where the ith term $(i = 1, 2, \ldots, K)$ is denoted by $Q_i(\mathbf{s}, a_k)$. By using the property of the state transition probability (3.15) and Assumption A 3.4.4, we have the following result:

$$Q_i(\mathbf{s}, a_k) = \sum_{\mathbf{s}', \mathbf{s}' \in \mathcal{S}_i} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}') =$$
$$\prod_{l=1}^{K} \mathbb{P}(h_l'|h_l) \prod_{l=1}^{K} \mathbb{P}(m_l'|m_l) \prod_{l=1, l\neq k}^{K} \mathbb{P}(b_l'|b_l) \mathbb{P}(b_k'|b_k + a_k) v_k^n(s'). \qquad \text{(C.3)}$$

In order to show the submodularity of $\sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}')$, we only need to prove the submodularity property of $Q_i(\mathbf{s}, a_k)$ for $i = 1, 2, \ldots, K$.

The proof of Theorem 3.4.1 needs the result from Lemma C.1.1 , whose proof will then be given after the proof of the Theorem.

Lemma C.1.1 *Under the conditions of Theorem 3.4.1, $Q_i(\mathbf{s}, a_k)$ is of the form $Q_i(\mathbf{s}, a_k) = \bar{Q}_i(b_k - a_k, \{\mathbf{h}, \mathbf{m}, \mathbf{b}_{-k}\})$. Function $\bar{Q}_i(x, \mathbf{y})$ is integer convex function of x for any given $\mathbf{y} = \{\mathbf{h}, \mathbf{m}, \mathbf{b}_{-k}\}$.*

According to Lemma C.1.1, $\bar{Q}_i(x, \mathbf{y})$ is integer convex in x, which can be written as:

$$\begin{aligned} & \bar{Q}_i(x'+\Delta,\mathbf{y}) - \bar{Q}_i(x',\mathbf{y}) \\ \leq\ & \bar{Q}_i(x+\Delta,\mathbf{y}) - \bar{Q}_i(x,\mathbf{y}) \end{aligned} \tag{C.4}$$

for $x' \geq x$ and $\Delta \geq 0$. Substitute in the above equation with $x' = b_k - a_k$, $x = b_k - a'_k$ and $\Delta = b'_k - b_k$ for $b'_k \geq b_k$ and $a'_k \geq a_k$, we obtain

$$\begin{aligned} & Q_i(\{b'_k, a_k\},\mathbf{y}) - Q_i(\{b_k, a_k\},\mathbf{y}) \\ \leq\ & Q_i(\{b'_k, a'_k\},\mathbf{y}) - Q_i(\{b_k, a'_k\},\mathbf{y}), \end{aligned} \tag{C.5}$$

which is the definition of submodularity of $Q_i(\mathbf{s}, a_k)$ in b_k and a_k. Furthermore, the linear combination of submodular functions still remains the submodular property. Thus, $\sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^n(\mathbf{s}')$ is submodular in the pair (b_k, a_k).

Based on the above results, we can prove the submodularity of the right hand side of equation (C.1). This proves the optimal transition action policy $\pi_k^*(\mathbf{s})^n$ is monotone nondecreasing on the buffer state b_k under fixed positive Lagrange constant.

According to [11], the optimal randomized policy with a general constraint is a mixed policy comprising of two pure policies that can be computed under two different Lagrange multipliers. As discussed above, we have already shown that each of these two pure policies is nondecreasing on the buffer state occupancy. Thus, the mixed policy also owns the nondecreasing property on the buffer state. This concludes the proof.

C.2 Proof of Lemma C.1.1

Let us first assume that $v_k^n(\mathbf{s}')$ is an integer convex function of b_k'. Then, according to Proposition 2 of [82] and A 3.4.4 in Section 3.4.2 along with (C.3), it can be concluded that $Q_i(\mathbf{s}, a_k)$ is an integer convex function of $b_k - a_k$.

Thus, we only need to prove that $v_k^n(\mathbf{s}')$ is an integer convex function of b_k', which can be proved via backward induction. The value vector can be updated by the value iteration algorithm as follows:

$$v_k^{(n)}(\mathbf{s}) = -c(\mathbf{s}, a_k) + \lambda_k^m \cdot d_k(\mathbf{s}, a_k) + \beta \sum_{\mathbf{s}'=1}^{|\mathcal{S}|} \mathbb{P}(\mathbf{s}'|\mathbf{s}, a_k) v_k^{n-1}(\mathbf{s}'). \tag{C.6}$$

If $v_k^{n-1}(\mathbf{s}')$ is integer convex in b_k', as well as the A 3.4.2 and A 3.4.3 from Section 3.4.2, it can be proved that $v_k^n(\mathbf{s})$ is integer convex in b_k by using the key Lemma 61 from [5]. The convergence of the value iteration algorithm is not affected by the initial values. Choose $v_k^0(\mathbf{s}')$ to be an integer convex function of b_k', $v_k^n(\mathbf{s})$ is integer convex in b_k follows by induction. This concludes the proof.