MODELING, FORECASTING AND RESOURCE ALLOCATION IN COGNITIVE RADIO NETWORKS

by

LUTFA AKTER

B.S., Bangladesh University of Engineering and Technology, 2002
M.S., Bangladesh University of Engineering and Technology, 2004

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2010

Abstract

With the explosive growth of wireless systems and services, bandwidth has become a treasured commodity. Traditionally, licensed frequency bands were exclusively reserved for use by the primary license holders (primary users), whereas, unlicensed frequency bands allow spectrum sharing. Recent spectrum measurements indicate that many licensed bands remain relatively unused for most of the time. Therefore, allowing secondary users (users without a license to operate in the band) to operate with minimal or no interference to primary users is one way of sharing spectrum to increase efficiency. Recently, Federal Communications Commission (FCC) has opened up licensed bands for opportunistic use by secondary users. A cognitive radio (CR) is one enabling technology for systems supporting opportunistic use. A cognitive radio adapts to the environment it operates in by sensing the spectrum and quickly decides on appropriate frequency bands and transmission parameters to use in order to achieve certain performance goals. A cognitive radio network (CRN) refers to a network of cognitive radios/secondary users.

In this dissertation, we consider a competitive CRN with multiple channels available for opportunistic use by multiple secondary users. We also assume that multiple secondary users may coexist in a channel and each secondary user (SU) can use multiple channels to satisfy their rate requirements. In this context, firstly, we introduce an integrated modeling and forecasting tool that provides an upper bound estimate of the number of secondary users that may be demanding access to each of the channels at the next instant. Assuming a continuous time Markov chain model for both primary and secondary users activities, we propose a Kalman filter based approach for estimating the number of primary and secondary users. These estimates are in turn used to predict the number of primary and secondary users in a future time instant. We extend the modeling and forecasting framework to the

case when SU traffic is governed by Erlangian process. Secondly, assuming that scheduling is complete and SUs have identified the channels to use, we propose two quality of service (QoS) constrained resource allocation frameworks. Our measures for QoS include signal to interference plus noise ratio (SINR) /bit error rate (BER) and total rate requirement. In the first framework, we determine the minimum transmit power that SUs should employ in order to maintain a certain SINR and use that result to calculate the optimal rate allocation strategy across channels. The rate allocation problem is formulated as a maximum flow problem in graph theory. We also propose a simple heuristic to determine the rate allocation. In the second framework, both transmit power and rate per channel are simultaneously optimized with the help of a bi-objective optimization problem formulation. Unlike prior efforts, we transform the BER requirement constraint into a convex constraint in order to guarantee optimality of resulting solutions. Thirdly, we borrow ideas from social behavioral models such as Homo Egualis (HE), Homo Parochius (HP) and Homo Reciprocan (HR) models and apply it to the resource management solutions to maintain fairness among SUs in a competitive CRN setting. Finally, we develop distributed user-based approaches based on "Dual Decomposition Theory" and "Game Theory" to solve the proposed resource allocation frameworks. In summary, our body of work represents significant ground breaking advances in the analysis of competitive CRNs.

MODELING, FORECASTING AND RESOURCE ALLOCATION IN COGNITIVE RADIO NETWORKS

by

LUTFA AKTER

B.S., Bangladesh University of Engineering and Technology, 2002
M.S., Bangladesh University of Engineering and Technology, 2004

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2010

Approved by:

Major Professor
Balasubramaniam Natarajan

Copyright

Abstract

With the explosive growth of wireless systems and services, bandwidth has become a treasured commodity. Traditionally, licensed frequency bands were exclusively reserved for use by the primary license holders (primary users), whereas, unlicensed frequency bands allow spectrum sharing. Recent spectrum measurements indicate that many licensed bands remain relatively unused for most of the time. Therefore, allowing secondary users (users without a license to operate in the band) to operate with minimal or no interference to primary users is one way of sharing spectrum to increase efficiency. Recently, Federal Communications Commission (FCC) has opened up licensed bands for opportunistic use by secondary users. A cognitive radio (CR) is one enabling technology for systems supporting opportunistic use. A cognitive radio adapts to the environment it operates in by sensing the spectrum and quickly decides on appropriate frequency bands and transmission parameters to use in order to achieve certain performance goals. A cognitive radio network (CRN) refers to a network of cognitive radios/secondary users.

In this dissertation, we consider a competitive CRN with multiple channels available for opportunistic use by multiple secondary users. We also assume that multiple secondary users may coexist in a channel and each secondary user (SU) can use multiple channels to satisfy their rate requirements. In this context, firstly, we introduce an integrated modeling and forecasting tool that provides an upper bound estimate of the number of secondary users that may be demanding access to each of the channels at the next instant. Assuming a continuous time Markov chain model for both primary and secondary users activities, we propose a Kalman filter based approach for estimating the number of primary and secondary users. These estimates are in turn used to predict the number of primary and secondary users in a future time instant. We extend the modeling and forecasting framework to the

case when SU traffic is governed by Erlangian process. Secondly, assuming that scheduling is complete and SUs have identified the channels to use, we propose two quality of service (QoS) constrained resource allocation frameworks. Our measures for QoS include signal to interference plus noise ratio (SINR) /bit error rate (BER) and total rate requirement. In the first framework, we determine the minimum transmit power that SUs should employ in order to maintain a certain SINR and use that result to calculate the optimal rate allocation strategy across channels. The rate allocation problem is formulated as a maximum flow problem in graph theory. We also propose a simple heuristic to determine the rate allocation. In the second framework, both transmit power and rate per channel are simultaneously optimized with the help of a bi-objective optimization problem formulation. Unlike prior efforts, we transform the BER requirement constraint into a convex constraint in order to guarantee optimality of resulting solutions. Thirdly, we borrow ideas from social behavioral models such as Homo Egualis (HE), Homo Parochius (HP) and Homo Reciprocan (HR) models and apply it to the resource management solutions to maintain fairness among SUs in a competitive CRN setting. Finally, we develop distributed user-based approaches based on "Dual Decomposition Theory" and "Game Theory" to solve the proposed resource allocation frameworks. In summary, our body of work represents significant ground breaking advances in the analysis of competitive CRNs.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude and profound indebtedness to Dr. Bala Natarajan for his constant guidance and endless patience throughout the progress of this work.

I would like to thank Dr. Caterina Scoglio for her continuous encouragement, inspiration and serving on my committee. I would like to thank Dr. Sanjoy Das, Dr. Jim Neil and Dr. Larry Weaver for serving on my committee, their encouragement and valuable suggestions.

Being a part of WiCom family, I have good memories only. Thanks to my research groupmates- Rajet, Krithika, Ahmad, Dalin, Narayanan, Anirudh, Mark, Sunitha and fellows Mina, Nikkie for their friendship, encouragement and support. My special thanks to Rajet for his friendship. Thanks to Farhana, Sohini, Saad, Shiplu to make my days in Manhattan so live.

Thanks to the authority who run the coffee shop "Cafe Q." I used to feel immensely good to start my day with a cup of "Cafe Q" coffee. To me, "Cafe Q" coffee is the best coffee in United States. I will miss coffee from "Cafe Q" in the rest of my life.

At this great moment, I would like to remember two other mentors of mine, Polock and Tuhin and say my thanks for their contributions and influences in my career and life.

Thanks to my sister in law Shanta, my cute niece Samarah, my siblings Anu and Muna, my parents in laws and my husband Farabi for their encouragement and support.

Dedication

To my parents and elder brother.

Chapter 1

Introduction

In this chapter, we provide a brief background on cognitive radio networks (CRNs). We introduce the challenges in CRNs and present an overview of prior efforts. We highlight the motivation for this research, followed by a summary or key contributions and organization of the dissertation.

1.1 Cognitive Radio Networks

The usage of radio spectrum and the regulation of radio emissions are coordinated by national regulatory bodies. In United States, the main authorities for radio spectrum regulation are the Federal Communications Commission (FCC) for commercial use and the National Telecommunications and Information Administration (NTIA) for government use. Historically, regulatory authorities divide spectrum into blocks. Licenses are issued for exclusive access for a given geographical region to some of the blocks. The blocks are termed as licensed spectrum/bands and users with the right to access the licensed bands are referred to as primary users. The regulatory authorities also allocate some spectrum blocks (e.g., 900 MHz band, 2.4 GHz Industrial, Scientific and Medical (ISM) band, 5 GHz Unlicensed National Information Infrastructure (UNII) band) where users can operate without any license. These blocks are called unlicensed bands. Traditionally, licensed bands are exclusively reserved for use by the primary license holders (primary users). Whereas, unlicensed bands promote coexistence of dissimilar radio systems in the same spectrum. As an

example, in the ISM band, Bluetooth Wireless Personal Area Network (WPAN) coexist with the IEEE802.11 Wireless Local Area Network (WLAN), cordless phones, radio frequency identification (RFID) cards and microwave ovens. As a result, there has been a growing interest of increasing spectral efficiency by shifting from "exclusive spectrum usage rights policy" to "shared spectrum policy".

Several recent measurements of spectrum usage indicate that many licensed spectrum bands remain relatively unused for most of the time [1, 2, 3, 4]. The measurements taken by Berkeley Wireless Research Center show that the allocated spectrum is vastly underutilized [1]. Measured results of spectrum usage activity (green means no activity) in downtown Berkeley, California are shown in Fig. 1.1. It has been reported in [2] that utilization in 30-300 MHz spectrum band is only 5.2%. In [3], it has been reported that utilization of spectrum below 3 GHz can be as low as 15%. The rest 85% of the time, unused spectrum can be allocated to "secondary users"- users without a license to operate in the band. Spectrum sharing has shown to increase spectrum utilization and has been proven to be successful and commercially practical. Allowing secondary users to operate with minimal or no interference to primary users is one way of spectrum sharing. Therefore, FCC has opened up licensed bands for opportunistic use by secondary users since 2004 [5].

In order to enable secondary users to coexist with other users in a frequency band, the radio receiver must be opportunistic [6]. Opportunistic users must quickly identify and exploit available frequency bands/channels. They must also be willing to be ready to be interrupted and look for other channels to complete transmission. This concept is called Dynamic Spectrum Access (DSA). DSA requires the radio to have the following features: (i) *Intelligence:* the radio must sense the environment and identify spatial, temporal or spectral voids. (ii) *Programmable:* the radio must be programmable to change transmission parameters such as power level, modulation order, operating frequency, transmission bandwidth, coding rate, frame size to achieve certain performance goals/QoS objectives. (iii) *Agility:* the radio must be able to hop quickly to available channels. (iv) *Broadband:* the radio must

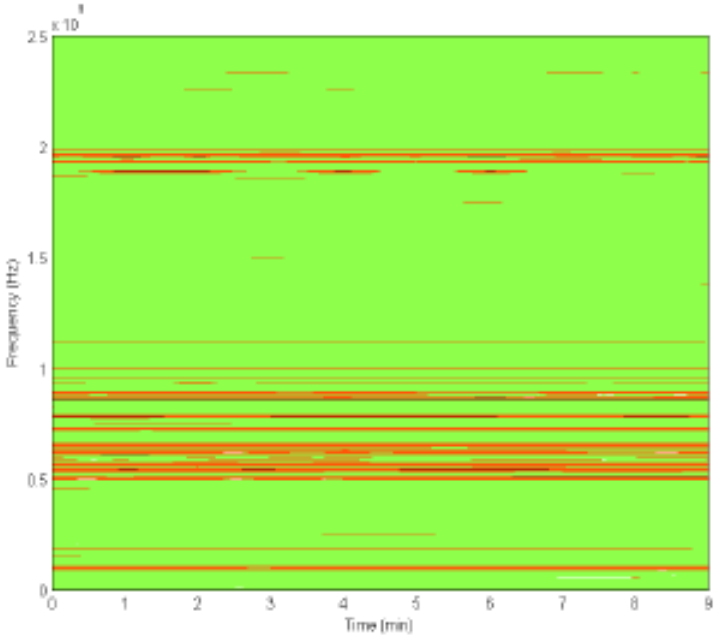

Figure 1.1: *Measurements of spectral usage activity in downtown, Berkeley ([1]).*

be able to scan a large number of channels in order to increase the probability of correctly identifying free channels to avoid long interruption in transmission. (v) *Low cost:* the cost of the radio has to be comparable with current technology. (vi) *Low power consumption:* power consumption of the radio has to be comparable with current wireless devices.

A cognitive radio (CR) has been considered as a possible enabling solution for a DSA system. The concept of cognitive radios was first introduced by Joseph Mitola III [7]. Joseph Mitola III defines cognitive radio as an extension of software defined radio (SDR) that employs model-based reasoning about users, multimedia content and communication context. The cognition cycle employed in investigated cognitive radio architectures at Royal Institute of Technology (KTH) is shown in Fig. 1.2. In [7], cognitive radio is defined as a goal-driven framework in which the radio senses the environment, infers context, assesses alternatives, generates plans, supervises multimedia services and learns from its mistakes. As a DSA compatible system, cognitive radio adapts to the environment by sensing the spectrum and takes quick decision on appropriate transmission parameters to achieve certain

performance goals. A cognitive radio can be envisioned as shown in Fig. 1.3 [8]. Figure 1.3 depicts how the environmental parameters (as dials) and transmission parameters (as knobs) interact and are used in a cognitive radio.

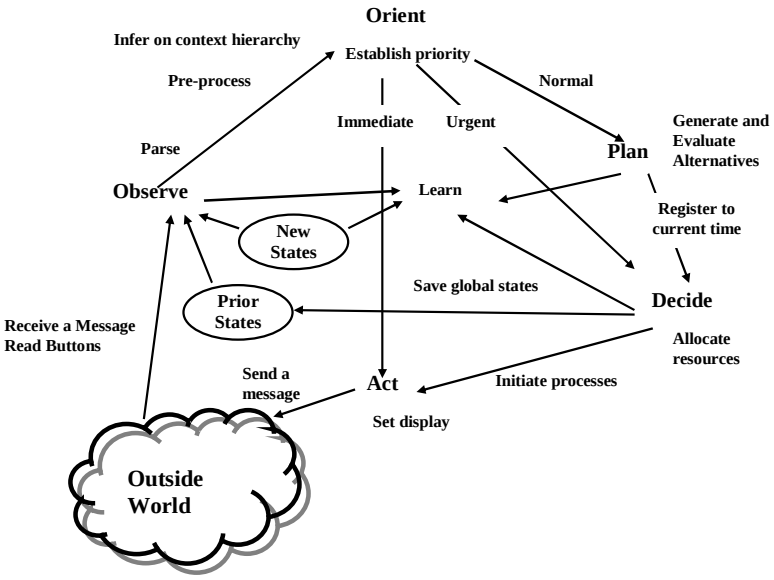

Figure 1.2: *Cognition cycle ([7]).*

A cognitive radio network (CRN) is defined as a network of cognitive radios/secondary users. The first task that a secondary user (SU) in CRN needs to perform is sensing the activity of primary users in its intended channel/channels. All SUs in CRNs maintain QoS through the transmission duration by dynamically seeking out the best transmission strategies (e.g., channel, rate, transmit power).

In the following section, we introduce some of the challenges in CRNs and describe related prior work.

1.2 Challenges in CRNs and Prior Work

The challenges in CRNs can be broadly listed as

1. Cognitive radio architecture and implementation issues [6]-[12],

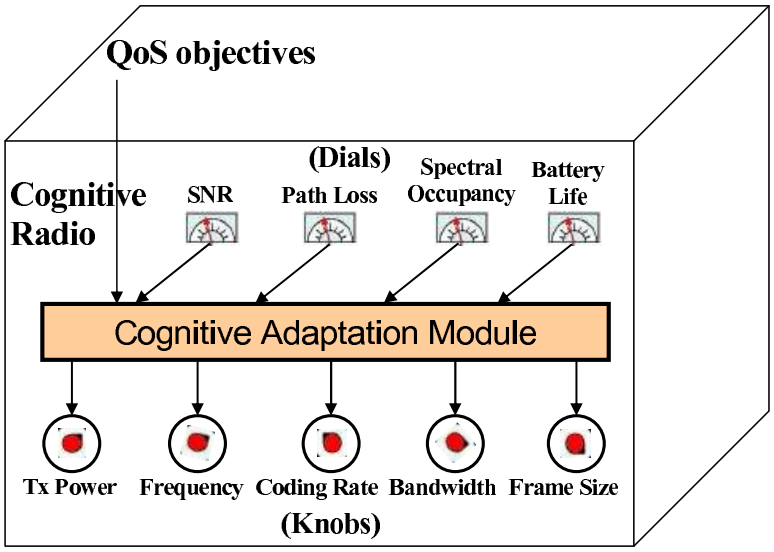

Figure 1.3: *Visual representation of cognitive radio dials and knobs ([8]).*

2. Spectrum sensing hardware requirements [6]-[11],

3. Spectrum sensing algorithms [13]-[29],

4. Resource management [30]-[53],

5. Fairness in resource allocation [54]-[66],

6. Policy challenges [67]-[71].

In this dissertation, we are primarily focused on addressing challenges in resource management and fairness which are implicitly related to spectrum sensing algorithms. Therefore, we present an overview of prior work in those areas in the following subsections.

1.2.1 Spectrum Sensing

Spectrum sensing is the most important and first task for a SU in CRNs. Spectrum sensing technique in CRNs has been widely studied. The sensing function requires the radio to step through a set of frequency bands and perform signal detection for each frequency band. Different sensing techniques such as matched filter, energy detector and cyclostationary

feature detection have been reviewed in [13]. Matched filtering [13] is the optimum method for detection of primary user (PU). Effectively, a matched filter does the demodulation of a PU signal. Hence, in matched filtering approach, SU is required to have *priori* knowledge of PU signal (such as modulation type, modulation order, pulse shaping, packet format). That is, a SU needs a dedicated receiver for every primary user class, which makes matched filtering approach impractical in CRNs. Energy detector [13] detects the signal by comparing the output of energy detector to a preset threshold and does not require any information of PU signal. Threshold setting requirement makes the energy detector vulnerable to noise and interference level. Cyclostationary feature detection approach [13] is based on properties of the modulated signal. Modulated signals are in general coupled with sine wave carriers, pulse trains or cyclic prefixes which result in built in periodicity. This periodicity helps to extract the information such as modulation, pulse shape of the received signal. In cyclostationary feature detection, instead of power spectral density, cyclic correlation function is used for detecting the signals present in the spectrum. A modified version of cyclostationary feature detector to improve spectrum sensing performance at low signal to noise ratio (SNR) is proposed in [14]. The modified detector in [14] performs the autocorrelation of received signal before the spectral correlation detection. The authors in [15] propose a sensing method to identify PU by estimating their radio frequency (RF) transmission parameters. The identification is done by matching the a *priori* information about PU transmission parameters to the features extracted from the received signal. In [16], the authors show a PU detection technique exploiting the local oscillator (LO) leakage power emitted by the RF front end of primary receivers. The authors in [18] derive a blind sensing algorithm based on oversampling the received signal or by employing multiple receive antennas. The proposed method in [18] does not require any information of PU signal or channel. The proposed method combines linear prediction and QR decomposition of the received signal matrix. Two signal statistics are computed from the oversampled received signal. The ratio of these two statistics indicates the presence/absence of the PU signal.

To increase spectrum sensing performance, network based sensing [24], cooperative sensing [25, 27, 28] have been proposed. The techniques in [24, 25, 27, 28] increase sensing time and are not well suited for practical implementation of cognitive radio in time sensitive operations. The authors in [29] propose the use of dedicated sensing receiver (DSR) that solely focuses on channel sensing and runs in parallel with a main receiver. Here, the authors also show that the DSR architecture provides up to a fivefold reduction in total mean detection time.

1.2.2 Resource Allocation

After the spectrum sensing phase, each SU in a CRN needs to identify its operating channel/channels, transmit power level, modulation type, modulation order, channel coding, spectral shaping etc. There have been significant research efforts related to determining optimal channel, transmit power, modulation type and rate for SUs [30, 31, 32, 33, 34, 35, 36, 8, 37, 38, 52, 53].

The authors in [30, 31] consider a CRN model with one PU and one SU coexisting in the same channel. In [30], power allocation strategies for the SU are developed with the objective of maximizing ergodic capacity under different constraints (such as limits on peak and average transmit and interference power). In [31], the authors develop a cognitive radio game to find optimal transmit power with the goal of minimizing total transmit power. Here, quality of service (QoS) is maintained for the PU (defined as minimum rate and maximum acceptable bit error rate (BER)). The authors claim that their formulation is applicable to the case when multiple SUs share a channel. In [32, 33], the authors consider a CRN model with one PU and multiple SUs coexisting in the same channel. Here, the authors develop distributed power allocation strategies for the SUs. The authors in [34, 35, 36] consider a system model where multiple SUs coexist in a channel. In [34], the authors design a power control game with a utility function of maximizing transmission rate to find transmit power. In [35], the authors study both centralized and distributed auction mechanisms to

allocate receive powers. They consider an objective function of maximizing utility which is a function of signal to interference plus noise ratio (SINR). The authors in [36] design a convex optimization problem to find optimal transmit power. A lower bound on SINR is used as a QoS constraint for secondary users. A distributed suboptimal joint coordination and power control mechanism to allocate transmit powers to secondary users is also presented in [36].

A genetic algorithm driven cognitive radio decision engine is employed to determine the optimal transmission parameters (transmit power, modulation type and rate (modulation order)) for a SU in both single and multi-carrier based CRNs in [8]. The approach presented in [8] suffers from numerous drawbacks. Genetic algorithms are notorious for slow convergence and high complexity. Therefore, their implementation is not suitable for time varying environments as well as delay sensitive applications. Additionally, the underlying optimization problem in [8] has non-convex fitness function which in turn implies that the optimality of the genetic algorithm based solution cannot be guaranteed.

In [37, 38], joint allocation of channel and transmit power for multi-channel multiuser CRN has been studied. The authors in [38] apply game theory to develop distributed power allocation algorithm. However, in [37], coexistence of multiple secondary users in a channel has not been considered. Also, in [37, 38], the QoS requirement of SUs has been ignored. In [72], the authors propose two game theoretic approaches using potential game framework and ϕ-no-regret-learning schemes to allocate available channels to secondary users. Both approaches show better performance compared to random channel allocation. In [52], the authors propose a stochastic channel selection based on learning automata technique to maximize the probability of successful transmissions and to avoid frequent channel switchings in CRN. A biologically-inspired spectrum sharing algorithm based on the adaptive task allocation model in insect colonies to select channel has been presented in [53]. However, in [52, 53], though the authors have considered multiple channels to start with, only one secondary user is eventually assigned a channel.

1.2.3 Fairness in Resource Allocation

When multiple SUs compete for a limited number of available channels/frequency bands in CRNs, fairness among SUs in resource allocation is another important consideration. Fairness issues in resource allocation has garnered some attention in recent years [54, 55, 56, 57, 58, 66].

The authors in [54] focus on deriving fair (in terms of airtime share) random access protocol for dissimilar radio systems in open spectrum access scenario. In their proposed fair random access protocol, each radio system contends for the spectrum with a finite probability. The authors also propose a Homo Egualis (HE) society model based distributed approach to determine the contending probability. In [55], the authors propose a fair opportunistic spectrum access using fast catch-up strategy that reduces the first passage time (first passage time is the amount of time after which all SUs have equal access right to the available channels). In [56], the authors study three variants of utility functions to allocate spectrum in CRN under protocol interference model. The variants are Max-sum-Reward, Max-min-Reward and Max-Proportional Fair utility functions. The authors map the different spectrum allocation problems into color-sensitive graph coloring (CSGC) problem and consider binary geometry interference model. As the optimal graph coloring problem is NP-hard, the authors also present heuristic to solve the allocation problems. In [57], the authors study the joint spectrum allocation and scheduling in CRN with the objective to achieve a tradeoff between throughput and fairness while ensuring interference-free transmission at any time (taking into account both protocol and physical interference models). The authors in [57] transform the joint allocation and scheduling problem into a problem of finding all possible transmission modes and the active time fraction for each transmission mode. A transmission mode is composed of a subset of user-channel pairs which can be active concurrently. They define three joint spectrum allocation and scheduling problems and these are MAximum throughput Spectrum allocation and Scheduling (MASS), Max-min fair MAximum throughput Spectrum allocation and Scheduling (MMASS) and Proportional

fAir Spectrum allocation and Scheduling (PASS). The MASS problem finds a feasible rate allocation vector, all transmission modes along with a feasible transmission schedule vector such that the throughput is maximized. To avoid starvation of some users due to maximizing throughput, the authors consider a new variable called Demand Satisfaction Factor (DSF) into the scheduling problem. The DSF of a user is defined as the ratio of rate allocated to that user over its traffic demand. In MMASS, the rate that minimizes the maximum dissatisfaction with respect to demand is evaluated. The PASS problem finds a feasible rate allocation vector, all transmission modes along with a feasible transmission schedule vector such that the summation of the logarithmic of DSF is maximized. The authors in [57] also conclude that PASS formulation provide a better tradeoff between throughput and fairness compared to MASS or MMASS.

The authors in [58] develop a set of resource allocation exploiting fairness axioms of game theory that provides fairness in allocating the "extra" resources available after satisfying the minimum requirements of primary users. In [66], the authors find optimal transmit power for users in wireless cellular and ad hoc networks considering proportional and minmax fairness. In proportional fairness resource allocation scheme formulation, the authors consider a static weight for each user and use the weight into resource allocation scheme. In minmax fairness resource allocation scheme formulation, the transmit power that maximizes the minimum signal to interference ratio is determined.

1.3 Motivation and Overview

We believe that in a practical CRN, (1) multiple channels may be available, and (2) multiple SUs may compete for available resources. To increase spectral efficiency, multiple SUs may coexist in a channel. Also, channels may be of different quality. Therefore, the SUs assigned to higher quality channels may hold an advantage over SUs assigned to the poorer channels and rate requirement of some secondary users may not be satisfied by allocating one channel to a user. That is, in practice, a single SU may occupy more than one channel.

Resource management in such a practical CRN is an important consideration that has not yet been addressed. The primary research question we answer in this dissertation is: *How can we optimize transmit power and rate (modulation order) in a competitive CRN while maintaining QoS for SUs?* Before we answer this question, the "coexistence of multiple SUs in a channel" motivates our first task in this dissertation. In a competitive environment where multiple SUs coexist in a single channel, one can expect the QoS of one user to depend on the number and behavior of other SUs. Much of the prior work in CRNs has primarily been focused on sensing primary users with very little emphasis on how multiple secondary users may compete for available spectrum. Therefore, forecasting the behavior of secondary users is equally critical in the successful operation of a CRN. For example, if a spectrum band is determined to be free and a large number of secondary users decide to use this spectrum band simultaneously, the QoS or BER performance of the secondary users will degrade due to high level of interference. Therefore, it is important to investigate strategies for enabling SUs to sense and predict the behavior of both primary and competing secondary users in a frequency band of interest. In this context, firstly we are motivated to *introduce an integrated modeling and forecasting framework for monitoring spectrum use by primary and secondary users in CRNs.*

Next, we assume scheduling is complete and SUs have identified the channels to use. Now, our goal is to determine the optimal transmit power and rate (modulation order) that competing SUs need to employ in each channel. In this context, we propose two resource allocation frameworks. In both frameworks, our objective is to determine the optimal distribution of power and rate that a secondary user has to employ across the channels that it uses in order to (1) minimize total power consumption; (2) maximize rate, and (3) maintain QoS. Our measures for QoS include SINR or BER and minimum rate requirement.

The proposed resource allocation frameworks provide the optimal transmit power and rate across the channels for all SUs for a given time instant. However, users may not be

satisfied with optimal allocation of resources based on instantaneous QoS. An example of dissatisfaction among SUs may arise when two SUs with different minimum rate requirements are allocated the same rate. Another example of dissatisfaction among SUs may arise when a user expends higher power relative to other users. Typically dissatisfaction is a feeling that develops over time. Hence, unlike prior efforts in resource allocation, it is imperative to consider user experience over time. Subsequently, the question that needs to be addressed is *"Can we maintain fairness in user experiences over time?"* To answer this question, we borrow ideas from social behavioral models and apply it to the resource management solutions in a competitive CRN setting.

The centralized solution of the proposed resource allocation frameworks demand extensive control signalling and is difficult to implement in practice if information exchange about all users and channels is limited. In this context, we are motivated to develop distributed user-based approaches to solve the proposed resource allocation frameworks.

1.4 Contributions

We consider a CRN with multiple channels available for opportunistic use by multiple SUs. We also assume that multiple SUs may coexist in a channel and each SU can use multiple channels to satisfy their rate requirements. The CRN operation is shown in Fig. 1.4. Figure 1.4 tells that each SU scans the spectrum at regular intervals and starts transmitting on particular channel/channels once it determines that the channel/channels will not be used by a PU. At any instant of time, SUs ceases transmission through channel/channels if PU enters in that particular channel/channels. The key contributions of this dissertation for such a CRN setting are summarized in this section.

1. Assuming that the spectrum usage of various channels are independent, *for the first time we present an integrated modeling and forecasting framework for monitoring spectrum use by primary and secondary users in a competitive CRN.*

   (a) Assuming a continuous time Markov chain model for both primary and secondary

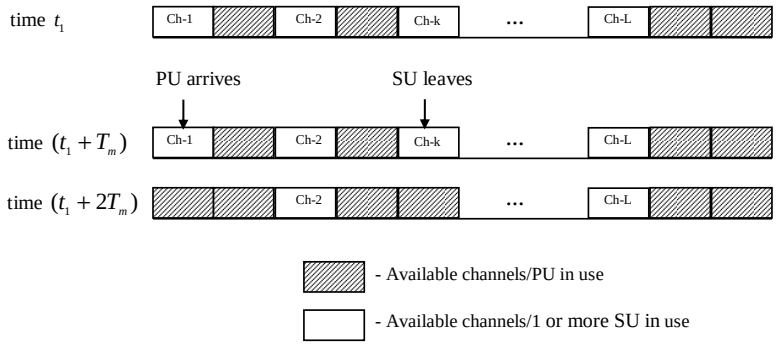

Figure 1.4: *CRN operation (T_m denotes measurement interval).*

users activities, we propose a Kalman Filter based approach for estimating the number of primary and secondary users. These estimates are in turn used to predict the number of primary and secondary users in a future time instant.

(b) Using both simulated data and measured power levels in the 2.4 GHz unlicensed band, we demonstrate the implementation of both the modeling and forecasting aspects of the proposed approach. We observe that our proposed forecasting technique, not only provides a good upper bound prediction for the number of primary and secondary user, it is also robust to model parameter estimation errors.

(c) We extend the modeling and forecasting framework to the case when SU traffic is governed by Erlangian process, (i.e., the traffic model incorporates bulk arrival or bulk departure scenarios).

(d) Knowledge of the upper bound provides valuable information to a SU interested in using a spectrum band that is already being used by other secondary users. Detailed analysis of proposed modeling and forecasting tools are provided in chapter 2 and in our papers [73, 74, 75].

2. *We propose two centralized resource allocation frameworks for resource allocation to secondary users in a competitive CRN.*

(a) In the first framework, we propose a two-stage process. In the first stage, optimal choice for transmit power is determined for all SUs subject to maintaining a given SINR in each channel used. Using this power/SINR result, the optimal distribution of rate (modulation order) is determined in the second stage. In the second stage, the rate distribution is formulated as a maximum flow problem in graph theory. We also propose a simple heuristic to determine the rate allocation.

(b) In the second framework, we jointly determine the best choice of power and rate distribution for every SU with the help of a bi-objective problem formulation.

(c) Unlike prior efforts, we have transformed the BER constraint in both frameworks into a convex constraint in order to ensure optimality of our resulting solutions.

(d) In both frameworks, we observe that optimal transmit power follows reverse water filling process and optimal rate allocation is proportional to SINR.

(e) In terms of total power (i.e., net transmission cost), the joint optimization framework is more economical relative to the two-stage optimization framework. This is because, the joint formulation offers more degrees of freedom with the ability to adapt both power and rate simultaneously in order to achieve a certain BER. In the two-stage optimization framework, either power or rate is available to adapt to achieve a certain SINR or BER, respectively. Detailed analysis of the proposed resource allocation frameworks are provided in chapters 3, 4 and in our paper [76].

3. In a competitive CRN, *for the first time we determine optimal power and rate distribution choices for each SU while maintaining fairness in current and prior history of user experience with respect to QoS among SUs.*

(a) We quantify user experience over time by introducing dynamic fairness weights for each SU in the resource allocation framework.

(b) The dynamics of the weights are governed by social behavioral models. We study

the effect of Homo Egualis (HE), Homo Parochius (HP) and Homo Reciprocan (HR) models.

(c) We observe that considering dynamic fairness weights in the resource allocation scheme provide a better system level fairness index (as defined by Jain in [77]) relative to the unweighted allocation scheme. Detailed analysis of the proposed resource allocation framework is provided in chapter 5 and in our papers [78, 79].

4. *With the help of dual decomposition theory [80], we develop three (3) user-based distributed approaches to solve the resource allocation framework introduced in 2.* We observe that the solution from each distributed implementation for both frameworks follows the centralized solution. Detailed analysis of the proposed distributed algorithms are provided in chapters 3, 4 and in our papers [81, 82].

5. Finally, *we develop game theory based implementation for joint resource allocation framework.* We analyze existence of Nash Equilibrium for the game. We develop an algorithm to reach Nash Equilibrium. Detailed analysis of the proposed distributed algorithm is provided in chapter 6 and in our paper [83].

1.5 Organization

The dissertation is organized into seven chapters. Chapter 2 presents the modeling and forecasting tool for secondary users activity. We develop estimation and forecasting tools for both Poissonian and Erlangian traffic model activities of secondary users. Numerical results on predictor performances for both traffic cases are also provided. In chapter 3, we introduce the two-stage resource allocation framework. We present dual based distributed approaches to solve stage 1. Optimal and heuristic rate distribution for stage 2 are provided. Additionally, a comparison on centralized and distributed power allocation, and optimal and heuristic rate allocation are shown. Chapter 4 contains the joint resource allocation framework along with its dual distributed implementations. Besides, a comparison on resource

allocation between two-stage and joint allocation schemes as well as between centralized and distributed joint allocation schemes are provided. In chapter 5, we describe the resource allocation framework incorporating dynamic fairness weights. We present the analogy between the social behavior of human beings and that of SUs in CRN, and the society models of interest to this work. Chapter 6 shows the game theoretic implementation of joint resource allocation framework. Here, we analyze the existence of the Nash Equilibrium. We develop an algorithm to reach the NE. Finally, chapter 7 concludes the dissertation and possible future directions and extensions of this work.

Chapter 2

Spectrum Usage Modeling and Forecasting

In this chapter, we seek a solution for the first task in chapter 1. Specifically, we introduce an integrated modeling and forecasting approach that SUs in a CRN can use to predict spectrum usage/availability based primarily on power level measurements. Our modeling and forecasting setup incorporates traffic behavior of both primary and competing secondary users in a spectrum band of interest. Firstly, by considering a continuous time Markov chain traffic model for PU and SUs, we propose a Kalman filter approach to estimate the number of primary and secondary users at a given time instant. Based on these estimates, we determine robust upper bound forecasts of the number of primary and secondary users for a future time instant. Secondly, we generalize the SU traffic model and develop estimation and forecasting tool accordingly.

It is important to remember that this chapter offers a modeling and forecasting tool and not algorithms for spectrum sharing among multiple secondary users. Additionally, note that even though we use the words spectrum band, frequency band and channel interchangeably throughout the chapter, they convey the same meaning.

The rest of the chapter is organized as follows. In Sec. 2.1, we describe the proposed spectrum usage model in detail for Poissonian traffic model. Section 2.2 presents the Kalman filtering techniques to estimate the number of primary and secondary users and Sec. 2.3

describes the proposed forecasting techniques for both primary and secondary users. Experimental results illustrating the application of the modeling and forecasting methods are provided in Sec. 2.4. Section 2.5 develops the estimation and prediction tool for SU generalized traffic model and also presents the performance of the predictors. Finally, conclusions are presented in Sec. 2.6.

2.1 System Model

We consider L channels within a CRN as shown in Fig. 1.4. Each of the L channels can be used by either a PU or one or more secondary users. Each SU uses the spectrum opportunistically. That is, every SU scans the spectrum at regular intervals and starts transmitting on a particular channel once it determines that the channel will not be used by a PU.

Typically, researchers focus on the sensing aspect of secondary users in order to determine if a channel is available for transmission. In this chapter, in addition to determining the presence or absence of the primary user in a given channel, we also consider the impact of multiple secondary users utilizing a single channel. For example, if a channel is determined to be free and a large number of secondary users decide to use this channel simultaneously, the QoS (SINR or BER) performance of the secondary users will be poor due to high level of interference from other secondary users. Consequently, it is important for every SU to monitor the spectrum usage by other secondary users and this aspect is captured in our proposed modeling and forecasting set up. Additionally, we assume that the spectrum usage of various channels are independent. Therefore, in the rest of the chapter, we restrict ourselves to the modeling and forecasting of spectrum use to one channel. It is easy to extend the idea presented in the following sections to the case of correlated channels.

We assume that both the PU and secondary users follow Poisson arrival process with rates, λ_p and λ_s, respectively. Their negative exponential service time distributions have rates, μ_p and μ_s, respectively. A similar model for arrival and departure processes of PU

and secondary users have been assumed in [54, 84]. The maximum number of PU and SU are N_p and N_s, respectively. For ease in presentation, N_p is assumed to be equal to 1, i.e., we assume that PU is either present or absent. In other words, PU follows a two-state ON-OFF Markov process. The two-dimensional state-transition-rate diagram for the CRN is shown in Fig. 2.1.

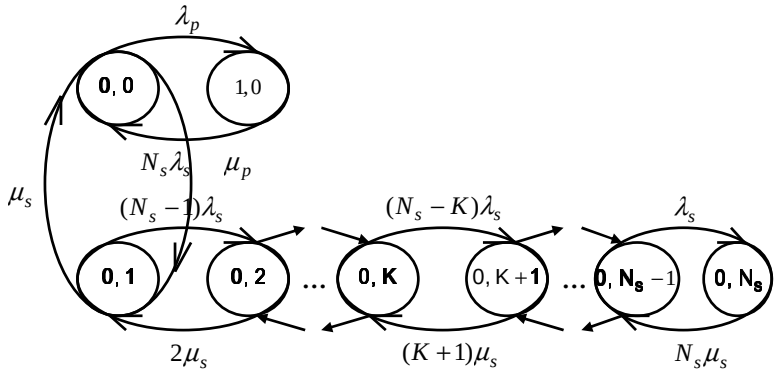

Figure 2.1: *Two-dimensional state-transition-rate diagram of PU and SU.*

Each state in the model is denoted as $(n_p,\ n_s)$, where n_p and n_s represent the number of primary user and secondary users, respectively. From this state-transition-rate diagram and concepts from queueing theory [85], the differential equations for the state probabilities, $pr_{i,k}(t)$ can be evaluated for both primary (with $i = p$) and secondary (with $i = s$) users. The state probability is defined as

$$pr_{i,k}(t) \triangleq \text{prob}\{x_i(t) = k\}, \tag{2.1}$$

where, $x_i(t)$ is the number of users at time t and k indicates that number. Since we have an ON-OFF traffic model for PU, $k = 0$ or 1 when $i = p$. On the other hand, for secondary users, i.e., when $i = s$, k can vary from 0 to N_s provided that there is no primary user in

the channel. In general, the differential equations for the state probabilities correspond to

$$\frac{d}{dt}pr_{i,0}(t) = \mu_i pr_{i,1}(t) - N_i\lambda_i pr_{i,0}(t), \tag{2.2}$$

$$\frac{d}{dt}pr_{i,k}(t) = (N_i - k + 1)\lambda_i pr_{i,(k-1)}(t) + (k+1)\mu_i pr_{i,(k+1)}(t) - (k\mu_i + (N_i - k)\lambda_i)pr_{i,k}(t),\ 1 \leq k\ <\ N_i, \tag{2.3}$$

$$\frac{d}{dt}pr_{i,N_i}(t) = \lambda pr_{i,(N_i-1)}(t) - N_i\mu_i pr_{i,N_i}(t), \text{ for } i = p,\ s. \tag{2.4}$$

Typically, we are interested in determining the expected number of primary and secondary users in a given time instant. Therefore, the $E\{x_i(t)\}$ can be written as

$$E\{x_i(t)\} = \sum_{k=0}^{N_i} k pr_{i,k}(t), \text{ for } i = p,\ s. \tag{2.5}$$

Hence,

$$\frac{d}{dt}E\{x_i(t)\} = \sum_{k=0}^{N_i} k\frac{d}{dt}pr_{i,k}(t), \text{ for } i = p,\ s. \tag{2.6}$$

Let $\mathbf{L}_Q = [0\ 1\ 2 \cdots\ N_i]^T$. From Eqs. (2.2)-(2.4), (2.6), we can write

$$\frac{d}{dt}E\{x_i(t)\} = \mathbf{L}_Q^T\dot{\mathbf{P}}_i = \mathbf{L}_Q^T\mathbf{Q}_i\mathbf{P}_i, \tag{2.7}$$

where,

$$\mathbf{Q}_i = \begin{bmatrix} -N_i\lambda_i & \mu_i & 0 & \cdot & 0 \\ N_i\lambda_i & -[(N_i-1)\lambda_i + \mu_i] & 2\mu_i & \cdot & 0 \\ 0 & (N_i-1)\lambda_i & -[(N_i-2)\lambda_i + 2\mu_i] & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \lambda_i & -N_i\mu_i \end{bmatrix}$$

and

$$\mathbf{P}_i = \begin{bmatrix} pr_{i,0}(t) \\ pr_{i,1}(t) \\ . \\ . \\ . \\ pr_{i,N_i}(t) \end{bmatrix}, \text{ for } i = p,\ s.$$

In Eq. (2.7), $(\cdot)^T$ indicates matrix or vector transpose operator. $\mathbf{L}_Q^T \mathbf{Q}_i \mathbf{P}_i$ is obtained as $N_i\lambda_i - (\lambda_i + \mu_i)E\{x_i(t)\}$. Therefore,

$$\frac{d}{dt}E\{x_i(t)\} = N_i\lambda_i - (\lambda_i + \mu_i)E\{x_i(t)\}. \tag{2.8}$$

We assume that measurements are performed at discrete time instants mT_m, $m = 1,\ 2,\ 3, \cdots$ for a given value T_m. Using the initial condition that the number of users at time $t = (m-1)T_m$ is $x_i(m-1)$, the solution of Eq. (2.8) is obtained as

$$E\left[x_i(m)|x_i(m-1)\right] = x_i(m-1)e^{-T_m(\lambda_i+\mu_i)} + \frac{N_i\lambda_i}{(\lambda_i+\mu_i)}\left[1 - e^{-T_m(\lambda_i+\mu_i)}\right]. \tag{2.9}$$

Therefore, it is possible to express the number of users (primary or secondary) at time mT_m in terms of the number of users (primary or secondary) at time $(m-1)T_m$ as

$$x_i(m) = A_i x_i(m-1) + B_i, \tag{2.10}$$

where,

$$A_i = e^{-T_m(\lambda_i+\mu_i)} \tag{2.11}$$

and

$$B_i = \frac{N_i\lambda_i}{(\lambda_i+\mu_i)}\left[1 - e^{-T_m(\lambda_i+\mu_i)}\right]. \tag{2.12}$$

Equation (2.10) shows us the relationship between the number of users at two successive measurement instants and in most general case, Eq. (2.10) corresponds to

$$x_i(m) = A_i x_i(m-1) + B_i u_i(m) + w_i(m), \text{ for } i = p,\ s. \tag{2.13}$$

Specifically, we can write down the state equations for primary and secondary users as

$$x_p(m) = A_p x_p(m-1) + B_p u_p(m) + w_p(m) \tag{2.14}$$

and

$$x_s(m) = A_s x_s(m-1) + B_s u_s(m) + w_s(m), \tag{2.15}$$

where, $x_p(m)$ and $x_s(m)$ represent the number of primary and secondary users using the spectrum, respectively, at the measurement instant m. The parameters B_p and B_s relate the optional control inputs $u_p(m)$ and $u_s(m)$, respectively to states. Equation (2.10) suggests that $u_p(m)$ and $u_s(m)$ are equal to 1 for our system model. $w_p(m)$ and $w_s(m)$ are the process noise and assumed to be zero mean Gaussian noise with variances σ_p^2 and σ_s^2, respectively. The parameters A_p and A_s relate the states at previous and current measurement instants, in the absence of either a driving function or process noise. A_p and A_s are assumed to be constant over the analysis or vary very slowly.

The received power at a secondary user terminal during the measurement instant m consists of relative power level increments caused by both primary and secondary users and in the most general case corresponds to,

$$y(m) = C_p x_p(m) + C_s x_s(m) + D + v(m), \tag{2.16}$$

where, $y(m)$ is received power in dBm; C_p and C_s represent the relative increase in power level (in dB) due to the presence of primary and secondary users, respectively; D represents the background thermal noise and $v(m)$ denotes the measurement noise which may arise due to miscalculation, misalignment of timings and is assumed to be zero mean Gaussian noise with variance σ_v^2. $y(m)$ is the only measurable variable in the system.

In a CRN, the primary user has the right of use. Therefore, modeling and forecasting tasks are more critical for secondary users. From a secondary user perspective, we assume two modes of operation. Once a SU decides to use a channel opportunistically, it will enter a "learning and modeling phase." In this phase, the SU measures the power levels in the

channel and estimates traffic parameters of primary and other secondary users. An example of this estimation process is discussed in Sec. 2.4 where a time series based approach is used for modeling. In the next phase, the SU becomes an "active" participant in opportunistic spectrum sharing process. In the "active" phase, each secondary user continues to sense the received power level in order to forecast the use of spectrum by other secondary users. This process is illustrated in Fig. 2.2.

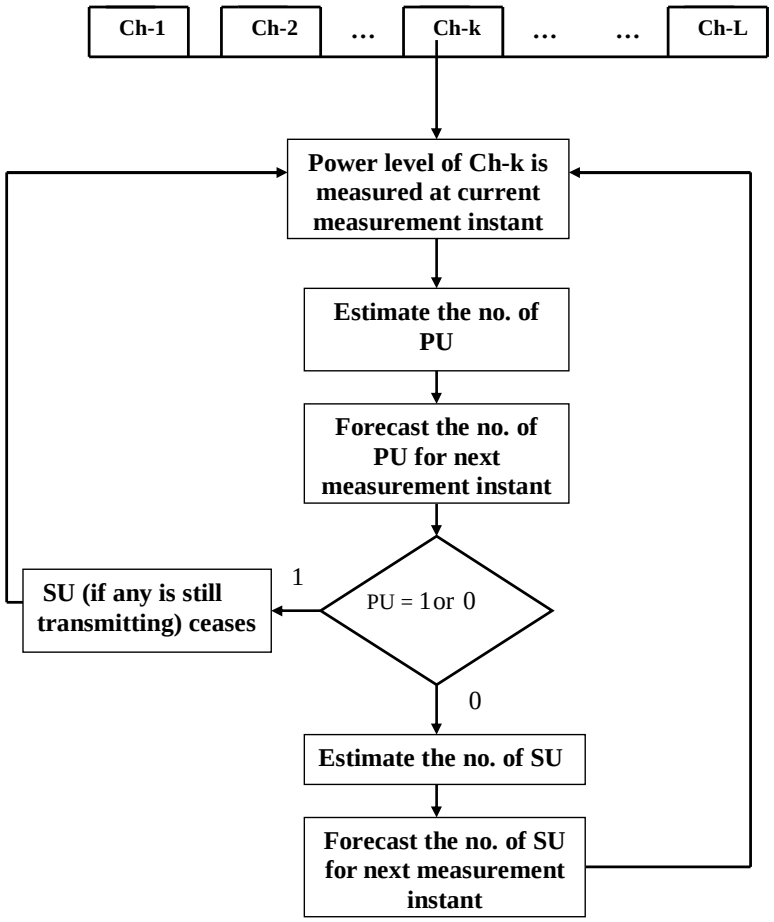

Figure 2.2: *Operation of the system on a single channel's spectrum usage in "active phase."*

2.2 Estimation of Spectrum Usage

In Sec. 2.1, we introduced the traffic models for the number of primary and secondary users using a given channel. In this section, we propose a Kalman filter based state estimation technique (i.e., estimating the number of primary and secondary users) based on the model from Sec. 2.1. The state estimation based on Kalman filter is summarized below:

State Equation: The state of the CRN can be represented by a vector $\mathbf{x} = [x_p \; x_s]^T$. The corresponding state equation is

$$\mathbf{x}(m) = \mathbf{A}\mathbf{x}(m-1) + \mathbf{B} + \mathbf{w}(m), \tag{2.17}$$

where,

$$\mathbf{A} = \begin{bmatrix} A_p & 0 \\ 0 & A_s \end{bmatrix}, \mathbf{B} = \begin{bmatrix} B_p \\ B_s \end{bmatrix}$$

and $\mathbf{w}(m)$ is a vector white Gaussian noise with mean $\mathbf{0}$ and covariance matrix,

$$\mathbf{\Sigma}_w = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_s^2 \end{bmatrix}.$$

Measurement Equation:

$$y(m) = \mathbf{C}^T\mathbf{x}(m) + D + v(m), \tag{2.18}$$

where $\mathbf{C}^T = [C_p \; C_s]$. Based on the state and measurement equations Eqs. (2.17) and (2.18), respectively, the Kalman filtering steps are given below:

Step 1: Initialization

$$\widehat{\mathbf{x}}(0|0) = \begin{bmatrix} E(x_p(0)) \\ E(x_s(0)) \end{bmatrix} \text{ and } \mathbf{M}(0|0) = \begin{bmatrix} \sigma_p^2(0) & 0 \\ 0 & \sigma_s^2(0) \end{bmatrix}. \tag{2.19}$$

Step 2: Prediction

$$\widehat{\mathbf{x}}(m|m-1) = \mathbf{A}\widehat{\mathbf{x}}(m-1|m-1) + \mathbf{B}, \forall\, m \tag{2.20}$$

$$\mathbf{M}(m|m-1) = \mathbf{A}\mathbf{M}(m-1|m-1)\mathbf{A}^T + \mathbf{\Sigma}_w, \forall\, m \tag{2.21}$$

Step 3: Kalman gain vector calculation

$$\mathbf{K}(m) = \mathbf{M}(m|m-1)\mathbf{C}\left(\mathbf{C}^T\mathbf{M}(m|m-1)\mathbf{C} + \sigma_v^2\right)^{-1}, \forall\, m \tag{2.22}$$

Step 4: Correction

$$\widehat{\mathbf{x}}(m|m) = \widehat{\mathbf{x}}(m|m-1) + \mathbf{K}(m)\left(y(m) - \mathbf{C}^T\widehat{\mathbf{x}}(m|m-1) - D\right), \forall\ m \quad (2.23)$$

$$\mathbf{M}(m|m) = \{\mathbf{I} - \mathbf{K}(m)\mathbf{C}^T\}\mathbf{M}(m|m-1), \forall\ m \quad (2.24)$$

From Eqs. (2.19)-(2.24), we estimate the number of primary and secondary users. The estimated value for the number of secondary users denoted as $\widehat{x}_s(m|m)$ is reset to 0 if the predicted value for the number of primary user for $(m+1)$ th instant is 1. This is also depicted in the block diagram in Fig. 2.2. The prediction methods for both primary and secondary users is described in the following section.

2.3 Forecasting Spectrum Usage

In this section, we describe the forecasting techniques used to predict the activity of primary and secondary users in the channel of interest.

One approach for forecasting is to determine the likely state at the next time instant given that we have the state estimate for the current instant. For this, we need to calculate the probability of transitioning to another state at time $(m+1)T_m$. The transitioning probabilities can be determined starting from the following differential equation,

$$\dot{\mathbf{P}}_i = \mathbf{Q}_i\mathbf{P}_i, \text{ for } i = p,\ s \quad (2.25)$$

within the time interval $mT_m < t < (m+1)T_m$. The $\mathbf{P}_i$ and $\mathbf{Q}_i$ are as defined in Sec. 2.1. Equation (2.25) governs the evolution of state transition probabilities. The solution of Eq. (2.25) gives the state transition probabilities from the m th measurement instant. The existence of the solution of Eq. (2.25) depends on two conditions. The first condition is the diagonalizable property of matrix $\mathbf{Q}_i$. The second one is non-positive definiteness of matrix $\mathbf{Q}_i$. The matrix $\mathbf{Q}_i$ satisfies both conditions. It is reducible to a diagonal form $\mathbf{Q}_i = \mathbf{E}_i\mathbf{\Gamma}_i\mathbf{E}_i^{-1}$, where $\mathbf{\Gamma}_i$ is a diagonal matrix with eigenvalues of $\mathbf{Q}_i$, and $\mathbf{E}_i$ is the matrix of corresponding right eigenvectors. The eigenvalues can be found to be $\gamma_r = -r(\lambda_i + \mu_i)$

for $r = 0, \cdots, N_i$. Hence, the solution, for $t \in (mT_m, (m+1)T_m]$, is given by

$$\mathbf{P}_i = \mathbf{E}_i e^{\mathbf{\Gamma}_i t} \mathbf{F}_i, \text{ for } i = p, \ s; \tag{2.26}$$

where $\mathbf{F}_i$ is a constant vector determined from the initial condition ($i.e.$, $\widehat{x}_i(m|m)$) as

$$\mathbf{F}_i = (e^{\mathbf{\Gamma}_i m T_m})^{-1} \mathbf{E}_i^{-1} \mathbf{P}_{mT_m(i)}, \tag{2.27}$$

where $\mathbf{P}_{mT_m(i)}$ is a vector with all 0's except the $\widehat{x}_i(m|m))$ th element which is 1. Now, we compute state transitioning probability values for the instant $(m+1)T_m$ by integrating the time varying state transitioning probability expressions (i.e., Eq. (2.26)) [86] as

$$\begin{aligned} \widetilde{\mathbf{P}}_i &= \frac{1}{T_m} \int_{mT_m}^{(m+1)T_m} \mathbf{P}_i dt \\ &= \frac{1}{T_m} \mathbf{E}_i \left(\int_{mT_m}^{(m+1)T_m} e^{\mathbf{\Gamma}_i t} dt \right) \mathbf{F}_i \\ &= \frac{1}{T_m} [\widetilde{pr}_{i,0} \ \widetilde{pr}_{i,1} \ \cdots \ \widetilde{pr}_{i,N_i}]^T, \text{ for } i = p, \ s. \end{aligned} \tag{2.28}$$

In the above integration notation, we have used the fact that the integral of a matrix is the integral of each element of the matrix. The elements $\widetilde{pr}_{i,k}$ of the vector $\widetilde{\mathbf{P}}_i$ denote the probabilities of transitioning to state k at instant $(m+1)T_m$, for $i = p, \ s$. Based on the state transitioning probabilities, we can now forecast the number of primary and secondary users. For ease in presentation, the forecasting method for spectrum usage of SU is described first.

2.3.1 Forecasting Spectrum Usage of SU

Forecasting spectrum usage by other secondary users is critical for the following reasons. Each SU can now determine if a particular channel is overcrowded with secondary users. If it is, the channel may be avoided as it may degrade the QoS. Additionally, by forecasting the number of secondary users in a channel, each SU can also determine how much power it needs to transmit without violating spectral emission limits (while maintaining its QoS).

In this dissertation, we propose an upper bound forecasting based on state transitioning probability matrix of SU similar to the approach taken for forecasting the number of flows in

Internet traffic [86]. We discuss the case when the number of PU is 0 and we are interested in forecasting the number of secondary users at the next instant. The optimal estimate of the number of secondary users i.e., $\widehat{x}_s(m|m)$, at time instant m is used to forecast the number of secondary users at $(m+1)$ th instant. Based on estimated number of secondary users, $\widehat{x}_s(m|m)$ at time mT_m, state transitioning probability values is computed from Eq. (2.28) and then prediction for $(m+1)$ th instant is done. The predicted state of SU for $(m+1)$ th instant at time mT_m corresponds to [86],

$$\widetilde{x}_s(m) = \min_{x_s \in [\widehat{x}_s(m|m),\ N_s]} x_s \text{ s.t. } \widetilde{pr}_{s,k} < \beta. \tag{2.29}$$

Here, β is pre-set probability value. Equation (2.29) can be understood with the help of an example. Suppose, $\widehat{x}_s(m|m)$ is obtained as 3 and N_s is 8. Based on the value of $\widehat{x}_s(m|m)$, a state transition probability vector i.e., $\widetilde{\mathbf{P}}_s = \frac{1}{T_m}[\widetilde{pr}_{s,0}\ \widetilde{pr}_{s,1}\ \cdots\ \widetilde{pr}_{s,8}]^T$ is obtained, which corresponds to the possible states of SU with number of users $[0\ 1\ \cdots\ 8]^T$, respectively. By observing this state transition probability vector, one can determine multiple states for which $\widetilde{pr}_{s,k} < \beta$. All states with number of users greater than 5 might satisfy this condition. This effectively suggests that the probability of $x_s(m+1)$ being 5 or more is going to be negligible. Therefore, the upper bound for the $x_s(m+1)$ should be 5. In general, the chosen state is the state with minimum number of user satisfying Eq. (2.29). As a result, Eq. (2.29) serves as a good upper bound for the number of secondary users at time $(m+1)$ based on measurements up to time m.

2.3.2 Forecasting Spectrum Usage of PU

We propose two ways to forecast the presence or absence of a primary user. The first method is Kalman filter (KF) based prediction. In this method, the number of primary user obtained from the prediction stage (Eq. (2.20)) is used as the forecasted number of primary user. For example, from the estimated value at m th instant, $\widehat{x}_p(m|m)$, the predicted value

for $(m+1)$ th instant at m th instant is taken as,

$$\begin{aligned} \widetilde{x}_p(m) &= \widehat{x}_p(m+1|m) \\ &= A_p \widehat{x}_p(m|m) + B_p. \end{aligned} \tag{2.30}$$

The second method is based on state transitioning probabilities as given in Eq. (2.28). In this method, state transition probabilities from the present estimated state $\widehat{x}_p(m|m)$ is computed in the same way as it is described for SU above and state with higher state transition probability is taken as predicted state $\widetilde{x}_p(m)$. For the proposed system model in this chapter, Markov chain of PU shows that it follows two states - state 0 (number of PU is 0) and state 1 (number of PU is 1). Suppose, $\widehat{x}_p(m|m)$ is obtained as 1 and based on this, a state transition probability vector i.e., $\widetilde{\mathbf{P}}_p = \frac{1}{T_m}[\widetilde{pr}_{p,0}\ \widetilde{pr}_{p,1}]^T$ is computed. The predicted state of PU for $(m+1)$ th instant at time mT_m is proposed as,

$$\widetilde{x}_p(m) = \begin{cases} 1, & \text{if } \widetilde{pr}_{p,1} > \widetilde{pr}_{p,0}, \\ 0, & \text{otherwise} \end{cases} . \tag{2.31}$$

2.4 Experimental Results

In this section, we illustrate the potency of the proposed techniques for estimating and forecasting spectrum usage in a cognitive radio network through simulation. We show the performance of proposed methods on two different sets of data - (1) simulated data from a CRN; (2) real time measurement data from 2.4 GHz ISM band.

2.4.1 Simulated CRN

We consider a CRN where PU follows an ON-OFF Markov process with arrival rate (λ_p) and departure rate (μ_p), 0.000625 sec^{-1} and 0.00125 sec^{-1}, respectively. For primary user, λ_p is set smaller compared to μ_p to reflect the assumption that the primary user arrives less frequently but once it comes, it stays longer. The above choices for λ_p and μ_p reflect a 32% use of the channel by the PU on average. We assume that a SU is in "active phase," i.e., SU knows all required traffic parameters. In our set up, secondary users follow Markov process

with 11 states. The arrival rate (λ_s) and departure rate (μ_s) are taken as, 0.005 sec^{-1} and 0.005 sec^{-1}, respectively. The state noise variances, σ_p^2 and σ_s^2 are considered as 0.2 and 1, respectively. These parameters are summarized in Table 2.1.

Table 2.1: *Statistics of primary and secondary users.*

	Primary User	Secondary User
Maximum Number	$N_p = 1$	$N_s = 10$
Power level increase (dB)	C_p=30	C_s =2
State Noise Variance	σ_p^2 =0.2	σ_s^2 =1
Arrival Rate (sec^{-1})	λ_p =0.000625	λ_s =0.005
Departure rate (sec^{-1})	μ_p =0.00125	μ_s=0.005

The plots of evolution of primary and secondary users, $x_p(m)$ and $x_s(m)$, respectively with time are shown in Figs. 2.3(a) and 2.3(b). The number of measurement instants is 1001. The measurement interval, T_m is 10 sec. At the terminal of a SU, attempting to use this channel, the received power, $y(m)$ is shown in Fig. 2.3(c). Background noise level D and measurement noise variance, σ_v^2 are assumed as -135 dBm and 3, respectively. From Fig. 2.3(c), it is evident that the arrival of a primary user results in sudden increase in received power level as expected.

From $y(m)$, we first estimate the number of primary and secondary users, $\widehat{x}_p(m|m)$ and $\widehat{x}_s(m|m)$, respectively, from Eqs. (2.19)-(2.24) and then use these estimates for forecasting. For estimation, the Kalman filter initialization parameters are set as

$$\widehat{\mathbf{x}}(0|0) = \begin{bmatrix} \frac{B_p}{1-A_p} \\ \frac{B_s}{1-A_s} \end{bmatrix}$$

and

$$\mathbf{M}(0|0) = \begin{bmatrix} \frac{\sigma_p^2}{1-A_p^2} & 0 \\ 0 & \frac{\sigma_s^2}{1-A_s^2} \end{bmatrix}.$$

After estimation, prediction for both the number of primary and secondary users are done.

The prediction performances for the activity of PU are shown in Fig. 2.4. Figure 2.4(a) shows the true activity of PU. In Fig. 2.4(b), performance of KF based prediction (using

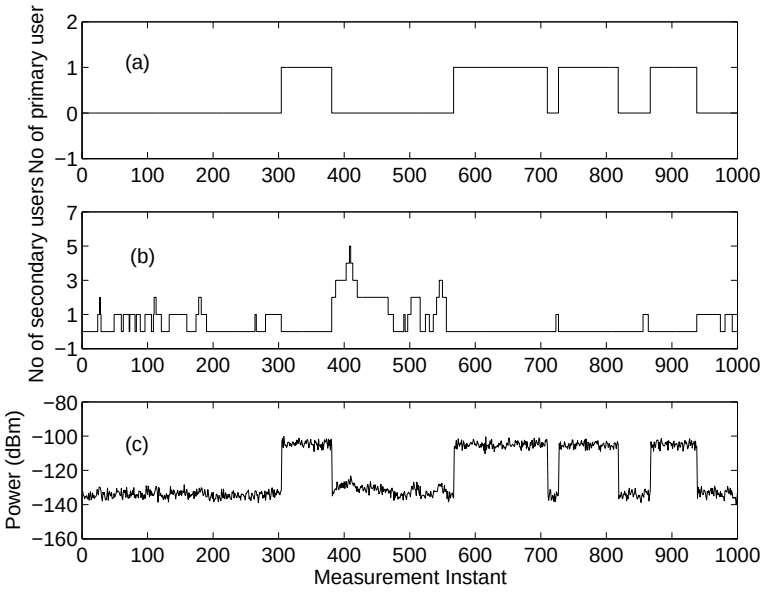

Figure 2.3: *Evolution of primary and secondary users, $x_p(m)$ and $x_s(m)$ along with power level variation, $y(m)$ with time.*

Eq. (2.30)) for the activity of PU is shown. This predictor solely depends on estimated value at the present instant and choices of A_p and B_p. In this set up, since A_p is 0.9814 and B_p is small, the predicted value for the next instant is equal to the current estimate. Therefore, with a single time instant lag, the KF predictor follows presence or absence of PU correctly. The state transitioning probability prediction method performs identical to the KF predictor as shown in Fig. 2.4(c). This method computes the probabilities $pr_{p,k}(t)$ of transitioning to all possible states from the current state (Eqs. (2.26)-(2.27)). But it is very easy to prove the probabilities $pr_{p,k}(t)$, $t \in (mT_m, (m+1)T_m]$, as defined in Eq. (2.1), are independent of m because arrivals and departures time follow exponential distribution (which is memory less). This simplification reduces the computation effort and time. The calculation of $\mathbf{F}_p$ and $\widetilde{\mathbf{P}}_p$ (in Eqs. (2.26) and (2.27), respectively) can be performed off-line for all possible states. Then the forecast process only involves this table-lookup to determine the next instant activity at each instant from a current instant activity based on Eq. (2.31).

As shown in the block diagram in Fig. 2.2, if the predicted value for PU, $\widetilde{x}_p(m)$ is 1, then

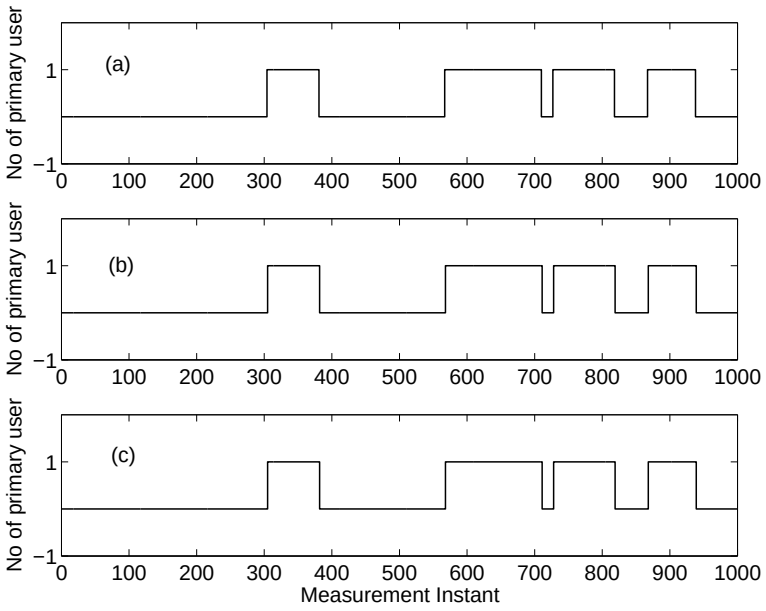

Figure 2.4: *Performance of prediction methods for PU; (a) True PU activity (absence or presence), (b) Prediction of activity by Method 1 and (c) Prediction of activity by Method 2.*

any SU if present, ceases to transmit. In order to reflect this in our simulation, $\widehat{x}_s(m|m)$ is reset to 0 and no prediction is done when $\widetilde{x}_p(m) = 1$. But if $\widetilde{x}_p(m)$ is computed as 0, prediction for SU is done based on current instant state estimate, $\widehat{x}_s(m|m)$ using the proposed method in Sec. 2.3. To predict the number of secondary users, the probabilities $pr_{s,k}(t)$ of transitioning to all possible states from the current state (Eqs. (2.26)-(2.27)) can be computed off-line for all possible states. Once again, these probabilities are independent of m. The forecast process only involves this table-lookup to determine the next state at each instant from the current state estimate based on Eq. (2.29). β is fixed at 0.1 for this simulation. This value of β indicates that the system has less than 10% chance to exceed the predicted state. Figure 2.5 shows the predicted number of secondary users, $\widetilde{x}_s(m)$ with true number of secondary users, $x_s(m)$. For clarity, only 300 to 600 measurement instants are shown in this figure. As expected, $\widetilde{x}_s(m)$ serves as a good upper bound predictor for the number of secondary users. For the same measurement window, the variation of predicted power level acting as an upper bound to true power level is shown in Fig. 2.6.

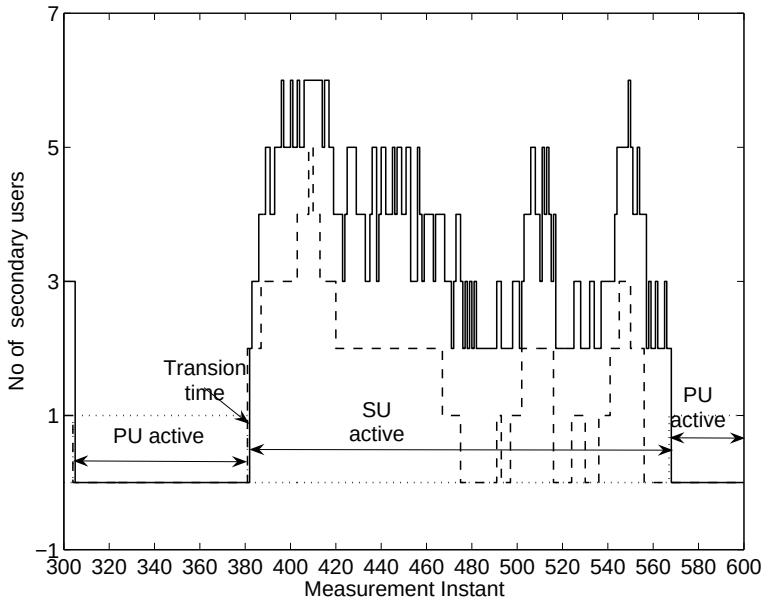

Figure 2.5: *Performance of the predictor for SU; where* $(\cdots)$, $(--)$ *and* $(-)$ *indicate true activity of primary user, true number of secondary users and predicted upper bound number of secondary users, respectively;* $\beta = 0.1$.

The performance of upper bound predictor for SU is affected by the preset value of β. For small values of β, the upper bound predictor performs satisfactorily. As β increases beyond a certain value, the upper bound begins to fail for a few time instants. As an example, Fig. 2.7 shows the performance of the predictor for $\beta = 0.20$. From this figure, it is evident that as β increases, the quality of upper bound predictor degrades.

The proposed estimation and forecasting process depend on traffic parameters λ_p, μ_p, λ_s and μ_s. The traffic parameters for PU are easier to find compared to SU as it follows ON-OFF traffic characteristics. Advances in sensing techniques [13, 18], [87] enable effective PU detection which in turn can be used to estimate λ_p and μ_p.

Estimating traffic parameters for SU is more involved. One approach is to use time series based Yule-Walker estimation as illustrated in next subsection. Therefore, it is important to evaluate the sensitivity of our proposed method to error in λ_s and μ_s estimates. In our simulation set up, the sensitivity performance of the predictor for SU is evaluated by using

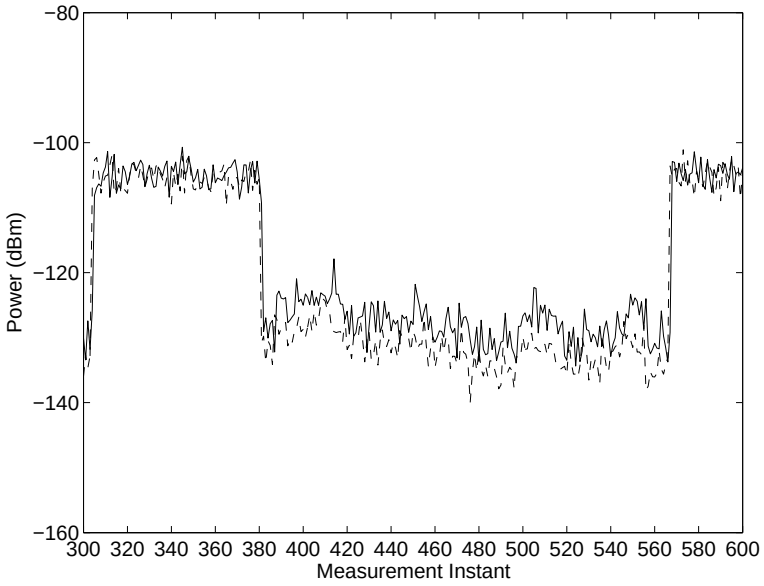

Figure 2.6: *Variation of predicted power level with true power level; where* $(--)$ *and* $(-)$ *indicate true and predicted power levels.*

erroneous values of λ_s and μ_s in Kalman filter estimation other than those used to generate traffic. Figure 2.8 shows the predictor performance for SU with true number of SU where value of λ_s is estimated as 25% lesser than its true value. From this figure, we observe that our proposed predictor continues to serve as a good upper bound for the number of SU. Figure 2.9 shows the predictor performance where values of λ_s and μ_s are overestimated by the same amount but performance of the predictor is still satisfactory. This shows that the predictor is relatively insensitive to inaccurate estimate of the values of λ_s and μ_s up to a certain limit.

2.4.2 Measured Data Analysis

In this set up, we show the effectiveness of proposed techniques on power level measurements from the ISM band. Although this band is not allocated for CR operations, we use the measurements to illustrate how the proposed methods can be practically implemented. The ISM band power level measurements are taken in Chicago by Shared Spectrum Company

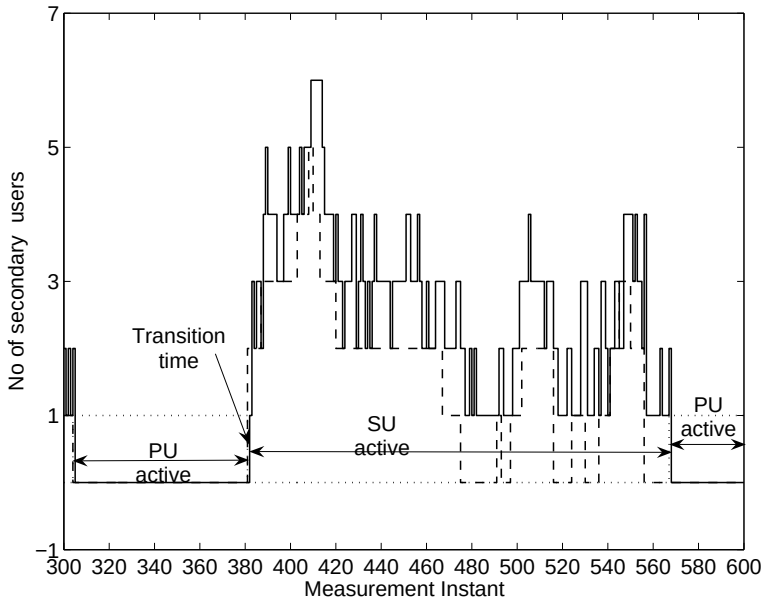

Figure 2.7: *Performance of the predictor for SU; where* $(\cdots)$, $(--)$ *and* $(-)$ *indicate true activity of primary user, true number of secondary users and predicted upper bound number of secondary users, respectively;* $\beta = 0.20$.

[88].

The ISM band ranges from 2.4 to 2.4835 GHz and is divided into 11 channels with carrier frequencies 2412, 2417, 2422, 2427, 2432, 2437, 2442, 2447, 2452, 2457, and 2462 MHz, respectively. This band is an unlicensed band. Since no primary user has exclusive access to these channels, we assume that the measured power is due to secondary users only. The received power measured for 2412 MHz frequency is shown in Fig. 2.10. The measurement interval, T_m is 2 sec. The data length is 1512. Any SU, interested to use one of the channels, needs to go through the "learning and modeling phase" first to find traffic parameters λ_s and μ_s for the corresponding channel. In this phase, SU can use time series approach to get these parameters. Assume that the models described in Eqs. (2.15)-(2.16) (with no PU) hold, the received signal at the SU terminal can be written as

$$y(m) = A_s y(m-1) + C_s B_s + (1 - A_s)D + [C_s w_s(m) + v(m) - A_s v(m-1)]. \quad (2.32)$$

Equation (2.32) indicates that the spectrum usage process is an auto regressive (AR)

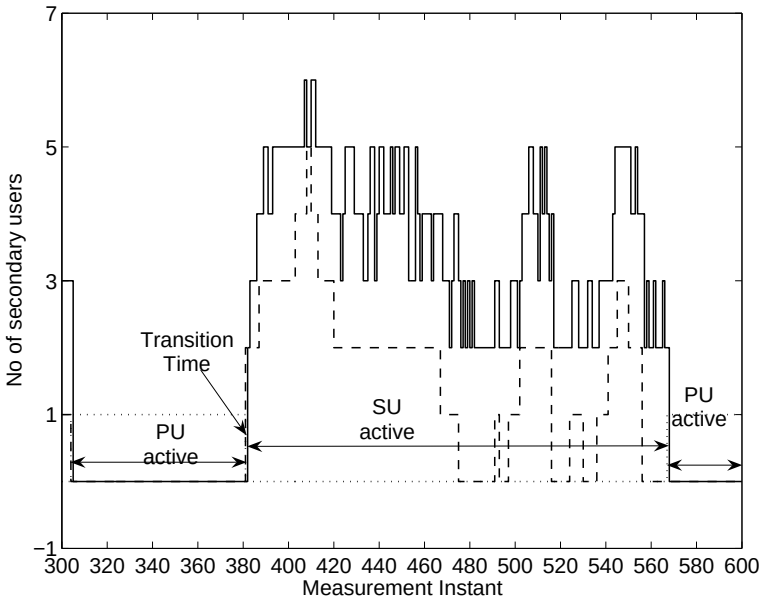

Figure 2.8: *Sensitivity of the predictor for SU (λ_s is under estimated by 25%); where $(\cdots)$, $(--)$ and $(-)$ indicate true PU activity, true number of secondary users and predicted upper bound number of secondary users, respectively.*

process of order 1. Any standard technique can be used to find parameters of AR process. In our case, we employ Yule-Walker estimation to find the model parameters. Based on some intuitive assumptions (e.g., $\lambda_s \leq \mu_s$, $\sigma_s^2 \leq \sigma_v^2$) along with Yule-Walker estimates; mean of the process, knowledge of D, N_s, C_s (considering N_s and C_s as regulatory constraints), we can solve for all traffic parameters A_s, B_s, λ_s, μ_s, σ_s^2 and σ_v^2. Specifically, in our simulation, we solve for all parameters assuming $\lambda_s = \mu_s$, $D = -130$ dB; and N_s and C_s to 20 and 2 dB, respectively. After finding model parameters, fitness of model is checked by finding residual partial auto correlation function (PACF) values. In general, if 95% of the residual PACF values are within the acceptable bound, the AR model parameters are valid. An example plot of the PACF of the residuals (the difference between the true and the model generated power level data) is shown in Fig. 2.11, where $(\cdots)$ and $(-)$ indicate the acceptable PACF bound and residual PACF values, respectively. If PACF values are satisfactory, model parameters obtained in "learning and modeling phase" is accepted and used in the "active

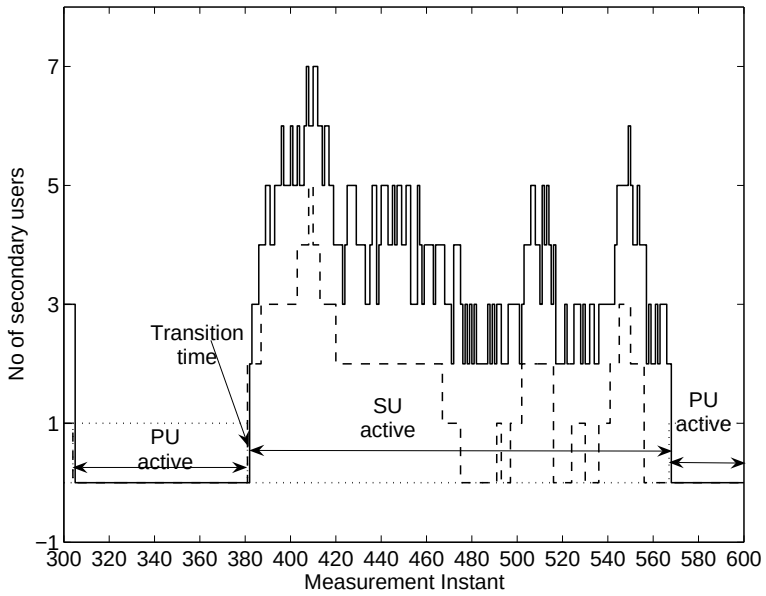

Figure 2.9: *Sensitivity of the predictor for SU (both λ_s and μ_s are over estimated by 25%); where $(\cdots)$, $(--)$ and $(-)$ indicate true PU activity, true number of secondary users and predicted upper bound number of secondary users, respectively.*

phase."

In the "active phase," SU attempts to use the channel opportunistically using the traffic parameters obtained in "learning and modeling phase." It is assumed that accepted model parameters from "learning and modeling phase" do not change during our simulation time. Once again, Eqs. (2.19)-(2.24) are used to estimate the number of secondary users and based on this estimation, Eqs. (2.26)-(2.29) are used to predict the number of secondary users. Figures 2.12 and 2.13 show the predicted number of users with the true one for two carrier frequencies, 2412 and 2437 MHz, respectively. It shows that the proposed forecasting technique also provides an good upper bound number of users for practically measured received power at user terminal.

In the following section, we generalize the SU traffic model. We consider an Erlangian process model for SU traffic. The model captures bulk arrival and departure scenarios in a frequency band that is open for opportunistic use by multiple secondary users. We develop a Kalman filter based estimation and forecasting strategy for the number of secondary users.

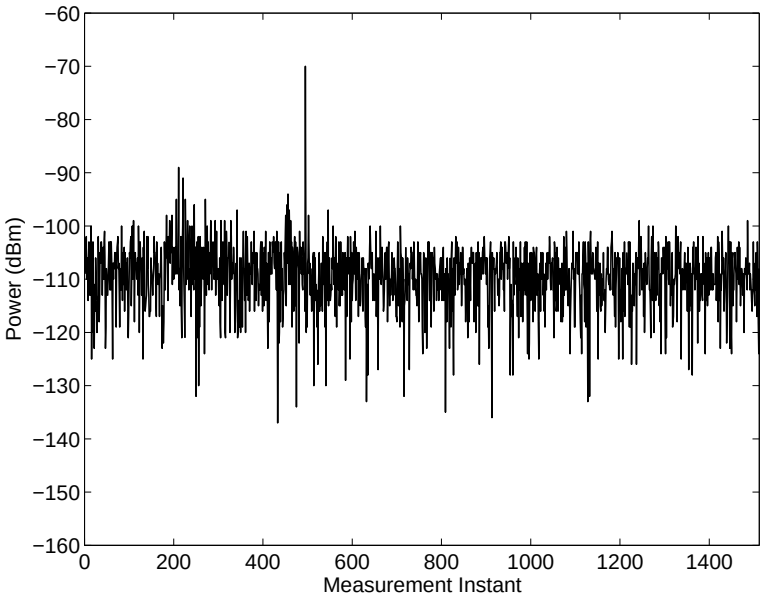

Figure 2.10: *Received power level for carrier frequency* 2412 *MHz.*

Using simulated data, we demonstrate that the proposed approach provides a robust upper bound prediction on the number of secondary users.

2.5 SU Generalized Traffic Model

2.5.1 System Model

As in Sec. 2.1, we assume that each channel in a CRN can be used by either a PU or one or more secondary users (once it is determined that the channel will not be used by a PU). We also assume that the PU follows Poisson arrival process with arrival and departure rates, λ_p and μ_p, respectively. The maximum number of PU is N_p. For ease in presentation, N_p is assumed to be equal to 1. In other words, PU follows a two-state ON-OFF Markov process. Now, unlike in Sec. 2.1, we assume that secondary users can arrive and leave the network as a bulk or group. This implies that non-nearest neighbor transitions are allowed. The maximum number of the secondary users is N_s and the acceptable maximum number of the secondary users in a bulk is N_b ($N_b < N_s$). Each of the secondary users in a bulk irrespective

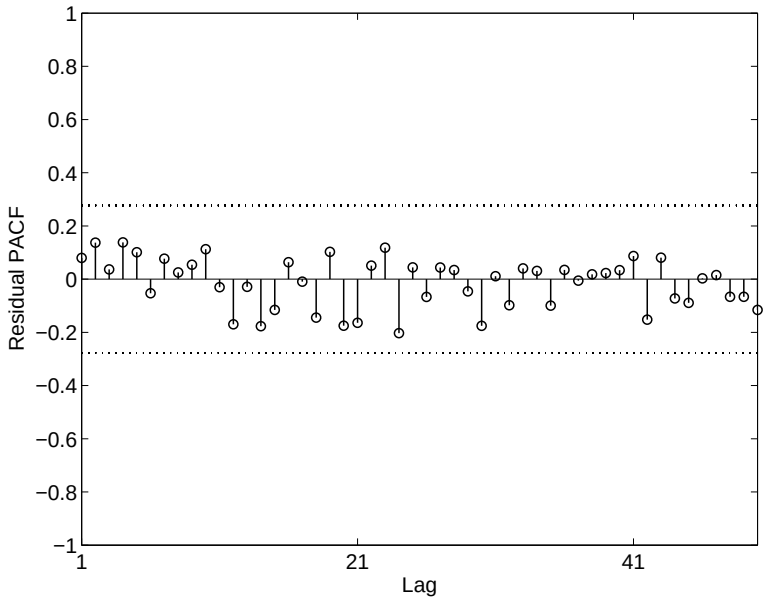

Figure 2.11: *Plot of residual PACF values for carrier frequency* 2412 *MHz; Model parameters correspond to this plot are* $\lambda_s = \mu_s = 0.4544$ sec^{-1}.

of bulk size has exponential inter-arrival and service time distributions with rates λ_s and μ_s, respectively. Based on these assumptions, the state-transition-rate diagram for k-th state of secondary users is shown in Fig. 2.14.

Our objective is to develop a prediction strategy for secondary users. To accomplish this, we need a dynamic model of secondary users activity. From the state-transition-rate diagram in Fig. 2.14 and concepts from queueing theory [85], the differential equations for the state probabilities $pr_{s,k}(t)$ is evaluated for secondary users. The state probability is defined as

$$pr_{s,k}(t) \triangleq \text{prob}\{x_s(t) = k\}, \tag{2.33}$$

where, $x_s(t)$ is the number of SUs at time t and k indicates that number. In general, the differential equations for the state probabilities for a system shown in Fig. 2.14 correspond

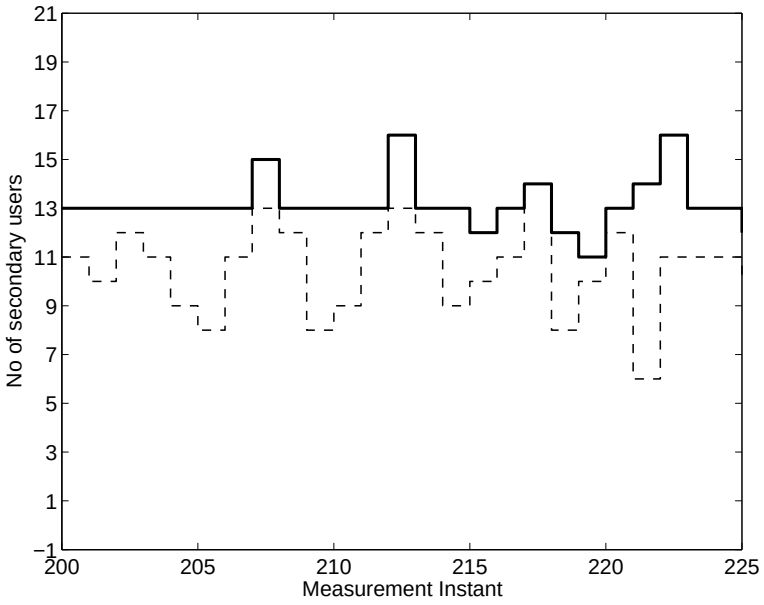

Figure 2.12: *Performance of upper bound predictor for carrier frequency* 2412 *MHz; where* $(--)$ *and* $(-)$ *indicate true and predicted upper bound number of users, respectively.*

to

$$\frac{d}{dt}pr_{s,0}(t) = \mu_s \sum_{j=1}^{\min(N_b,\ N_s)} pr_{s,j}(t) j g_j - N_s \lambda_s pr_{s,0}(t) \sum_{j=1}^{\min(N_b,\ N_s)} g_j \tag{2.34}$$

.

.

.

$$\frac{d}{dt}pr_{s,k}(t) = \lambda_s \sum_{j=1}^{\min(N_b,\ k)} pr_{s,(k-j)}(t)(N_s - k + j) g_j + \mu_i \sum_{j=1}^{\min(N_b,\ (N_s-k))} pr_{s,(k+j)}(t)(k+j) g_j$$
$$-pr_{s,k}(t) \left(k\mu_s \sum_{j=1}^{\min(N_b,\ k)} g_j + (N_s - k)\lambda_s \sum_{j=1}^{\min(N_b,\ (N_s-k))} g_j \right),\ 1 \leq k\ <\ N_s, \tag{2.35}$$

.

.

.

$$\frac{d}{dt}pr_{s,N_s}(t) = \lambda_s \sum_{j=1}^{\min(N_b,\ N_s)} pr_{s,(N_s-j)}(t) j g_j - N_s \mu_i pr_{s,N_s}(t) \sum_{j=1}^{\min(N_b,\ N_s)} g_j. \tag{2.36}$$

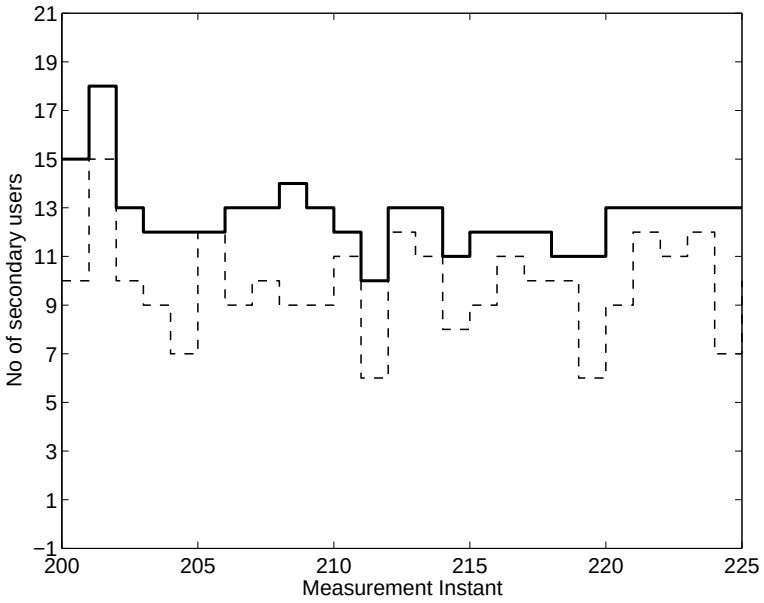

Figure 2.13: *Performance of upper bound predictor for carrier frequency* 2437 *MHz; where* $(--)$ *and* $(-)$ *indicate true and predicted upper bound number of users, respectively.*

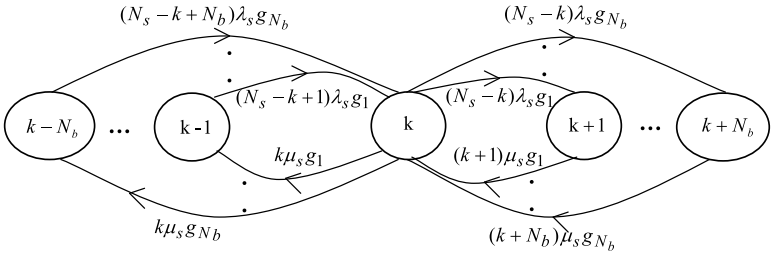

Figure 2.14: *State-transition-rate diagram of* k*-th state of secondary users.*

Here,

$$g_j \triangleq \text{Probability of [group size of users} = j\}]. \tag{2.37}$$

$E\{x_s(t)\}$ can be written as

$$E\{x_s(t)\} = \sum_{k=0}^{N_s} k pr_{s,k}(t). \tag{2.38}$$

Hence,

$$\frac{d}{dt}E\{x_s(t)\} = \sum_{k=0}^{N_s} k \frac{d}{dt} pr_{s,k}(t). \tag{2.39}$$

Let $\mathbf{L}_Q = [0\ 1\ 2 \cdots\ N_s]^T$. From Eqs. (2.34)-(2.36) and (2.39), we can write

$$\frac{d}{dt}E\{x_s(t)\} = \mathbf{L}_Q^T \dot{\mathbf{P}}_s = \mathbf{L}^T \mathbf{Q}_s \mathbf{P}_s, \tag{2.40}$$

where,

$$\mathbf{Q}_s = \begin{bmatrix} -N_s\lambda_s \sum_{j=1}^{N_b} g_j & \mu_s g_1 & \cdot & \cdot \\ N_s\lambda_s g_1 & -\left((N_s-1)\lambda_s \sum_{j=1}^{N_b} g_j + \mu_s g_1\right) & \cdot & \cdot \\ N_s\lambda_s g_2 & (N_s-1)\lambda_s g_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & -N_s\mu_s \sum_{j=1}^{N_b} g_j \end{bmatrix}$$

and

$$\mathbf{P}_s = \begin{bmatrix} pr_{s,0}(t) \\ pr_{s,1}(t) \\ \cdot \\ \cdot \\ \cdot \\ pr_{s,N_s}(t) \end{bmatrix}.$$

In Eq. (2.40), $(\cdot)^T$ indicates matrix or vector transpose operator. It is easy to show that $\mathbf{L}_Q^T \mathbf{Q}_s \mathbf{P}_s$ corresponds to $-(\lambda_s + \mu_s)E\{x_s(t)\} + f_s(t)$. Therefore,

$$\frac{d}{dt}E\{x_s(t)\} = -(\lambda_s + \mu_s)E\{x_s(t)\} + f_s(t), \tag{2.41}$$

where,

$$f_s(t) = \sum_{k=0}^{N_s} pr_{s,k}(t) \left[(\lambda_s + \mu_s)k + k\mu_s \sum_{j=1}^{\min(N_b,\ k)} (-jg_j) + (N_s - k)\lambda_s \sum_{j=1}^{\min(N_b,\ (N_s-k))} jg_j\right]. \tag{2.42}$$

We assume that measurements are performed at discrete time instants mT_m, $m = 1,\ 2,\ 3, \cdots$ for a given value T_m. Using the initial condition that the number of users at time $t = (m-1)T_m$ is $x_s(m-1)$, the solution of Eq. (2.41) is obtained as

$$\begin{aligned} E\left[x_s(m)|x_s(m-1)\right] &= e^{-T_m(\lambda_s+\mu_s)}x_s(m-1) \\ &\quad + e^{-T_m(\lambda_s+\mu_s)}\left[\int_{(m-1)T_m}^{mT_m} f_s(t)e^{(\lambda_s+\mu_s)(t-(m-1)T_m)}dt\right]. \end{aligned} \tag{2.43}$$

Therefore, it is possible to express the number of secondary users at time mT_m in terms of the number of secondary users at time $(m-1)T_m$ as

$$x_s(m) = A_s x_s(m-1) + B_s(m), \tag{2.44}$$

where,

$$A_s = e^{-T_m(\lambda_s+\mu_s)} \tag{2.45}$$

and

$$B_s(m) \;=\; e^{-T_m(\lambda_s+\mu_s)} \left[\int_{(m-1)T_m}^{mT_m} f_s(t) e^{(\lambda_s+\mu_s)(t-(m-1)T_m)} dt \right]. \tag{2.46}$$

Equation (2.44) establishes the relationship between the number of users at two successive measurement instants and in the most general case corresponds to

$$x_s(m) = A_s x_s(m-1) + B_s(m) u_s(m) + w_s(m). \tag{2.47}$$

Equation (2.47) can be considered the state equation where, $x_s(m)$ represents the number of secondary users using the spectrum at the measurement instant m. The parameter $B_s(m)$ relates the optional control input, $u_s(m)$ to state. Equation (2.44) suggests that $u_s(m)$ is equal to 1. $w_s(m)$ is the process noise and assumed to be zero mean Gaussian noise with variance σ_s^2. The parameter A_s relates the state at previous and current measurement instants, in the absence of either a driving function or process noise. A_s is assumed to be constant over the analysis or varies very slowly.

The received power at a secondary user terminal during the measurement instant m consists of relative power level increments caused by secondary users and in the most general case corresponds to,

$$y(m) = C_s x_s(m) + D + v(m), \tag{2.48}$$

where, $y(m)$ is received power in dBm; C_s represent the relative increase in power level (in dB) due to the presence of one secondary user; D represents the background thermal

noise and $v(m)$ denotes the measurement noise which may arise due to miscalculation, misalignment of timings and is assumed to be zero mean Gaussian noise with variance σ_v^2. $y(m)$ is the only measurable variable in the system.

It is important to note that unlike in SU Poissonian traffic model case (Eq. (2.15)), the parameter $B_s(m)$ relating optional control input, $u_s(m)$ to the state $x_s(m)$ is not a constant and is time-dependent. The computation of time-dependent $B_s(m)$ is discussed in Sec. 2.5.2.

2.5.2 Estimation of Spectrum Usage

In this section, we develop a Kalman filter based state estimation technique based on the model from Sec. 2.5.1. An opportunistic SU or a central controller can use this technique to estimate the number of secondary users (once the traffic parameters are determined in learning phase).

The state estimation based on Kalman filter is summarized below:

State Equation: The state equation is

$$x_s(m) = A_s x_s(m-1) + B_s(m) + w_s(m), \tag{2.49}$$

where, $w_s(m)$ is a white Gaussian noise with mean 0 and variance, σ_s^2.

Measurement Equation:

$$y(m) = C_s x_s(m) + D + v(m), \tag{2.50}$$

where, $v(m)$ is a white Gaussian noise with mean 0 and variance, σ_v^2. Based on Eqs. (2.49) and (2.50), the Kalman filtering steps are given below:

Step 1: Initialization

$$\widehat{x}_s(0|0) = E\{x_s(0)\} \tag{2.51}$$

$$M_s(0|0) = \sigma_s^2(0) \tag{2.52}$$

Step 2: Prediction

$$\widehat{x}_s(m|m-1) = A_s\widehat{x}_s(m-1|m-1) + B_s(m|m-1), \tag{2.53}$$

$$M_s(m|m-1) = A_s M_s(m-1|m-1)A_s + \sigma_s^2, \forall\ m \tag{2.54}$$

Step 3: Kalman gain calculation

$$k_s(m) = M_s(m|m-1)C_s\left(C_s M_s(m|m-1)C_s + \sigma_v^2\right)^{-1}, \forall\ m \tag{2.55}$$

Step 4: Correction

$$\widehat{x}_s(m|m) = \widehat{x}_s(m|m-1) + k_s(m)\left(y(m) - C_s\widehat{x}_s(m|m-1) - D\right), \tag{2.56}$$

$$M_s(m|m) = \{1 - k_s(m)C_s\}M_s(m|m-1), \forall\ m \tag{2.57}$$

Step 5: Computation of $B_s(m)$, $\forall\ m$

From Eq. (2.46), we observe that $B_s(m)$ is $e^{-T_m(\lambda_s+\mu_s)}$ multiplied by the integration of $f_s(t)e^{(\lambda_s+\mu_s)(t-(m-1)T_m)}$ between $(m-1)T_m$ to mT_m. $f_s(t)$ in turn can be written as $\mathbf{a}^T\mathbf{P}_s$, where, $\mathbf{P}_s$ is as defined in Sec. 2.5.1 and $\mathbf{a}^T$ is equal to

$$\left[N_s\lambda_s\sum_{j=1}^{N_b} jg_j \cdots (\lambda_s+\mu_s)N_s + k\mu_s\sum_{j=1}^{N_b}(-jg_j)\right]. \tag{2.58}$$

The state probabilities $\mathbf{P}_s$ can be computed from the differential equation

$$\dot{\mathbf{P}}_s = \mathbf{Q}_s\mathbf{P}_s. \tag{2.59}$$

If the matrix $\mathbf{Q}_s$ has unique eigenvalues then the solution of Eq. (2.59) for $t \in ((m-1)T_m, mT_m]$ is given by

$$\mathbf{P}_s = \mathbf{E}_s e^{\mathbf{\Gamma}_s t}\mathbf{F}_s. \tag{2.60}$$

Here, $\mathbf{\Gamma}_s$ is a diagonal matrix with eigenvalues of $\mathbf{Q}_s$; $\mathbf{E}_s$ is the matrix of corresponding right eigenvectors, and $\mathbf{F}_s$ is a constant vector determined from the initial condition (*i.e*, $\widehat{x}_s(m-1|m-1)$) as

$$\mathbf{F}_s = (e^{\mathbf{\Gamma}_s(m-1)T_m})^{-1}\mathbf{E}_s^{-1}\mathbf{P}_{(m-1)T_m(s)}. \tag{2.61}$$

Here, $\mathbf{P}_{(m-1)T_m(s)}$ is a vector with all zeros except the $\widehat{x}_s(m-1|m-1))$ th element which is 1. It is very easy to show that at mth instant,

$$B_s(m) = e^{-T_m(\lambda_s+\mu_s)}\mathbf{a}^T\mathbf{I}_s, \tag{2.62}$$

where

$$\mathbf{I}_s = \mathbf{E}_s\left(\int_{(m-1)T_m}^{mT_m} e^{\mathbf{\Gamma}_{sb}t}dt\right)\mathbf{F}_s. \tag{2.63}$$

Here, $\mathbf{\Gamma}_{sb} = (\mathbf{\Gamma}_s + (\lambda_s + \mu_s))$ and $\mathbf{F}_s$ is computed as

$$\mathbf{F}_s = (e^{\mathbf{\Gamma}_{sb}(m-1)T_m})^{-1}\mathbf{E}_s^{-1}\mathbf{P}_{(m-1)T_m(s)}. \tag{2.64}$$

In Eq. (2.63), we have used the fact that the integral of a matrix is the integral of each element of the matrix.

2.5.3 Forecasting Spectrum Usage

In Sec. 2.3.1, we develop the forecasting tool for SU Poissonian traffic model assumption. For SU Erlangian traffic model assumption, the same tool can be used by an opportunistic SU for forecasting the number of secondary users at a future time instant. Here, we briefly review the forecasting tool.

The approach for forecasting is to determine the most probable state at the next time instant given that we have the current instant state estimate. To do this, we need to calculate the probability of transitioning to another state at time $(m+1)T_m$. The state transitioning probability values for the instant $(m+1)T_m$ are computed by integrating the time varying state transitioning probability expressions (i.e., Eq. (2.60)) as

$$\begin{aligned}
\widetilde{\mathbf{P}}_s &= \frac{1}{T_m}\int_{mT_m}^{(m+1)T_m}\mathbf{P}_s dt \\
&= \frac{1}{T_m}\mathbf{E}_s\left(\int_{mT}^{(m+1)T_m} e^{\mathbf{\Gamma}_s t}dt\right)\mathbf{F}_s \qquad (2.65)\\
&= \frac{1}{T_m}[\widetilde{pr}_{s,0}\ \widetilde{pr}_{s,1}\ \cdots\ \widetilde{pr}_{s,N_s}]^T. \qquad (2.66)
\end{aligned}$$

Here, $\mathbf{F}_s$ is a constant vector determined from the initial condition ($i.e.$, $\widehat{x}_s(m|m)$) as

$$\mathbf{F}_s = (e^{\mathbf{\Gamma}_s m T_m})^{-1} \mathbf{E}_s^{-1} \mathbf{P}_{mT_m(s)}. \tag{2.67}$$

Here, $\mathbf{P}_{mT_m(s)}$ is a vector with all zeros except the $\widehat{x}_s(m|m))$ th element which is 1. The elements $\widetilde{pr}_{s,k}$ of the vector $\widetilde{\mathbf{P}}_s$ denote the probabilities of transitioning to state k at instant $(m+1)T_m$. Based on estimated number of secondary users, $\widehat{x}_s(m|m)$ at time mT_m, state transitioning probability values are computed from Eq. (2.66) and then prediction for $(m+1)$ th instant is done. The predicted state of SU for $(m+1)$ th instant at time mT_m corresponds to

$$\widetilde{x}_s(m) = \min_{x_s \in [\widehat{x}_s(m|m),\ N_s]} x_s \text{ s.t. } \widetilde{pr}_{s,k} < \beta. \tag{2.68}$$

Here, β is a threshold similar to that defined in Sec. 2.3.1.

2.5.4 Experimental Results

We illustrate the performance of the proposed Kalman filter based estimate and upper bound predictor on simulated CRN data.

We consider a CRN during the time when PU is absent and SU starts to use channel opportunistically. As mentioned before, secondary users follow Erlangian process. The maximum number of secondary users, N_s is taken as 20. The arrival and departure rates, λ_s and μ_s of each of the secondary users are taken as, 0.0019 sec^{-1} and 0.0025 sec^{-1}, respectively. The state noise variance σ_s^2 is set to 1. N_b is considered as 4. The group probabilities g_1, g_2, g_3 and g_4 are set as 0.65, 0.20, 0.10 and 0.05, respectively. This choice of probabilities reflects a reasonable assumption that a group with 1 secondary user has the maximum probability and a group with N_b number secondary users has the minimum probability to arrive or to depart the network.

The evolution of secondary users, $x_s(m)$ with measurement instant are shown in Fig. 2.15(a). The number of measurement instants is 3001. The measurement interval, T_m is 10 sec. At the terminal of a SU, attempting to use this channel, the received power, $y(m)$ is

shown in Fig. 2.15(b). Background noise level D and measurement noise variance, σ_v^2 are assumed as -135 dBm and 3, respectively.

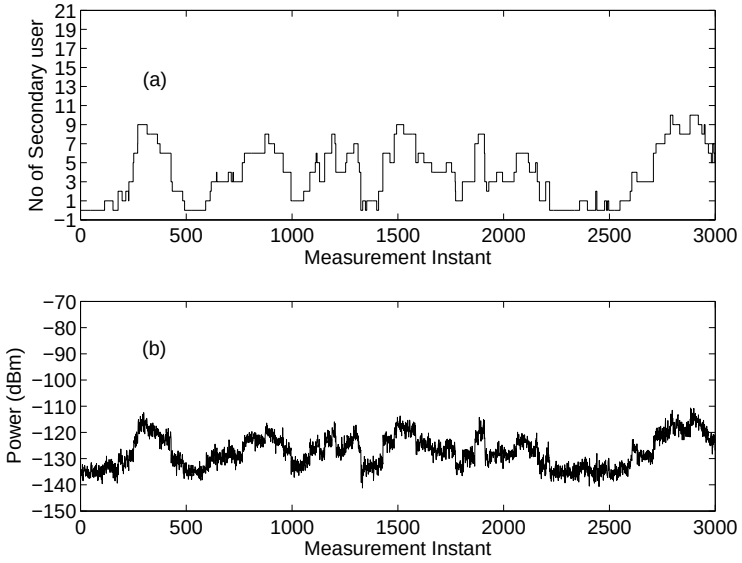

Figure 2.15: *Evolution of secondary users, $x_s(m)$ and power level variation, $y(m)$ with time.*

From $y(m)$, we first estimate the number of secondary users, $\widehat{x}_s(m|m)$ from Eqs. (2.51)-(2.57) and then use this estimate for forecasting. The Kalman filter initialization parameters are set as $\widehat{x}_s(0|0) = B_s(0)/(1 - A_s)$ and $M_s(0|0) = \sigma_s^2/(1 - A_s^2)$. $B_s(0)$ is evaluated (using Eq. (2.62)) assuming $x(-1) = 0$. After estimation, prediction for the number of secondary users is done.

As in Sec. 2.4, the forecast process only involves a table-lookup to determine the next state at each instant from the current state estimate based on Eq. (2.68). β is fixed at 0.006 for this simulation. This value of β indicates that the system has less than 0.6% chance to exceed the predicted state. Figure 2.16 shows the predicted upper bound number of secondary users, $\widetilde{x}_s(m)$ with true number of secondary users, $x_s(m)$. For clarity, only 1200 to 1500 measurement instants are shown in this figure. From Fig. 2.16, it is evident that the prediction tool provides a good upper bound for the number of secondary users.

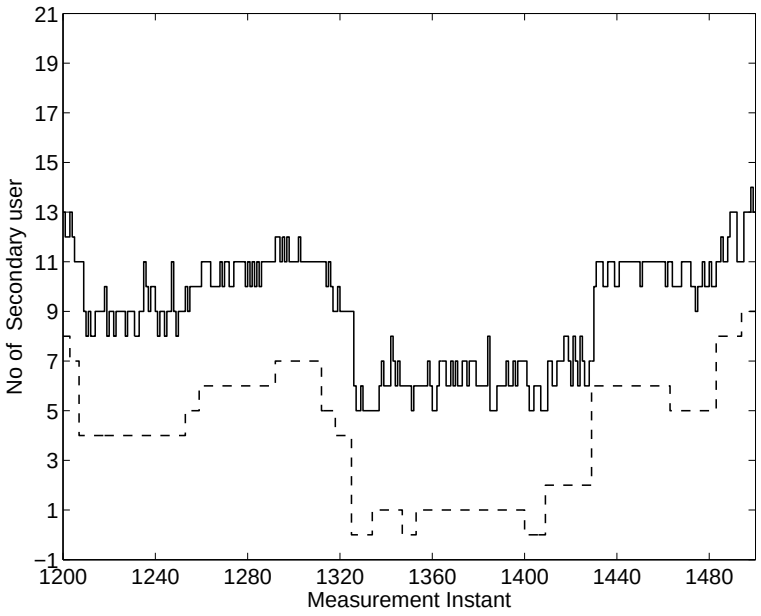

Figure 2.16: *Performance of the predictor; where* $(--)$ *and* $(-)$ *indicate the true and predicted upper bound number of secondary users, respectively;* $\beta = 0.006$.

In the analysis and simulation thus far, we assumed that secondary users have an accurate estimate of traffic parameters λ_s and μ_s. In practice, this may not be possible. Figures 2.17 and 2.18 show the robustness of the predictor with erroneous estimate of traffic parameters λ_s and μ_s. In Fig. 2.17, the parameters λ_s and μ_s are assumed to have been overestimated by 10%. These over estimated λ_s and μ_s values are used in Kalman filter estimator and then in predictor. In Fig. 2.18, the parameter estimates are assumed to be 10% smaller than their true values. In each case, the proposed upper bound predictor still performs satisfactorily. The predictor shows relatively low sensitivity to erroneous estimate of traffic parameters.

2.6 Summary

In this chapter, we present an integrated spectrum usage model and forecasting strategy for both primary and secondary users in CRNs. Firstly, assuming that primary and secondary users follow a continuous time Markov chain model, we develop a Kalman filter approach to

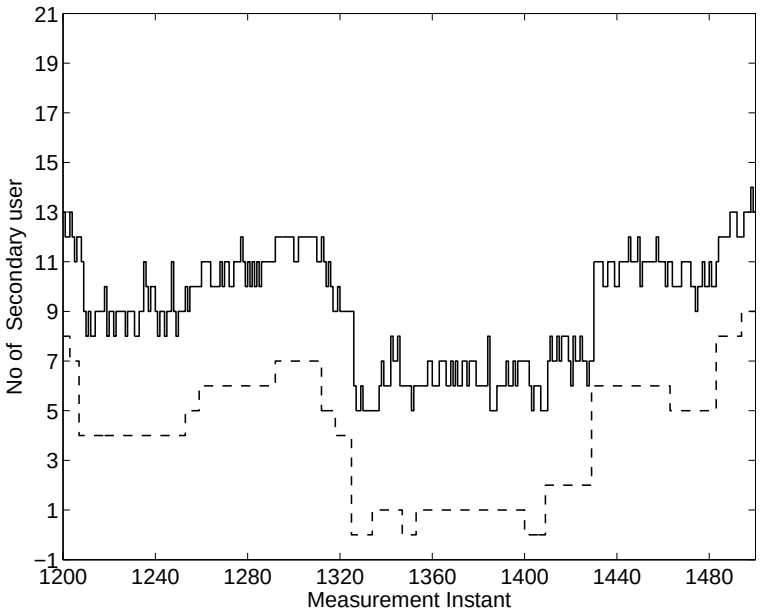

Figure 2.17: *Sensitivity of the predictor (both λ_s and μ_s are overestimated by 10%); where $(--)$ and $(-)$ indicate the true and predicted upper bound number of secondary users, respectively; $\beta = 0.006$.*

estimate the number of users solely based on power level measurements at the radio terminal of an opportunistic user. These estimates in turn are used to determine upper bound predictors for a future time instant. In a sense, this forecasting strategy provides secondary users not only the ability to predict if a spectrum band will be used by a primary user but also determine if a band will be overcrowded with secondary users. The effectiveness of the proposed architecture is demonstrated using experiments on both practically measured as well as simulated data. Secondly, we assume that secondary user traffic in a CRN is governed by Erlangian process, i.e., the traffic model incorporates bulk arrival or bulk departure scenarios. Accordingly, we develop a Kalman filter based state estimation technique to estimate the number of secondary users based on power level measurements. This estimate is used for upper bound prediction of the number of secondary users at future time instant. Simulation results show that the proposed forecasting strategy provides robust upper bound predictor for the number of secondary users.

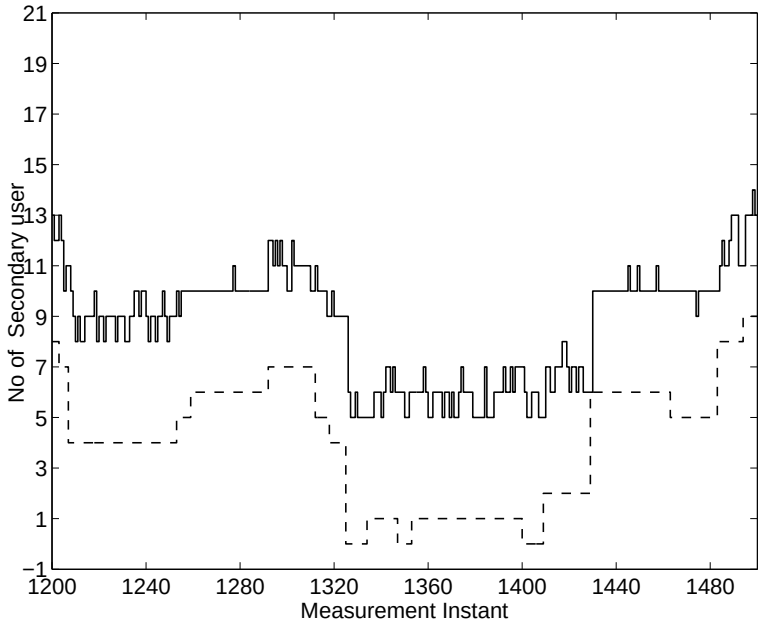

Figure 2.18: *Sensitivity of the predictor (both λ_s and μ_s are underestimated by 10%); where $(--)$ and $(-)$ indicate the true and predicted upper bound number of secondary users, respectively; $\beta = 0.006$.*

Chapter 3

Two-Stage Resource Allocation

In chapter 2, we proposed an integrated modeling and forecasting strategy that can be used to predict the number of secondary users accessing a spectral band of interest. In this chapter, we consider a multi-channel CRN where multiple secondary users share a single channel and multiple channels are simultaneously used by a single secondary user. In each channel, the number of users is determined through proposed modeling and forecasting strategy in chapter 2. Here, we are interested to find the transmission parameters for the users in each of the channels. Specifically, we propose a resource allocation framework for transmit power and rate for secondary users. Our objective is to determine the optimal distribution of power and rate that a secondary user has to employ across the channels that it uses in order to (1) minimize total power consumption; (2) maximize rate, and (3) maintain QoS. A two-stage approach is presented in this chapter. In the first stage, optimal choice for transmit power is determined for all SUs subject to maintaining a given SINR in each channel used. Firstly, stage 1 optimization problem is solved in a centralized manner and then we employ dual decomposition theory to derive three different distributed solutions. Using power/SINR result in stage 1, the optimal distribution of bits/channel is determined in the second stage. Specifically, we formulate the rate distribution problem as a maximum flow problem in graph theory. We also develop a heuristic approach to determine the rate distribution. It is important to note that unlike prior efforts, we have transformed the BER constraint into a convex constraint in order to ensure optimality of our resulting

solution.

The rest of the chapter is organized as follows. In Sec. 3.1, along with the system model, we state all of our assumptions and notations used in the rest of the chapter. Section 3.2 describes the proposed two-stage resource allocation framework along with dual distributed versions in detail. Numerical results are presented in Sec. 3.3. Finally, Sec. 3.4 summarizes the chapter.

3.1 System Model

Consider a CRN with a total of M secondary users and L free channels available for opportunistic use (determined after spectrum sensing) by multiple SUs. We assume that each channel can be used simultaneously by multiple secondary users via some form of non-orthogonal multiple access scheme, and a single secondary user can use several channels at the same time to meet their rate requirements. Our interest in this chapter is to maintain QoS for these competing SUs via effective resource allocation. We consider SINR or BER and minimum rate requirement as measures to indicate QoS. In order to enable mathematical tractability of the resource allocation framework, we invoke the following assumptions: (1) We assume that we have a central cognitive network controller that will perform the resource allocation and has access to all SUs channel and interference parameters; (2) Every active SU radio has an upper limit on power and rate (bits/channel use) at which it can transmit; (3) All SUs employ M-ary quadrature amplitude modulation (QAM) scheme with an adaptable modulation order M; (4) Simple path loss model for channel has been assumed; (5) Each channel has a maximum rate (bits/channel use) that it can support, and (6) Each user has a minimum rate and SINR or BER constraint that needs to be maintained.

Additionally, we enforce an interference temperature threshold to protect possible primary user transmission on any channel. Interference temperature is defined as the total RF power measured at a receiving antenna per unit bandwidth of primary user. Recently, Federal communications committee has removed the interference temperature limit as the

quantifying metric for characterizing interference [89]. However, we feel that there is value to having this limit in place especially during the transition stages. That is the interference threshold will ensure that when a PU enters a channel used by SUs, there is a certain limit to the interference that it experiences before the SUs vacate the channel. Finally, we assume that at each time instant for resource allocation, we have an estimate of the number of secondary users that may be demanding access to each of the channels denoted by $\widetilde{N}_s(k)$ where $k = 1,\ 2, \cdots,\ L$. This information can be obtained using traffic models along with a Kalman filter based predictor developed in chapter 2.

Under this system model, we propose a resource allocation framework to find transmit power and rate. It is important to note that we use the terms rate and bits/channel use interchangeably throughout the chapter. One can also visualize the bits/channel use measure to indicate the modulation order employed by the SU in a channel. Figure 3.1 summarizes the overall resource allocation architecture with information on all the constraints considered. Table 3.1 defines most of the relevant terms used throughout the chapter.

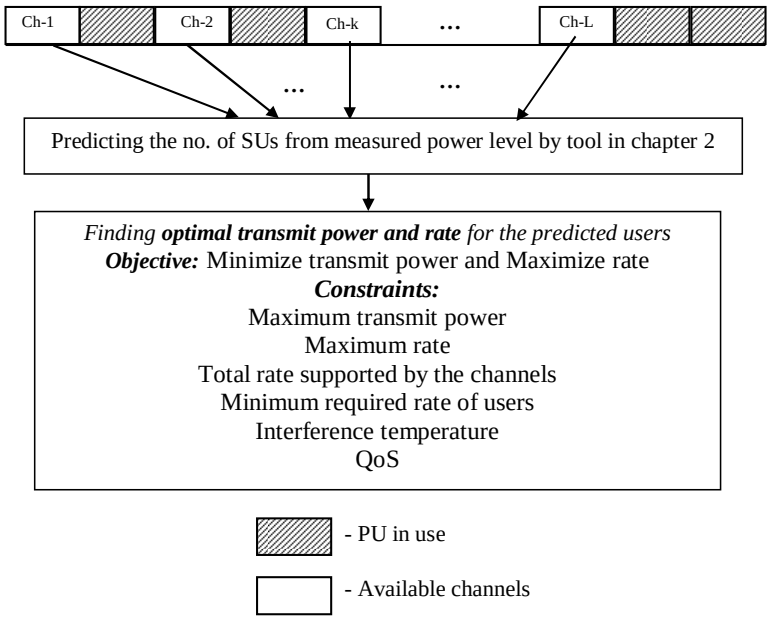

Figure 3.1: *Resource allocation model in cognitive radio network.*

Table 3.1: *Notations.*

$\widetilde{N}_s(k)$	Predicted number of users for k-th channel
$\sigma^2(k)$	Noise variance in k-th channel
$\rho_{j,i}$	Orthogonality factor between users j and i
$h_{i,i}(k)$	Power gain from i-th transmitter to i-th receiver in k-th channel
$h_{i,m}(k)$	Power gain from i-th transmitter at location m in k-th channel
$p_i(k)$	Transmit power per bit of i-th user in k-th channel
$p_i^{max}(k)$	Maximum transmit power per bit of i-th user in k-th channel
$I_{th}(k)$	Interference temperature constraint in k-th channel
$b_i(k)$	Rate of i-th user in k-th channel
$b_i^{max}(k)$	Maximum rate of i-th user in k-th channel
$R_{ch}^u(k)$	Maximum rate supported by k-th channel
R_i^l	Minimum required rate for i-th user
$p_{e,i}(k)$	BER for i-th user in k-th channel
$p_{e,i}^{th}$	BER threshold at receiver for i-th user in any channel
$\gamma_i(k)$	SINR per bit for i-th user in k-th channel
$\gamma_i^{th}(k)$	SINR per bit threshold at receiver for i-th user in k-th channel

3.2 Optimization Problem Formulation

In this section, we describe the proposed resource allocation framework that we solve in order to determine the best strategies for SUs from a resource utilization and QoS standpoint. We consider that the central controller has prior knowledge (based on traffic models and our proposed forecasting method in chapter 2) on which users are currently occupying each of the available channels. In the two-stage resource allocation framework, we decompose the minimization of total transmit power and maximization of total rate of SUs into two separate stages. The proposed resource allocation framework is shown as a block diagram in Fig. 3.2.

3.2.1 Stage 1: Centralized Power Allocation

In stage 1, our objective is to determine the best choice for SU transmit powers in each channel such that the overall power consumed is minimized and QoS (defined in terms of

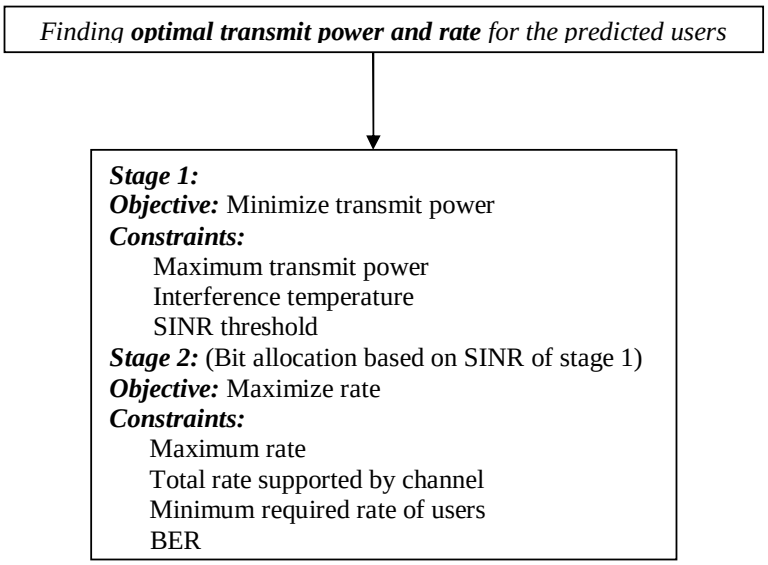

Figure 3.2: *Proposed two-stage resource allocation framework*

SINR) is maintained. The mathematical description of stage 1 corresponds to:

$$\begin{aligned}
\text{Determine} \quad & \mathbf{p} = [p_1(1) \cdots p_{\widetilde{N}_s(1)}(1) \cdots p_1(L) \cdots p_{\widetilde{N}_s(L)}(L)]^T \\
\text{To Minimize} \quad & \sum_{k=1}^{L} \sum_{i=1}^{\widetilde{N}_s(k)} p_i(k) \\
\text{subject to} \quad & \\
& C11: \ 0 \leq p_i(k) \leq p_i^{max}(k), \ \forall \ i, \ k; \\
& C12: \ \sum_{i=1}^{\widetilde{N}_s(k)} p_i(k) h_{i,m}(k) \leq I_{th}(k), \ \forall \ k; \\
& C13: \ \gamma_i(k) \geq \gamma_i^{th}(k), \ \forall \ i, \ k;
\end{aligned} \tag{3.1}$$

where

$$\gamma_i(k) = \frac{p_i(k) h_{i,i}(k)}{\sum_{j=1, \ j \neq i}^{\widetilde{N}_s(k)} p_j(k) h_{j,i}(k) {\rho_{j,i}}^2 + \sigma^2(k)}, \ \forall \ i, \ k. \tag{3.2}$$

Here, $C11$ indicates limit on transmit power; $C12$ indicates the interference temperature constraint, and $C13$ is SINR constraint required to guarantee desired QoS. This is a convex optimization (linear programming) problem. The following theorem shows the convexity of QoS/SINR constraint ($C13$).

Theorem 3.2.1. *$\gamma_i(k) \geq \gamma_i^{th}(k),\ \forall\ i,\ k$ is a convex constraint.*

Proof. From Eqs. (3.1) and (3.2), we can write the constraint as

$$p_i(k)h_{i,i}(k) \geq \gamma_i^{th}(k) \left(\sum_{j=1,\ j\neq i}^{\widetilde{N}_s(k)} p_j(k)h_{j,i}(k){\rho_{j,i}}^2 + \sigma^2(k) \right). \tag{3.3}$$

Equation (3.3) is equivalent to

$$\gamma_i^{th}(k) \left(\sum_{j=1,\ j\neq i}^{\widetilde{N}_s(k)} p_j(k)h_{j,i}(k){\rho_{j,i}}^2 + \sigma^2(k) \right) - p_i(k)h_{i,i}(k) \leq 0. \tag{3.4}$$

This inequality is a linear combination of the variables $p_i(k)$. Hence, the inequality is linear, can be treated as convex. □

Solution to this optimization problem provides optimal transmit powers that every secondary user needs to use in the channels that they are operating in.

In the following subsection, we derive the user-based distributed approach to solve the above proposed optimization problem.

3.2.2 Stage 1: Distributed Power Allocation

The centralized solution of the first stage requires a central controller and information about all users and channels. That is, centralized power allocation demands extensive control signalling and is difficult to implement in practice. Hence, we develop a distributed user-based approach to solve stage 1 of our proposed resource allocation framework. We use the dual decomposition of stage 1 optimization problem in order to derive the user-based power allocation algorithm.

For ease in presentation, we assume that there are equal number of users in each of the channels. The discussion below can be easily extended to the case when there are different number of users in each channel. Stage 1 of the proposed resource allocation problem (Eq. (3.1)) has one coupled constraint ($C12$) and one cross power term ($\sum_{j=1,\ j\neq i}^{\widetilde{N}_s(k)} p_j(k)h_{j,i}(k){\rho_{j,i}}^2$) in constraint $C13$. Introducing an auxiliary variable $in_i(k)$ (representing the interference

power that user i experiences in k-th channel) for the cross power term, the optimization problem can be restated as

$$\begin{aligned}
&\text{Determine} \quad [\mathbf{p}^T\ \mathbf{in}^T]^T \\
&\text{To Minimize} \\
&\qquad \sum_{k=1}^{L}\sum_{i=1}^{M} p_i(k) \\
&\text{subject to} \\
&\qquad C11:\ 0{\leq}p_i(k){\leq}p_i^{max}(k),\ \forall\ i,\ k; \\
&\qquad C12:\ \sum_{i=1}^{M} p_i(k)h_{i,m}(k){\leq}I_{th}(k),\ \forall\ k; \\
&\qquad C13:\ p_i(k)h_{i,i}(k){\geq}(in_i(k)+\sigma^2(k))\gamma_i^{th}(k),\forall\ i,\ k; \\
&\qquad C14:\ in_i(k){\geq}C_i(k),\ \forall\ i,\ k; \qquad (3.5)
\end{aligned}$$

where, $\mathbf{in} = [in_1(1)\cdots in_M(1)\cdots in_M(L)]^T$ and $C_i(k)$ equal to $\sum_{j=1,\ j\neq i}^{M} p_j(k)h_{j,i}(k){\rho_{j,i}}^2$ is the lower bound for $in_i(k)$. From (3.5), the Lagrangian of stage 1 optimization problem can be written as

$$\begin{aligned}
&\text{Determine} \quad [\mathbf{p}^T\ \mathbf{in}^T]^T \\
&\text{To Minimize:} \\
&\qquad \mathcal{L}_{ts} = \sum_{k=1}^{L}\sum_{i=1}^{M} p_i(k) + \sum_{k=1}^{L}\lambda(k)\left(\sum_{i=1}^{M} p_i(k)h_{i,m}(k) - I_{th}(k)\right) \\
&\qquad + \sum_{k=1}^{L}\sum_{i=1}^{M}\nu_i^{ts}(k)\left(in_i(k) - C_i(k)\right) \\
&\text{subject to} \\
&\qquad CD11:\ 0{\leq}p_i(k){\leq}p_i^{max}(k),\ \forall\ i,\ k; \\
&\qquad CD12:\ p_i(k)h_{i,i}(k){\geq}(in_i(k)+\sigma^2(k))\gamma_i^{th}(k),\forall\ i,\ k; \\
&\qquad CD13:\ in_i(k){\geq}C_i(k),\ \forall\ i,\ k; \qquad (3.6)
\end{aligned}$$

Here, $\lambda(k)$ and $\nu_i^{ts}(k)$ are dual variables. Rearranging (3.6) results in

$$\begin{aligned}
&\text{Determine} \quad [\mathbf{p}^T\ \mathbf{in}^T]^T \\
&\text{To Minimize:} \\
&\mathcal{L}_{ts} = \sum_{i=1}^{M}\sum_{k=1}^{L}\left(p_i(k) + \lambda(k)p_i(k)h_{i,m}(k) + \nu_i^{ts}(k)in_i(k)\right) - \sum_{k=1}^{L}\lambda(k)I_{th}(k) \\
&-\sum_{k=1}^{L}\sum_{i=1}^{M}\nu_i^{ts}(k)C_i(k) \\
&\text{subject to} \\
&CD11,\ CD12,\ CD13. \qquad (3.7)
\end{aligned}$$

Now, we can easily decompose the optimization problem (3.7) into M subproblems. Based on how we model the impact of $in_i(k)$ in each of the subproblems, three formulations for decomposed problem from Lagrangian (3.7) can be derived.

CASE 1: For the scenario when $in_i(k)$ is assumed constant but measurable, each of the subproblems can be written as

$$\begin{aligned}
&\text{Determine} \quad \mathbf{p}_i \\
&\text{To Minimize:} \quad g_i(\mathbf{p}_i,\ \boldsymbol{\lambda}) = \sum_{k=1}^{L} p_i(k)\left(1 + \lambda(k)h_{i,m}(k)\right) \\
&\text{subject to} \\
&CDL11:\ 0 \leq p_i(k) \leq p_i^{max}(k),\ \forall\ k; \\
&CDL12:\ p_i(k)h_{i,i}(k) \geq (in_i(k) + \sigma^2(k))\gamma_i^{th}(k), \forall\ k; \qquad (3.8)
\end{aligned}$$

where,

$$\mathbf{p}_i = [p_i(1)\ p_i(2)\cdots p_i(L)]^T, \qquad (3.9)$$

$$\boldsymbol{\lambda} = [\lambda(1)\ \lambda(2)\cdots\lambda(L)]^T, \qquad (3.10)$$

$$in_i(k) = \sum_{j=1,\ j\neq i}^{M} p_j(k)h_{j,i}(k){\rho_{j,i}}^2, \qquad (3.11)$$

$\mathbf{p}_i$ and $g_i(\mathbf{p}_i,\ \boldsymbol{\lambda})$ are the transmit powers across different channels and the Lagrangian

function for user i, respectively. The corresponding master dual problem is

$$\begin{aligned} &\text{Determine} && \boldsymbol{\lambda} \\ &\text{To Minimize:} && D_{ts,1}(\boldsymbol{\lambda}) = \sum_{i=1}^{M} g_i(\mathbf{p}_i,\ \boldsymbol{\lambda}) - \sum_{k=1}^{L} \lambda(k) I_{th}(k) \\ &\text{subject to} && \\ & && \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{3.12}$$

The user-based distributed power allocation algorithm using the gradient projection method (where the power and dual variables are adjusted in the opposite direction to the gradient $\nabla D_{ts,1}(\boldsymbol{\lambda})$) can be summarized as follows. Dual variables $\boldsymbol{\lambda}$ are initialized. $\mathbf{in}_i$ are measured. Each user executes one optimization subproblem to compute transmit power for each of its intended channels. At regular intervals, each user measures $\mathbf{in}_i$ and updates the dual variables. Each user continues to do the same until it achieves desired SINR along with satisfying system constraint $(C12)$. The pseudo code for the algorithm is shown in Algorithm 3.1. In Algorithm 3.1, t is the iteration counter, α_{ts} is a sufficiently small positive step-size.

Algorithm 3.1: Dual Algorithm to solve (3.7) based on **CASE 1**

Initialization: $\boldsymbol{p}(\mathbf{0}),\ \boldsymbol{\lambda}(\mathbf{0})$;
while termination criterion is not true **do**
▷ *% Execute subproblems*
for $i = 1,\ 2, \cdots,\ M$ **do**
Measure $\mathbf{in}_i$;
Solve optimization subproblem (3.8) for $\boldsymbol{p}_i^t$;
end for
▷ *% Update* $\boldsymbol{\lambda}$
for $k = 1,\ 2, \cdots,\ L$ **do**
if $(\sum_{i=1}^{M} p_i(k) h_{i,m}(k) > I_{th}(k))$ **then**
$\lambda^{t+1}(k) = [\lambda^t(k) - \alpha_{ts}(-\sum_{i=1}^{M} p_i(k) h_{i,m}(k) + I_{th}(k))]$;
else
$\lambda^{t+1}(k) = \lambda^t(k)$;
end if
end for
end while

The power sequences generated from Algorithm 3.1 converge to the optimal power solu-

tion. This implies that $D_{ts,1}(\boldsymbol{\lambda})$ is a Lipschitz function which guarantees the convergence of gradient projection algorithms [90].

Lemma 3.2.1. *(Lipschitz Continuity of $\nabla D_{ts,1}(\boldsymbol{\lambda})$) The function $D_{ts,1}(\boldsymbol{\lambda})$ is differentiable and there exists a constant $K > 0$ such that,*

$$||D_{ts,1}(\boldsymbol{\lambda}_a) - D_{ts,1}(\boldsymbol{\lambda}_b)||_2 \leq K||\boldsymbol{\lambda}_a - \boldsymbol{\lambda}_b||_2 \ \forall \boldsymbol{\lambda}_a,\ \boldsymbol{\lambda}_b\ \in\ \Re_+^L. \tag{3.13}$$

Proposition 3.2.1. *(Convergence of Algorithm 3.1) If $D_{ts,1}(\boldsymbol{\lambda})$ satisfies Lemma 3.2.1 and $0 < \alpha_{ts} < 2/K$, then starting with any power $\mathbf{0} \leq \boldsymbol{p}_i(\mathbf{0}) \leq \boldsymbol{p}_i^{max}$, $\forall i$ and dual variables $\boldsymbol{\lambda}(\mathbf{0}) \geq \mathbf{0}$, each point $(\boldsymbol{p},\ \boldsymbol{\lambda})$ of sequence $(\boldsymbol{p}^t,\ \boldsymbol{\lambda}^t)$ generated by algorithm is converged. The details on convergence of gradient projection algorithms can be found in [80, 91].*

It is important to note that the distributed approach does not fully avoid central control. This is due to the requirement of updating dual variables $\boldsymbol{\lambda}$. The dual variables capture information regarding how well the interference temperature threshold constraint is being satisfied. If the interference temperature threshold constraint is violated, then the corresponding dual variable increases in magnitude. This increase forces the objective function in our resource allocation problem to increase. To counter this effect, the resource allocation variables (power of users) are reduced which in turn improves the ability of satisfying the interference temperature threshold constraint.

CASE 2: Consider the case when $in_i(k)$ is assumed a variable. However, in each iteration of distributed approach, a lower bound for this interference is measurable. In this case, each of the subproblems can be written as,

$$\begin{aligned}
&\text{Determine} && [\mathbf{p}_i{}^T\ \mathbf{in}_i{}^T]^T \\
&\text{To Minimize:} && g_i(\mathbf{p}_i,\ \mathbf{in}_i,\ \boldsymbol{\lambda}) = \sum_{k=1}^{L} p_i(k)\,(1 + \lambda(k) h_{i,m}(k)) \\
&\text{subject to} && \\
& && CDL11,\ CDL12, \\
& && CDL13:\ in_i(k) \geq C_i(k),\ \forall\ k,
\end{aligned} \tag{3.14}$$

where, $C_i(k)$ is the lower bound for $in_i(k)$ and equal to $\sum_{j=1,\ j\neq i}^{M} p_j(k)h_{j,i}(k){\rho_{j,i}}^2$. The corresponding master dual problem is

$$\begin{aligned} &\text{Determine} && \boldsymbol{\lambda} \\ &\text{To Minimize:} && D_{ts,2}(\boldsymbol{\lambda}) = \sum_{i=1}^{M} g_i(\mathbf{p}_i,\ \mathbf{in}_i,\ \boldsymbol{\lambda}) - \sum_{k=1}^{L} \lambda(k) I_{th}(k) \\ &\text{subject to} && \\ & && \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{3.15}$$

The user-based distributed power allocation algorithm using the gradient projection method (where the power and dual variables are adjusted in the opposite direction to the gradient $\nabla D_{ts,2}(\boldsymbol{\lambda})$) can be summarized as follows. Dual variables $\boldsymbol{\lambda}$ are initialized. At the beginning of each iteration, each user measures the lower bound of expected interference from other users $\mathbf{C}_i$ (equals to $[C_i(1)\ C_i(2) \cdots C_i(L)]^T$). Each user executes one optimization subproblem to compute transmit power for each of its intended channels. At regular intervals, each user measures $\mathbf{C}_i$ and updates the dual variables. Each user continues to do the same until it achieves desired SINR along with satisfying system constraint ($C12$). The pseudo code for the algorithm is shown in Algorithm 3.2. Here, t and α_{ts} are as defined above. The pseudo code for the corresponding algorithm is shown in Algorithm 3.2.

Algorithm 3.2: Dual Algorithm to solve (3.7) based on **CASE 2**

Initialization: $\boldsymbol{p}(\mathbf{0}),\ \boldsymbol{\lambda}(\mathbf{0})$;
while termination criterion is not true **do**
▷ *% Execute subproblems*
for $i = 1,\ 2, \cdots,\ M$ **do**
Measure $\mathbf{C}_i$;
Solve optimization subproblem (3.14) for $\boldsymbol{p}_i^t$;
end for
▷ *% Update* $\boldsymbol{\lambda}$
for $k = 1,\ 2, \cdots,\ L$ **do**
if $(\sum_{i=1}^{M} p_i(k)h_{i,m}(k) > I_{th}(k))$ **then**
$\lambda^{t+1}(k) = [\lambda^t(k) - \alpha_{ts}(-\sum_{i=1}^{M} p_i(k)h_{i,m}(k) + I_{th}(k))]$;
else
$\lambda^{t+1}(k) = \lambda^t(k)$;
end if
end for
end while

The power sequences generated from Algorithm 3.2 converge to the optimal power solution. This implies that $D_{ts,2}(\boldsymbol{\lambda})$ is a Lipschitz function which guarantees the convergence of gradient projection algorithms [90].

Lemma 3.2.2. *(Lipschitz Continuity of $\nabla D_{ts,2}(\boldsymbol{\lambda})$) The function $D_{ts,2}(\boldsymbol{\lambda})$ is differentiable and there exists a constant $K > 0$ such that,*

$$||D_{ts,2}(\boldsymbol{\lambda}_a) - D_{ts,2}(\boldsymbol{\lambda}_b)||_2 \leq K||\boldsymbol{\lambda}_a - \boldsymbol{\lambda}_b||_2\ \forall \boldsymbol{\lambda}_a,\ \boldsymbol{\lambda}_b\ \in\ \Re_+^L. \tag{3.16}$$

Proposition 3.2.2. *(Convergence of Algorithm 3.2) If $D_{ts,2}(\boldsymbol{\lambda})$ satisfies Lemma 3.2.2 and $0 < \alpha_{ts} < 2/K$, then starting with any power $\mathbf{0} \leq \boldsymbol{p}_i(\mathbf{0}) \leq \boldsymbol{p}_i^{max}\ \forall i$ and dual variables $\boldsymbol{\lambda}(\mathbf{0}) \geq \mathbf{0}$, each point $(\boldsymbol{p},\ \boldsymbol{\lambda})$ of sequence $(\boldsymbol{p}^t,\ \boldsymbol{\lambda}^t)$ generated by algorithm is converged.*

CASE 3: An alternative formulation can be created by absorbing the constraint $CD13$

in the objective function. The subproblems for this case can be formulated as

$$\begin{aligned}
&\text{Determine} && [\mathbf{p}_i{}^T\ \mathbf{in}_i{}^T]^T \\
&\text{To Minimize:} && g_i(\mathbf{p}_i,\ \mathbf{in}_i,\ \boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}) = \sum_{k=1}^{L} \left(p_i(k)(1+\lambda(k)h_{i,m}(k)) + \nu_i^{ts}(k) in_i(k)\right) \\
&\text{subject to} && \\
& && CDL11,\ CDL12,\ CDL13.
\end{aligned} \tag{3.17}$$

The corresponding master dual problem is

$$\begin{aligned}
&\text{Determine} && [\boldsymbol{\lambda}^T\ \boldsymbol{\nu}_1^{ts,T}\ \cdots\ \boldsymbol{\nu}_i^{ts,T}\ \cdots\ \boldsymbol{\nu}_M^{ts,T}]^T \\
&\text{To Minimize:} && D_{ts,3}(\boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}\ \forall\ i) \\
& && = \sum_{i=1}^{M} g_i(\mathbf{p}_i,\ \mathbf{in}_i,\ \boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}) - \sum_{k=1}^{L} \lambda(k) I_{th}(k) - \sum_{k=1}^{L}\sum_{i=1}^{M} \nu_i^{ts}(k) C_i(k) \\
&\text{subject to} && \\
& && \boldsymbol{\lambda} \geq \mathbf{0}; \\
& && \boldsymbol{\nu}_i^{ts} \geq \mathbf{0};
\end{aligned} \tag{3.18}$$

where,

$$\boldsymbol{\nu}_i^{ts} = [\nu_i^{ts}(1)\ \nu_i^{ts}(2) \cdots \nu_i^{ts}(L)]^T. \tag{3.19}$$

The user-based distributed power allocation algorithm using the gradient projection method (where the power and dual variables are adjusted in the opposite direction to the gradient $D_{ts,3}(\boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}\ \forall\ i)$) can be summarized as follows. Dual variables $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}_i^{ts}\ \forall i$ are initialized. At the beginning of each iteration, each user measures the lower bound of expected interference from other users $\mathbf{C}_i$ (equals to $[C_i(1)\ C_i(2) \cdots C_i(L)]^T$). Each user executes one optimization subproblem to compute transmit power for each of its intended channels. At regular intervals, each user measures $\mathbf{C}_i$ and updates the dual variables. Each user continues to do the same until it achieves desired SINR along with satisfying system constraint ($C12$). The pseudo code for the algorithm is shown in Algorithm 3.3. Here, t and α_{ts} are

as defined above. The pseudo code for the corresponding algorithm is shown in Algorithm 3.3. In Algorithm 3.3, β_{ts} is also a sufficiently small positive step-size and $[\cdot]^+$ denotes the projection onto nonnegative orthant.

Algorithm 3.3: Dual Algorithm to solve (3.7) based on **CASE 3**

Initialization: $\boldsymbol{p}(\mathbf{0}),\ \boldsymbol{\lambda}(\mathbf{0})$;
Initialization: $\boldsymbol{\nu}_1^{ts}(\mathbf{0}),\ \cdots,\ \boldsymbol{\nu}_i^{ts}(\mathbf{0}),\ \cdots,\ \boldsymbol{\nu}_M^{ts}(\mathbf{0})$;
Measure $\mathbf{C}_1,\ \cdots,\ \mathbf{C}_i,\ \cdots,\ \mathbf{C}_M$;
while termination criterion is not true **do**
▷ *% Execute subproblems*
for $i = 1,\ 2, \cdots,\ M$ **do**
Solve optimization subproblem (3.17) for $\boldsymbol{p}_i^t$;
end for
▷ *% Update* $\boldsymbol{\lambda}$
for $k = 1,\ 2, \cdots,\ L$ **do**
if $(\sum_{i=1}^{M} p_i(k)h_{i,m}(k) > I_{th}(k))$ **then**
$\lambda^{t+1}(k) = [\lambda^t(k) - \alpha_{ts}(-\sum_{i=1}^{M} p_i(k)h_{i,m}(k) + I_{th}(k))]$;
else
$\lambda^{t+1}(k) = \lambda^t(k)$;
end if
end for
▷ *% Update* $\boldsymbol{\nu}_1^{ts}, \cdots,\ \boldsymbol{\nu}_i^{ts}, \cdots,\ \boldsymbol{\nu}_M^{ts}$
for $i = 1, 2, \cdots, M$ **do**
Measure $\mathbf{C}_i$;
for $k = 1,\ 2, \cdots,\ L$ **do**
$\nu_i^{ts,t+1}(k) = [\nu_i^{ts,t}(k) - \beta_{ts}(-in_i(k) + C_i(k))]^+$;
end for
end for
end while

The power sequences generated from Algorithm 3.3 converge to the optimal power solution. This implies that $D_{ts,3}(\boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}\ \forall\ i)$ is a Lipschitz function which guarantees the convergence of gradient projection algorithms [90].

Lemma 3.2.3. *(Lipschitz Continuity of $D_{ts,3}(\boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}\ \forall\ i)$) The function $D_{ts,3}(\boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}\ \forall\ i)$*

is differentiable and there exists a constant $K > 0$ *such that,*

$$||D_{ts,3}(\boldsymbol{\lambda}_a,\ \boldsymbol{\nu}_{a,i}^{ts}\ \forall\ i) - D_{ts,3}(\boldsymbol{\lambda}_b,\ \boldsymbol{\nu}_{b,i}^{ts}\ \forall\ i)||_2$$
$$\leq K||[\boldsymbol{\lambda}_a^T\ \boldsymbol{\nu}_{a,i}^{ts,T}\ \forall\ i]^T - [\boldsymbol{\lambda}_b^T\ \boldsymbol{\nu}_{b,i}^{ts,T}\ \forall\ i]^T||_2\ \forall \boldsymbol{\lambda}_b,\ \boldsymbol{\lambda}_b,\ \boldsymbol{\nu}_{a,i}^{ts}\ \forall\ i,\ \boldsymbol{\nu}_{b,i}^{ts}\ \forall\ i\ \in\ \Re_+^L. \quad (3.20)$$

Proposition 3.2.3. *(Convergence of Algorithm 3.3) If* $D_{ts,3}(\boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}\ \forall\ i)$ *satisfies Lemma 3.2.3 and* $0 < \alpha_{ts},\ \beta_{ts} < 2/K$*, then starting with any power* $\mathbf{0} \leq \boldsymbol{p}_i(\mathbf{0}) \leq \boldsymbol{p}_i^{max}\ \forall\ i$ *and dual variables* $\boldsymbol{\lambda}(\mathbf{0}),\ \boldsymbol{\nu}_i^{ts}(\mathbf{0})\ \forall\ i \geq \mathbf{0}$*, each point* $(\boldsymbol{p},\ \boldsymbol{\lambda},\ \boldsymbol{\nu}_i^{ts}\ \forall\ i)$ *of sequence* $(\boldsymbol{p}_i^t,\ \boldsymbol{\lambda}^t,\ \boldsymbol{\nu}_i^{ts,t}\ \forall\ i)$ *generated by algorithm is converged.*

In summary, based on a priori information or ability to measure interference power, we can formulate the different versions of distributed implementation of stage 1. It is also important to note that initializing dual variables and choice of step sizes are critical for convergence speed of the distributed solution [80, 91].

3.2.3 Stage 2: Centralized Rate Allocation

In the second stage, we attempt to satisfy the rate requirement for each secondary user. Our goal in this stage is to determine how each SU distributes its information across the multiple channels in a way that the overall rate is maximized and the individual rate requirement is met. Employing the optimal transmit powers and SINRs from first stage, the following rate

allocation problem is proposed:

$$\begin{aligned}
&\text{Determine} \quad \mathbf{b} \\
&\text{To Maximize} \\
&\qquad \sum_{k=1}^{L} \sum_{i=1}^{\widetilde{N}_s(k)} b_i(k) \\
&\text{subject to} \\
&\qquad C15: \ b_i(k) \in [1, \cdots, \ b_i^{max}(k)], \ \forall \ i, \ k; \\
&\qquad C16: \ \sum_{i=1}^{\widetilde{N}_s(k)} b_i(k) \leq R_{ch}^u(k), \ \forall \ k; \\
&\qquad C17: \ \sum_{k=1}^{L} b_i(k) \geq R_i^l \ \forall \ i; \\
&\qquad C18: \ p_{e,i}(k) \leq p_{e,i}^{th}, \ \forall \ i, \ k; \qquad (3.21)
\end{aligned}$$

where,

$$p_{e,i}(k) = \frac{4}{b_i(k)} \left(1 - 2^{-\frac{b_i(k)}{2}}\right) Q\left(\sqrt{\frac{3b_i(k)\gamma_i(k)}{(2^{b_i(k)}-1)}}\right), \forall \ i, \ k, \ \text{even } b_i(k); \qquad (3.22)$$

$$p_{e,i}(k) \leq \frac{4}{b_i(k)} Q\left(\sqrt{\frac{3b_i(k)\gamma_i(k)}{(2^{b_i(k)}-1)}}\right), \ \forall \ i, \ k, \ \text{odd } b_i(k). \qquad (3.23)$$

Here, $\mathbf{b} = [b_1(1) \cdots b_{\widetilde{N}_s(1)}(1) \cdots b_{\widetilde{N}_s(L)}(L)]^T$; $C15$ indicates limit on rate; $C16$ indicates total rate that a channel can support, $C17$ captures the rate requirements for each SU, and $C18$ is BER requirement for every SU. It is to be noted that $Q(x)$ is defined as $\int_x^\infty e^{-\zeta^2/2} d\zeta$. It is very easy to show that to achieve a certain BER, constraint $C18$ is equivalent to the following constraint

$$C19: \ -\gamma_i(k) \leq -C_{qarg}(2^{b_i(k)} - 1), \ \forall \ i, \ k; \qquad (3.24)$$

where, C_{qarg} is a constant and can be determined using (1) minimum rate, $b_i^{min}(k) = 1$ (in our system); (2) $b_i^{max}(k)$, and (3) value of $p_{e,i}^{th}$ from $C18$. As an example, with $b_i^{min}(k) = 1$ to achieve $p_{e,i}^{th} = 10^{-3}$, Eq. (3.23) suggests that $\gamma_i(k)/(2^{b_i(k)} - 1)$ has to be greater than

4.08 and with $b_i^{max}(k) = 6$, Eq. (3.22) suggests that $\gamma_i(k)/(2^{b_i(k)} - 1)$ has to be greater than 0.50. From this, we can conclude that by setting $C_{qarg} = 4.08$, we can guarantee a BER that is less than or equal to 10^{-3} for the feasible values of $b_i(k)$. For ease in presentation, we define $Q_{qarg,i}(k) = -\gamma_i(k)/(2^{b_i(k)} - 1)$; so that constraint $C19$ can be rewritten as

$$C19: \ Q_{qarg,i}(k) \leq -C_{qarg}. \tag{3.25}$$

Optimal Allocation

The rate allocation problem (Eq. (3.21)) can be formulated as a maximum flow problem in a directed network in graph theory. A directed network is expressed as $G = (N, A)$ defined by a set N of n nodes and a set A of m directed links [92]. Each link $(i, j) \in A$ is associated with a capacity u_{ij} that denotes the maximum amount that can flow on the link and a lower bound l_{ij} that denotes the minimum amount that must flow on the link. The maximum flow problem seeks a feasible solution that sends the maximum amount of flow from a specified source node s to another specified sink node d in such a directed network. The rate measure in our problem of interest takes the role of flow in the maximum flow problem formulation. Therefore, the equivalent graph formulation of Eq. (3.21) corresponds to Fig. 3.3.

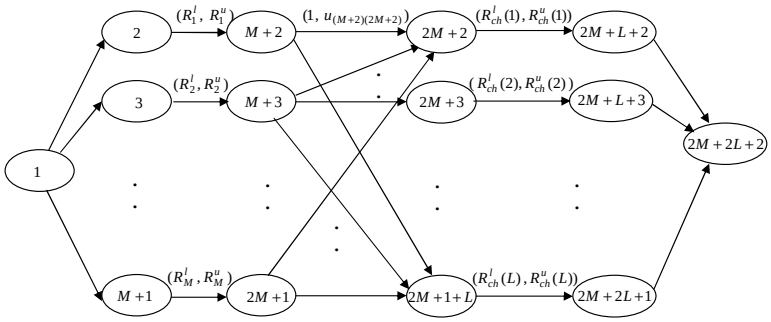

Figure 3.3: *Rate distribution problem as a maximum flow problem in graph theory.*

In the graph shown in Fig. 3.3, nodes 1 and $2M + 2L + 2$ are the source and sink nodes, respectively. Nodes $(2, 3, \cdots, M{+}1, M{+}2, \cdots, 2M{+}1)$ and $(2M{+}2, 2M{+}3, \cdots, 2M{+}1{+}L, 2M{+}L{+}2, \cdots, 2M{+}2L{+}1)$ represent users and channels, respectively. The lower bounds

on link capacities between nodes $(2,\ 3,\cdots,\ M+1)$ and $(M+2,\cdots,\ 2M+1)$ can be obtained from the minimum rate requirement of the users (constraint $C17$). The upper bounds on link capacities for these links can be set to some reasonable high values. As for example, the upper bound can be set to the value obtained by multiplying the maximum capacity of the links between nodes $(M+2,\ M+3,\cdots,\ 2M+1)$ and $(2M+2,\ 2M+3,\cdots,\ 2M+1+L)$ by the number of total available channels. The upper bounds on link capacities between nodes $(2M+2,\ 2M+3,\cdots,\ 2M+1+L)$ and $(2M+L+2,\ 2M+L+3,\cdots,\ 2M+2L+1)$ can be obtained from the maximum rate supporting capabilities of the channels (constraint $C16$). The lower bounds on link capacities for these links i.e., $R^l_{ch}(k)$'s can be set to 0. The lower and upper bounds on link capacities between nodes $(M+2,\ M+3,\cdots,\ 2M+1)$ and $(2M+2,\ 2M+3,\cdots,\ 2M+1+L)$ can be computed from the constraints $C15$ and $C19$ with obtained SINR in first stage. The upper bounds on the link capacities between nodes 1 to $(2,\ 3,\cdots,\ M+1)$ and $(2M+L+2,\ 2M+L+3,\cdots,\ 2M+2L+1)$ to $2M+2L+2$ can be set to the total capacity of the links between nodes between $(M+2,\ M+3,\cdots,\ 2M+1)$ to $(2M+2,\ 2M+3,\cdots,\ 2M+1+L)$.

There are several algorithms to solve maximum flow problem such as labeling algorithm, capacity scaling algorithm, successive shortest path algorithm. The running time of the labeling algorithm, capacity scaling algorithm and successive shortest path algorithm are $O(nmU)$, $O(nmlogU)$ and $O(n^2m)$, respectively [92]. Here, U is the maximum capacity of the links in the network. The running time may increase with a high number of nodes (n) or links (m) or maximum capacity (U) of the link in the network. The running time of the above mentioned algorithms for finding maximum flow in a network corresponding to Fig. 3.3 is shown as $O(M^2LU)$, $O(M^2LlogU)$ and $O(M^2L)$, respectively. In the following, we develop an heuristic to solve problem (3.21).

Heuristic Allocation

The rate allocation algorithm that can be employed to solve rate allocation problem (3.21) as follows. First, we allocate the maximum feasible rate (i.e., maximum $b_i(k)$ that satisfies

Eq. (3.25)) to all users across channels. Based on this allocated rate, the average $Q_{qarg,i}(k)$ is calculated for all users and compared with $-C_{qarg}$ in the next step. For a specific user, if average $Q_{qarg,i}(k)$ does not satisfy constraint $C19$, the maximum rate allocated to a channel for that user is reduced by 1. This process is repeated until constraint $C19$ is satisfied or a maximum number of iterations (l_1^{max}) are completed. In the latter case, the average $Q_{qarg,i}(k)$ (though not satisfactory) is the achievable $Q_{qarg,i}(k)$ for that user. In the following step, the algorithm checks if the total rate limits that are set for all channels are violated. If the rate constraint per channel (constraint $C16$) is not met, the maximum rate allocated to a user in that channel is reduced by 1. This process is repeated until constraint $C16$ is satisfied or a maximum number of iterations (l_2^{max}) are completed. The pseudo code for the algorithm is provided below:

Algorithm 3.4: Rate allocation algorithm

▷ % *Initialize rate*

for $k = 1,\ 2, \cdots,\ L$ **do**

  for $i = 1,\ 2, \cdots,\ \widetilde{N}_s(k)$ **do**

    $b_i(k)$=max$(1, \cdots,\ b_i^{max}(k))$ that satisfies Eq. (3.25);

  end for

end for

▷ % *Checking average bit error rate*

for $i = 1,\ 2, \cdots,\ M$ **do**

  Stopping counter, $l_1 = 1$;

  while $l_1 < l_1^{max}$ **do**

    Compute $\bar{Q}_{qarg,i} = \frac{\sum_{k=1}^{L} b_i(k) Q_{qarg,i}(k)}{\sum_{k=1}^{L} b_i(k)}$;

    if $\bar{Q}_{qarg,i} > -C_{qarg}$ **then**

      Reduce the highest rate (in a channel) by 1;

      $l_1 = l_1 + 1$;

    else

      $l_1 = l_1^{max}$;

    end if

  end while

end for

▷ % *Checking rate supported by a channel*

for $k = 1,\ 2, \cdots,\ L$ **do**

  Stopping counter, $l_2 = 1$;

  while $l_2 < l_2^{max}$ **do**

    if $\sum_{i=1}^{\widetilde{N}_s(k)} b_i(k) > R_{ch}^{u}(k)$ **then**

      Reduce the highest rate (allocated to the user) by 1;

      $l_2 = l_2 + 1$;

    else

      $l_2 = l_2^{max}$;

    end if

  end while

end for

The running time of our developed heuristic is shown as $O(M(max\{L,\ l_1^{max},\ l_2^{max}\}))$ which is less than that of the optimal graph theoretic algorithms.

In summary, the two-stage resource allocation framework decomposes the power calculation and rate allocation into two stages. In the first stage, the transmit power for every

SU is governed by the SINR threshold and in the second stage, we attempt to maximize the rate for each SU given the BER requirement.

3.3 Numerical Results

In this section, we quantify the performance of the proposed two-stage resource allocation framework. We assume a CRN with $L = 11$ available channels and a total of $M = 10$ secondary users. We assume a usage pattern as shown in Table 3.2, where a 1 indicates that the corresponding channel is being used by the SU. Table 3.3 provides information on the channel quality for all L channels. Table 3.4 lists the minimum rate requirement for each SU. Finally, Table 3.5 contains all other system parameters that are relevant to our resource allocation framework. Based on all this information, our objective is to find the optimal transmit power and rate that each of the M SUs should employ to guarantee their QoS.

Table 3.2: *Usage pattern across channels.*

Channel, k	1	2	3	4	5	6	7	8	9	10	11
User, 1	1	0	1	0	0	1	1	0	0	0	1
User, 2	1	1	1	0	0	1	1	0	1	0	1
User, 3	0	1	1	0	1	1	0	0	1	0	1
User, 4	0	1	1	0	1	1	1	0	1	1	1
User, 5	0	1	1	1	1	1	1	1	1	1	1
User, 6	0	1	1	1	1	0	1	1	1	1	1
User, 7	0	1	0	1	1	0	1	1	0	1	1
User, 8	1	1	1	0	1	0	1	1	0	1	1
User, 9	1	0	1	0	1	0	1	1	0	1	1
User, 10	1	0	1	0	1	1	0	1	0	1	1

Table 3.3: *Channel quality parameters.*

Channel, k	1	2	3	4	5	6	7	8	9	10	11
$\sigma^2(k)$, $(\times 10^{-3})$	5	4	3	2	2.5	6	4	4	5	3.5	4.5

Table 3.4: *Minimum rate requirement of users.*

User, u	1	2	3	4	5	6	7	8	9	10
R_i^l	3	8	4	12	9	7	14	5	10	8

Table 3.5: *System parameters.*

$p_i^{max}(k)\ \forall\ i,\ k$	5
$b_i^{max}(k)\ \forall\ i,\ k$	6
$p_{e,i}^{th}\ \forall\ i$	10^{-3}
$I_{th}(k)\ \forall\ k$	$200 \times \sigma^2(k)$
$R_{ch}^u(k)\ \forall\ k$	20
$\rho_{j,i}$	0.03125

As discussed in section 3.2, the first stage of resource allocation framework is a linear programming (LP) problem, and any LP solver can be used to find the solution. In this work, we use the "Linear Interior Point Solver (LIPSOL)" to solve stage 1. LIPSOL method has been briefly discussed in Appendix A. We use "Mixed Integer Programming (MIP)" to solve second stage for optimal rate distribution.We set the SINR threshold, $\gamma_i^{th}(k)$ to 12 dB. Based on the system parameters defined earlier, we can calculate C_{qarg} at 4.5. For second stage, we set $R_{ch}^u(k)$ to 20 for all channels.

Figure 3.4 illustrates the transmit power and rate allocation across channels for users 1 and 8. The channel noise variance and resulting SINR are also plotted for reference. Here, user 1 operates on channels 1, 3, 6, 7 and 11; user 8 operates on channels 1, 2, 3, 5, 7, 8, 10 and 11. From Fig. 3.4, it is evident that for both users, higher transmit powers are allocated to channels with higher noise variance. In other words, optimal transmit power allocation follows "reverse water filling" process since the goal is to satisfy the minimum SINR threshold. Figure 3.4 also indicates that the SINR threshold is attained in every channel. The allocated rate across channels directly follows SINR and since SINR is maintained at the threshold value, the rate allocation is also a constant across channels. The allocated power

and rate for other users follow the pattern presented for users 1 and 8. Figure 3.5 shows the total transmit power and rate allocation across users. From this figure, we can conclude that the proposed resource allocation framework has been successful in meeting the rate requirement for every active SU. Figure 3.6 captures the effects of increasing number of users on the total transmit power and rate for user 1. It is clear that with increase in the number of users in the system (i.e., increasing the number of users in the channels based on usage pattern), user 1 is forced to use higher transmit power. This is because, with increase in number of users, interference increases and therefore more power is required to satisfy the SINR threshold. Since, SINR is maintained at the threshold level irrespective of the number of users, the total rate (that is a function of SINR) for user 1 remains unchanged as seen in Fig. 3.6.

Figure 3.7 shows us the allocation of transmit power across users obtained from three formulations of distributed approach, i.e., CASE 1, CASE 2 and CASE 3 with centralized solution. In each case, we initialize dual variable $\lambda(k)$ to 0 for all channels. The step size α_{ts} is set to 0.1. For CASE 3, $\nu_i^{ts}(k)$ is initialized to 0 for all users and channels and the step size β_{ts} is set to 0.1. From Fig. 3.7, we can conclude that the solution from each of the distributed formulations converges to the centralized solution.

The distributed formulations require measurement of interference power ($\mathbf{in}_i$ for CASE 1, $\mathbf{C}_i$ for CASE 2 and 3). Figure 3.8 shows the number of iterations that the three formulations of distributed approach need to converge with error (overestimation) in measurements. This figure illustrates that with increase in percentage of error, the number of iteration decreases. The reason behind this behavior can be better understood by observing the impact of erroneous measurement of interference power in each of the subproblems. When $\mathbf{in}_i$ for CASE 1, $\mathbf{C}_i$ for CASE 2 and 3 are overestimated, then it causes an increase in the magnitude of optimizing variables $\mathbf{p}_i$. The increase in magnitude of optimizing variables improves the ability to satisfy the SINR constraints at a faster rate.

Figure 3.9 illustrates the rate distribution resulting from our proposed heuristic with

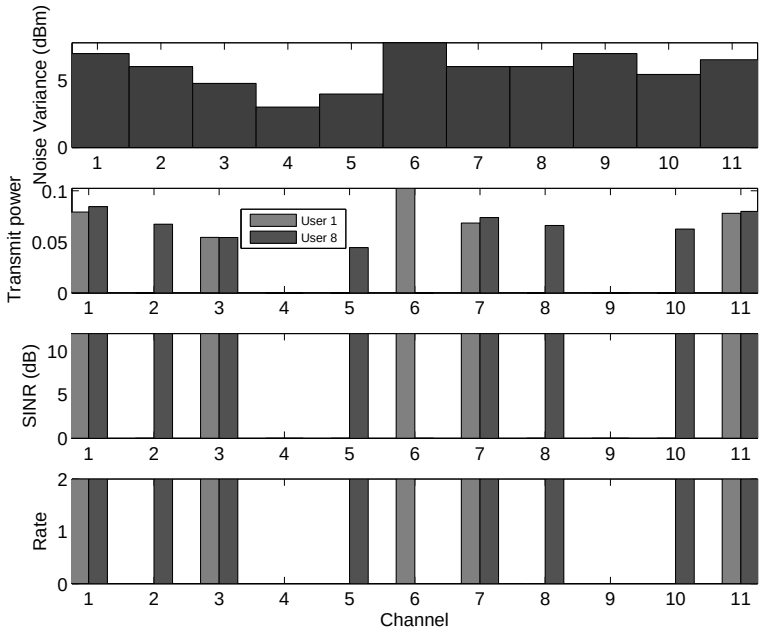

Figure 3.4: *Allocation of transmit power and rate with channel noise variance and SINR for users* 1 *and* 8.

optimal graph theoretic approach. We consider three example cases. In example 1, $R^u_{ch}(k)$ is set as 20 for all channels; In example 2, $R^u_{ch}(k)$ is set as 15 for all channels; In example 3, the maximum rate supporting capabilities of the channels are set as 10, 12, 14, 18, 15, 8, 11, 11, 8, 14 and 14, respectively for channels 1 through 11. The total minimum rate requirement for all users in the system is 80 in all examples. In example 1, both heuristic and graph theoretic approaches result in identical rate distribution for users. However in examples 2 and 3, the rate distribution profile from both approaches are different. In all three example cases, the total rate $\left(\sum_{k=1}^{L}\sum_{i=1}^{\widetilde{N}_s(k)} b_i^{opt}(k)\right)$ supported is found to be equal. From Fig. 3.9, we can conclude that our proposed heuristic performs comparable to optimal graph theoretic approach.

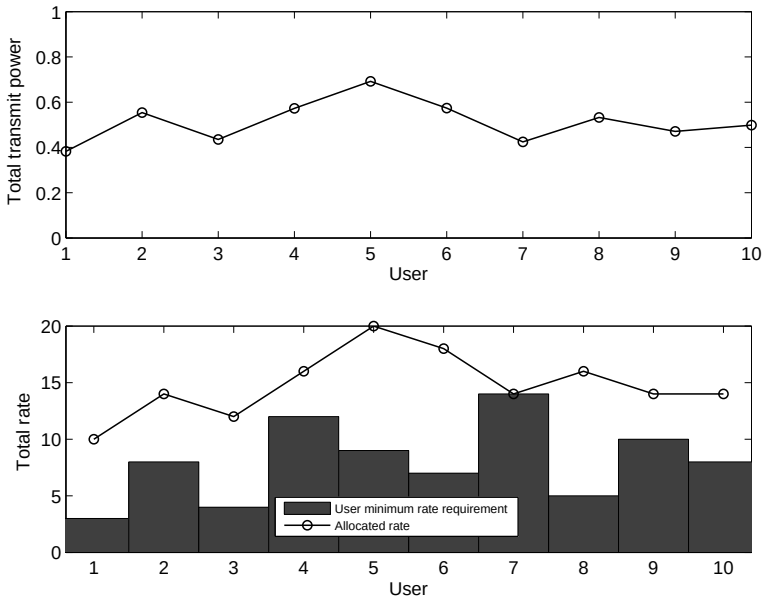

Figure 3.5: *Allocation of total transmit power and total rate across users.*

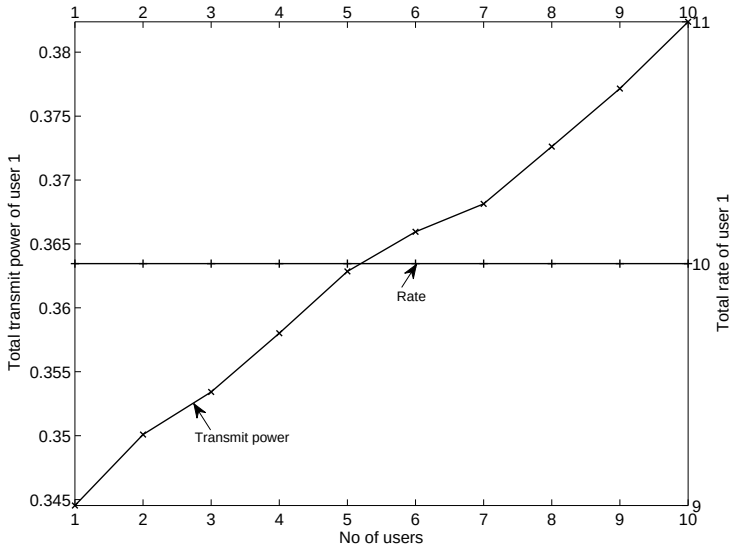

Figure 3.6: *Total transmit power and total rate for user* 1 *with number of users.*

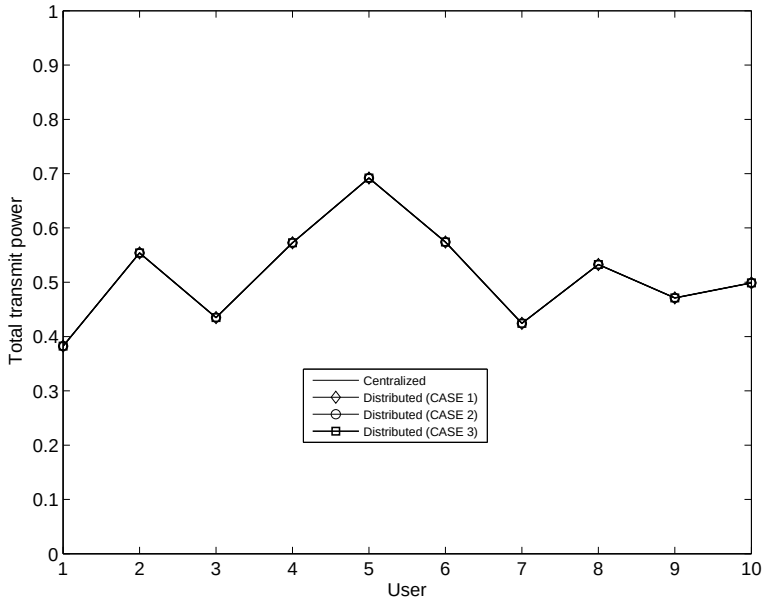

Figure 3.7: *Allocation of total transmit power across users from different distributed approaches.*

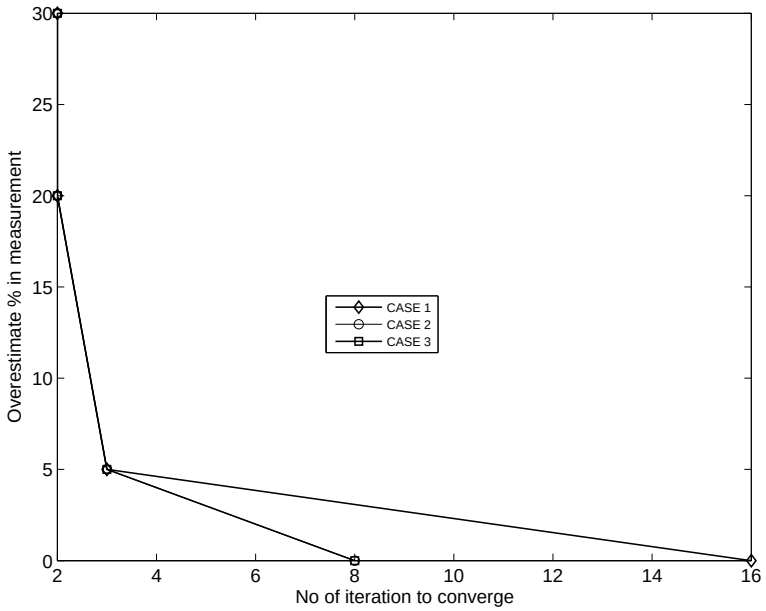

Figure 3.8: *Convergence speed of the distributed approach with imperfect measurement of interference power of adjacent users.*

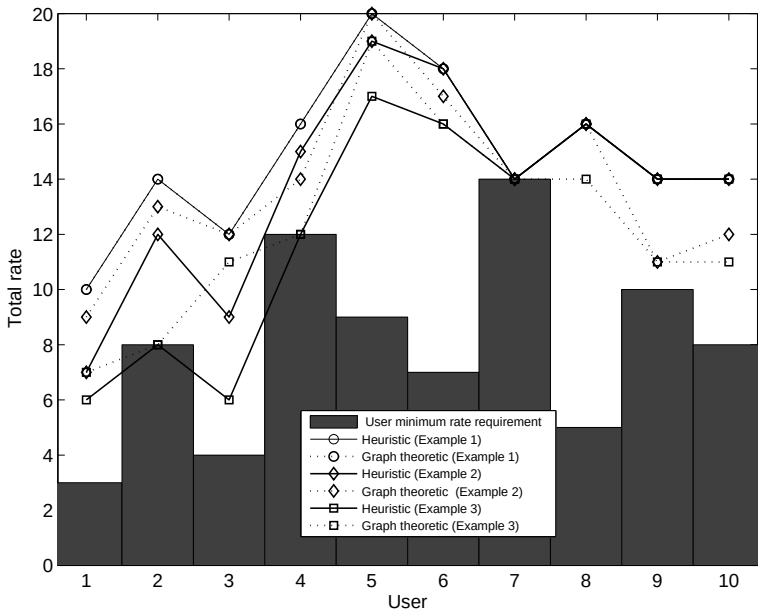

Figure 3.9: *Total rate allocation across users from proposed heuristic and graph theoretic analysis.*

3.4 Summary

In this chapter, we propose a two-stage resource allocation framework that provides the optimal transmit power and rate distribution that each SU needs to employ while maintaining QoS in a multi-channel CRN. We assume that multiple secondary users may coexist in a single channel and a single secondary user can simultaneously employ multiple channels to meet its rate requirements. We show that optimal transmit power follows reverse water filling process and optimal rate allocation is proportional to SINR. We also observe that the dual decomposed user-based distributed solution of stage 1 converges to centralized solution and rate distribution in stage 2 based on our proposed heuristic is close to optimal graph theoretic solution.

Chapter 4

Joint Resource Allocation

In chapter 3, we proposed a two-stage resource allocation framework to determine transmit power and rate separately for secondary users in a multi-channel multiuser competitive CRN. In this chapter, we jointly determine the best choice of power and rate distribution for every SU with the help of a bi-objective optimization problem formulation. We adopt the same system model and notations introduced in chapter 3. As in two-stage resource allocation framework, our objective is to determine the optimal distribution of power and rate that a secondary user has to employ across the channels that it uses in order to (1) minimize total power consumption; (2) maximize rate, and (3) maintain QoS. Firstly, resource allocation problem is solved in a centralized manner and then we employ dual decomposition theory to derive three different distributed solutions.

The rest of the chapter is organized as follows. Section 4.1 describes the proposed resource allocation framework. Section 4.2 shows the dual distributed versions in detail. Numerical results are presented in Sec. 4.3. Finally, Sec. 4.4 summarizes the chapter.

4.1 Optimization Problem Formulation

The objectives of the joint resource allocation framework are to (1) minimize the total transmit power, and (2) maximize the total rate while satisfying the QoS requirements of all active SUs. The mathematical description of the proposed bi-objective resource allocation

corresponds to:

$$
\begin{aligned}
&\text{Determine} && [\mathbf{p}^T\ \mathbf{b}^T]^T \\
&\text{To Minimize:} && F_1 = \sum_{k=1}^{L} \sum_{i=1}^{\widetilde{N}_s(k)} p_i(k) \text{ and} \\
&\text{Maximize:} && F_2 = \sum_{k=1}^{L} \sum_{i=1}^{\widetilde{N}_s(k)} b_i(k) \\
&\text{subject to} && \\
& && C21: \ 0 \leq p_i(k) \leq p_i^{max}(k), \ \forall\ i,\ k; \\
& && C22: \ b_i(k) \in [1, \cdots, \ b_i^{max}(k)], \ \forall\ i,\ k; \\
& && C23: \ \sum_{i=1}^{\widetilde{N}_s(k)} p_i(k) h_{i,m}(k) \leq I_{th}(k), \ \forall\ k; \\
& && C24: \ \sum_{i=1}^{\widetilde{N}_s(k)} b_i(k) \leq R_{ch}^{u}(k), \ \forall\ k; \\
& && C25: \ \sum_{k=1}^{L} b_i(k) \geq R_i^{l}, \ \forall\ i; \\
& && C26: \ p_{e,i}(k) \leq p_{e,i}^{th}, \ \forall\ i,\ k.
\end{aligned} \tag{4.1}
$$

Here, $\mathbf{p}$ and $\mathbf{b}$ are as defined in chapter 3; $p_{e,i}(k)$ is as defined earlier in Eqs. (3.22)- (3.23); constraints $C21$ and $C22$ indicate limits on transmit power and rate, respectively; $C23$ indicates the interference temperature constraint; $C24$ indicates the total rate supported by a channel; $C25$ represents the required rate of users and finally $C26$ is QoS/BER constraint. Since $b_i(k)$ is discrete and constraint $C26$ is nonlinear, the optimization problem formulation presented above is a constrained multi-objective mixed integer nonlinear programming (multi-objective MINLP) optimization problem, which is NP-hard in general. Relaxing the integer constraint on rate, $b_i(k)$ (as assumed in [93]) and assuming $b_i(k)$ as continuous variable, the above optimization problem can be restated with $C22$ as:

$$
C22: 1 \leq b_i(k) \leq b_i^{max}(k), \ \forall\ i,\ k. \tag{4.2}
$$

This optimization problem is non-convex due to constraint $C26$. We transform $C26$ into a simplified form (similar to the approximation of $C18$ to $C19$ in two-stage resource allocation) and then perform convex approximation to ensure optimality of the resulting solution. Constraint $C26$ can be rewritten as

$$C27: \quad -\gamma_i(k) \leq -C_{qarg}(2^{b_i(k)}-1), \ \forall \ i, \ k; \tag{4.3}$$

where, C_{qarg} is determined using the same procedure as discussed in two-stage resource allocation framework. The following theorem discusses the convex approximation of constraint $C27$.

Theorem 4.1.1. *$-\gamma_i(k) \leq -C_{qarg}(2^{b_i(k)}-1), \ \forall \ i, \ k$ is a convex constraint.*

Proof. Equation (4.3) can be written as

$$\gamma_i(k) \geq C_{qarg}(2^{b_i(k)}-1). \tag{4.4}$$

From Eqs. (3.2) and (4.4), we can write

$$\begin{aligned} p_i(k)h_{i,i}(k) \geq & \quad C_{qarg}(2^{b_i(k)}-1)\left(\sum_{j=1 \ j\neq i}^{\widetilde{N}_s(k)} p_j(k)h_{j,i}(k){\rho_{j,i}}^2 + \sigma^2(k)\right) \\ = & \quad C_{qarg}\sum_{j=1 \ j\neq i}^{\widetilde{N}_s(k)} p_j(k)2^{b_i(k)}h_{j,i}(k){\rho_{j,i}}^2 + C_{qarg}2^{b_i(k)}\sigma^2(k) \\ & -C_{qarg}\sum_{j=1 \ j\neq i}^{\widetilde{N}_s(k)} p_j(k)h_{j,i}(k){\rho_{j,i}}^2 - C_{qarg}\sigma^2(k). \end{aligned} \tag{4.5}$$

Finally, rearranging Eq. (4.5), we get

$$\begin{aligned} C_{qarg}\sum_{j=1 \ j\neq i}^{\widetilde{N}_s(k)} p_j(k)2^{b_i(k)}h_{j,i}(k){\rho_{j,i}}^2 + C_{qarg}2^{b_i(k)}\sigma^2(k) - C_{qarg}\sum_{j=1 \ j\neq i}^{\widetilde{N}_s(k)} p_j(k)h_{j,i}(k){\rho_{j,i}}^2 \\ -p_i(k)h_{i,i}(k) - C_{qarg}\sigma^2(k) \leq 0. \end{aligned} \tag{4.6}$$

Here, $b_i(k)$, $p_i(k)$ and $p_j(k)$ are the optimization variables. $2^{b_i(k)}$ is a convex function. The second term is convex as it is a function of $2^{b_i(k)}$. The third and fourth terms are convex as these are linear functions of $p_j(k)$ and $p_i(k)$, respectively. The first term vanishes if $\rho_{j,i} = 0$ and the entire inequality becomes convex i.e., users are orthogonal and constraint $C27$ is convex. But if $\rho_{j,i}$ is not equal to zero, the component functions $p_j(k)2^{b_i(k)}$ can be

linearized via Taylor expansion around a point of interest $[p_j^t(k)\ b_i^t(k)]^T$. In that case, the entire inequality (Eq. (4.6)) can be considered a convex inequality corresponding to

$$C_{qarg}\sum_{j=1\ j\neq i}^{\widetilde{N}_s(k)}\left(p_j^t(k)2^{b_i^t(k)}+2^{b_i^t(k)}\left(p_j(k)-p_j^t(k)\right)+p_j^t(k)2^{b_i^t(k)}log_e 2\left(b_i(k)-b_i^t(k)\right)\right)$$
$$\times h_{j,i}(k){\rho_{j,i}}^2+C_{qarg}2^{b_i(k)}\sigma^2(k)-C_{qarg}\sum_{j=1\ j\neq i}^{\widetilde{N}_s(k)}p_j(k)h_{j,i}(k){\rho_{j,i}}^2$$
$$-p_i(k)h_{i,i}(k)-C_{qarg}\sigma^2(k)\leq 0. \tag{4.7}$$

□

The optimization problem defined above has two objective functions F_1 and F_2 that work against each other. That is, as each SU attempts to increase the rate (bits/channel use) in order to maximize F_2, the constraint $C27$ becomes difficult to satisfy unless more transmit power is used. Therefore, F_1 will increase if we attempt to increase F_2 and vice-versa. It is common to combine such mutually conflicting objectives into a single objective function using the "weighted sum" approach [94] and look at pareto optimal solutions. The optimization problem with combined single objective can now be rewritten as:

$$\text{Minimize } \tau_1 F_1-\tau_2 F_2 \tag{4.8}$$
$$\text{subject to}$$
$$C21,\ C22,\ C23,\ C24,\ C25,\ C27.$$

The parameters τ_1 and τ_2 in the combined objective function are the scalarization factors and can be set following the discussion in [94]. Finally, we use the solution obtained from the convex formulation (Eq. (4.8)) as a starting point to search in the neighborhood for the optimal discrete valued $b_i(k)$ (denoted as $\mathbf{b}^{opt}$). Based on the new discrete solution, the optimal transmit power $\mathbf{p}^{opt}$ is recalculated using Eq. (4.3).

In the next section, we derive the user-based distributed approach to solve the proposed resource allocation problem.

4.2 Distributed Implementation

The centralized solution of the proposed resource allocation framework demands extensive control signalling and is difficult to implement in practice as information about all users and channels is needed at the central controller. Hence, we develop a distributed user-based approach to solve the proposed resource allocation problem. We use dual decomposition theory to derive the user-based joint power and rate allocation algorithm.

For ease in presentation, we assume that there are equal number of users in each of the channels. The discussion below can be easily extended to the case when there are different number of users in each channel. The proposed resource allocation problem has two coupled constraints ($C23$ and $C24$) and one cross power term ($\sum_{j=1,\ j\neq i}^{\widetilde{N}_s(k)} p_j(k)h_{j,i}(k){\rho_{j,i}}^2$) in constraint $C27$. Introducing an auxiliary variable $in_i(k)$ (representing the interference power that user i experiences in k-th channel) for the cross power term, the resource allocation problem (Eq. (4.8)) can be restated as

$$
\begin{aligned}
\text{Determine} \quad & [\mathbf{p}^T\ \mathbf{in}^T\ \mathbf{b}^T]^T \\
\text{To Minimize:} \quad & \tau_1 \sum_{k=1}^{L}\sum_{i=1}^{M} p_i(k) - \tau_2 \sum_{k=1}^{L}\sum_{i=1}^{M} b_i(k) \\
\text{subject to} \quad & \\
& C21:\ 0 \leq p_i(k) \leq p_i^{max}(k),\ \forall\ i,\ k; \\
& C22: 1 \leq b_i(k) \leq b_i^{max}(k),\ \forall\ i,\ k; \\
& C23:\ \sum_{i=1}^{M} p_i(k)h_{i,m}(k) \leq I_{th}(k),\ \forall\ k; \\
& C24:\ \sum_{i=1}^{M} b_i(k) \leq R_{ch}^{u}(k),\ \forall\ k; \\
& C25:\ \sum_{k=1}^{L} b_i(k) \geq R_i^{l},\ \forall\ i; \\
& C27:\ -p_i(k)h_{i,i}(k) \leq -C_{qarg}(in_i(k) + \sigma^2(k))(2^{b_i(k)} - 1),\ \forall\ i,\ k
\end{aligned}
\tag{4.9}
$$

where, $\mathbf{in} = [in_1(1)\cdots in_M(1)\cdots in_M(L)]^T$ and $C_i(k)$ (corresponding to $\sum_{j=1,\ j\neq i}^{M} p_j(k)h_{j,i}(k){\rho_{j,i}}^2$)

is the lower bound for $in_i(k)$. The Lagrangian of the proposed resource allocation problem (Eq. (4.9)) can be written as

$$\begin{aligned}
&\text{Determine} \quad [\mathbf{p}^T\ \mathbf{in}^T\ \mathbf{b}^T]^T \\
&\text{To Minimize:} \\
&\mathcal{L}_j = \tau_1 \sum_{k=1}^{L}\sum_{i=1}^{M} p_i(k) - \tau_2 \sum_{k=1}^{L}\sum_{i=1}^{M} b_i(k) \\
&+ \sum_{k=1}^{L} \lambda_1(k) \left(\sum_{i=1}^{M} p_i(k) h_{i,m}(k) - I_{th}(k) \right) \\
&- \sum_{k=1}^{L} \lambda_2(k) \left(\sum_{i=1}^{M} b_i(k) - R^u_{ch}(k) \right) + \sum_{k=1}^{L}\sum_{i=1}^{M} \nu_i^j(k)(in_i(k) - C_i(k)) \\
&\text{subject to} \\
&CD21: \ 0 \leq p_i(k) \leq p_i^{max}(k), \ \forall\ i,\ k; \\
&CD22: 1 \leq b_i(k) \leq b_i^{max}(k), \ \forall\ i,\ k; \\
&CD23: \ -p_i(k) h_{i,i}(k) \leq -C_{qarg}(in_i(k) + \sigma^2(k)) \\
&\times (2^{b_i(k)} - 1), \ \forall\ i,\ k; \\
&CD24: \ \sum_{k=1}^{L} b_i(k) \geq R_i^l, \ \forall\ i; \\
&CD25: \ in_i(k) \geq C_i(k), \ \forall\ i,\ k.
\end{aligned} \tag{4.10}$$

Here, $\lambda_1(k)$, $\lambda_2(k)$ and $\nu_i^j(k)$ are dual variables. Note that in $CD25$, $C_i(k)$ corresponds to $\sum_{j=1,\ j\neq i}^{M} p_j(k) h_{j,i}(k) {\rho_{j,i}}^2$, and is therefore dependent on allocated power for other users. Within a resource allocation period, $p_j(k)$ for other users may change. Therefore, taking a conservative approach, we set $C_i(k)$ as a lower bound for the variable $in_i(k)$ that represents the expected interference experienced by user i from other users in the channel k. In a sense, the lower bound reflects an optimistic guess at what the interference may be at the next iteration. It is prudent for us to consider this as a lower bound as the actual interference

could be higher. Rearranging (4.10) results in

$$\begin{aligned}
&\text{Determine} \quad [\mathbf{p}^T\ \mathbf{in}^T\ \mathbf{b}^T]^T \\
&\text{To Minimize:} \\
&\mathcal{L}_j = \sum_{i=1}^{M}\sum_{k=1}^{L}\left(\tau_1 + \lambda_1(k)h_{i,m}(k)\right)p_i(k) + \sum_{i=1}^{M}\sum_{k=1}^{L}\nu_i^j(k)in_i(k) \\
&-\sum_{i=1}^{M}\sum_{k=1}^{L}\left(\tau_2 + \lambda_2(k)\right)b_i(k) - \sum_{k=1}^{L}\lambda_1(k)I_{th}(k) - \sum_{i=1}^{M}\sum_{k=1}^{L}\nu_i^j(k)C_i(k) \\
&+\sum_{k=1}^{L}\lambda_2(k)R_{ch}^u(k) \\
&\text{subject to} \\
&CD21,\ CD22,\ CD23,\ CD24,\ CD25. \qquad (4.11)
\end{aligned}$$

Now, we can easily decompose the resource allocation problem (4.11) into M subproblems. Based on how we model the impact of $in_i(k)$ in each of the subproblems, three different decomposed formulations of (4.11) can be derived.

CASE 1: For the scenario when $in_i(k)$ is assumed constant but measurable, each of the subproblems can be written as

$$\begin{aligned}
&\text{Determine} \quad [\mathbf{p}_i^T\ \mathbf{b}_i^T]^T \\
&\text{To Minimize:} \quad g_i(\mathbf{p}_i,\ \mathbf{b}_i,\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2) = \sum_{k=1}^{L}\left(\tau_1 + \lambda_1(k)h_{i,m}(k)\right)p_i(k) - \sum_{k=1}^{L}\left(\tau_2 + \lambda_2(k)\right)b_i(k) \\
&\text{subject to} \\
&CDL21:\ 0 \leq p_i(k) \leq p_i^{max}(k),\ \forall\ k; \\
&CDL22: 1 \leq b_i(k) \leq b_i^{max}(k),\ \forall\ k\ nonumber \qquad (4.12) \\
&CDL23:\ -p_i(k)h_{i,i}(k) \leq -C_{qarg}(in_i(k) + \sigma^2(k))(2^{b_i(k)} - 1),\ \forall\ k; \\
&CDL24:\ \sum_{k=1}^{L} b_i(k) \geq R_i^l; \qquad (4.13)
\end{aligned}$$

where,

$$\mathbf{p}_i = [p_i(1)\ p_i(2) \cdots p_i(L)]^T, \tag{4.14}$$

$$\mathbf{b}_i = [b_i(1)\ b_i(2) \cdots b_i(L)]^T, \tag{4.15}$$

$$\boldsymbol{\lambda}_1 = [\lambda_1(1)\ \lambda_1(2) \cdots \lambda_1(L)]^T, \tag{4.16}$$

$$\boldsymbol{\lambda}_2 = [\lambda_2(1)\ \lambda_2(2) \cdots \lambda_2(L)]^T, \tag{4.17}$$

$$in_i(k) = \sum_{j=1,\ j\neq i}^{M} p_j(k) h_{j,i}(k) {\rho_{j,i}}^2. \tag{4.18}$$

Here, $\mathbf{p}_i$ and $\mathbf{b}_i$ are the transmit power and rate across different channels, respectively; and $g_i(\mathbf{p}_i,\ \mathbf{b}_i,\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ is the Lagrangian function for user i. The corresponding master dual problem is

$$\begin{aligned}
&\text{Determine} && [\boldsymbol{\lambda}_1^T\ \boldsymbol{\lambda}_2^T]^T \\
&\text{To Minimize:} && D_{j,1}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2) = \sum_{i=1}^{M} g_i(\mathbf{p}_i,\ \mathbf{b}_i,\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2) - \sum_{k=1}^{L} \lambda_1(k) I_{th}(k) + \sum_{k=1}^{L} \lambda_2(k) R_{ch}^u(k) \\
&\text{subject to} && \\
& && \lambda_1(k) \geq 0,\ \forall\ k; \\
& && \lambda_2(k) \geq -\tau_2,\ \forall\ k.
\end{aligned} \tag{4.19}$$

The pseudo code for user-based distributed joint power and rate allocation algorithm using the gradient projection method (where the primal variables and dual variables are adjusted in the opposite direction to the gradient $\nabla D_{j,1}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$) for CASE 1 can be summarized in Algorithm 4.1. Dual variables $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are initialized. At the beginning of an iteration, each user measures the interference $\mathbf{in}_i$ (equals to $[in_i(1),\ in_i(2), \cdots,\ in_i(L)]^T$). Then, each user executes the corresponding resource allocation subproblem to compute transmit power and rate for all its channels. The corresponding dual variables $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are updated. Each user continues to do (1) measure $\mathbf{in}_i$, (2) solve subproblem and (3) update the dual variables until the desired QoS is achieved and all system constraints (C23 and C24) are satisfied. In Algorithm 4.1, t is the iteration counter, $\alpha_{j,1}$ and $\alpha_{j,2}$ are sufficiently small positive step size, used to evolve the Lagrange multipliers.

Algorithm 4.1: Dual Algorithm to solve (4.11) based on **CASE 1**

Initialization: $\boldsymbol{p}(\mathbf{0}),\ \boldsymbol{\lambda}_1(\mathbf{0}),\ \boldsymbol{\lambda}_2(\mathbf{0})$;
while termination criterion is not true **do**
▷ *% Execute subproblems*
for $i = 1,\ 2, \cdots,\ M$ **do**
Measure $\mathbf{in}_i$;
Solve optimization subproblem (4.13) for $\boldsymbol{p}_i^t$ and for $\boldsymbol{b}_i^t$;
end for
▷ *% Update* $\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2$
for $k = 1,\ 2, \cdots,\ L$ **do**
if $(\sum_{i=1}^{M} p_i(k)h_{i,m}(k) > I_{th}(k))$ **then**
$\lambda_1^{t+1}(k) = [\lambda_1^t(k) - \alpha_{j,1}(-\sum_{i=1}^{M} p_i(k)h_{i,m}(k) + I_{th}(k))]$;
else
$\lambda_1^{t+1}(k) = \lambda_1^t(k)$;
end if
if $(\sum_{i=1}^{M} b_i(k) > R_{ch}^u(k))$ **then**
$\lambda_2^{t+1}(k) = min[-\tau_2,\ \lambda_2^t(k) + \alpha_{j,2}(-\sum_{i=1}^{M} b_i(k) + R_{ch}^u(k))]$;
else
$\lambda_2^{t+1}(k) = \lambda_2^t(k)$;
end if
end for
end while

The power and rate sequences generated from Algorithm 4.1 converge to the optimal power and rate solution. This implies that $D_{j,1}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ is a Lipschitz function which guarantees the convergence of gradient projection algorithms [90].

Lemma 4.2.1. *(Lipschitz Continuity of $D_{j,1}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$) The function $D_{j,1}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ is differentiable and there exists a constant $K > 0$ such that,*

$$||D_{j,1}(\boldsymbol{\lambda}_{a,1},\ \boldsymbol{\lambda}_{a,2}) - D_{j,1}(\boldsymbol{\lambda}_{b,1},\ \boldsymbol{\lambda}_{b,2}))||_2 \leq ||K[\boldsymbol{\lambda}_{a,1}^T\ \boldsymbol{\lambda}_{a,2}^T]^T - [\boldsymbol{\lambda}_{b,1}^T\ \boldsymbol{\lambda}_{b,2}^T]^T||_2$$
$$\forall \boldsymbol{\lambda}_{a,1},\ \boldsymbol{\lambda}_{a,2},\ \boldsymbol{\lambda}_{b,1},\ \boldsymbol{\lambda}_{b,2}\ \in\ \Re_+^L. \tag{4.20}$$

Proposition 4.2.1. *(Convergence of Algorithm 4.1) If $D_{j,1}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ satisfies Lemma 4.2.1 and $0 < \alpha_{j,1},\ \alpha_{j,2} < 2/K$, then starting with any power $\mathbf{0} \leq \boldsymbol{p}_i(\mathbf{0}) \leq \boldsymbol{p}_i^{max}\ \forall\ i$ and dual variables $\boldsymbol{\lambda}_1(\mathbf{0}),\ \boldsymbol{\lambda}_2(\mathbf{0}) \geq \mathbf{0}$, each point $(\boldsymbol{p},\ \boldsymbol{b},\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ of sequence $(\boldsymbol{p}^t,\ \boldsymbol{b}^t,\ \boldsymbol{\lambda}_1^t,\ \boldsymbol{\lambda}_2^t)$ generated*

by algorithm is converged. The details on convergence of gradient projection algorithms can be found in [80, 91].

It is important to note that the distributed approach does not fully avoid central control. This is due to the requirement of updating dual variables $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. The dual variables capture information regarding how well the interference temperature threshold constraints and upper bound of maximum rate supporting capability (of the channel) constraints are being satisfied. If the interference temperature threshold constraint is violated, then the corresponding dual variable increases in magnitude. This increase forces the objective function in our resource allocation problem to increase. To counter this effect, the resource allocation variables (power of users) are reduced which in turn improves the ability of satisfying the interference temperature threshold constraint (constraint C23). If the upper bound of channel maximum rate supporting capability constraint is not maintained, then the corresponding dual variable decreases in magnitude. This decrease causes the objective function in our resource allocation problem to increase. To counter this effect, the resource allocation variables (rate of users) are reduced which in turn improves the ability of satisfying the maximum rate constraint (constraint C24).

Finally, each user search in the neighborhood for the optimal discrete valued $b_i(k)$ (denoted as $\mathbf{b}_i^{opt}$) and optimal transmit power $\mathbf{p}_i^{opt}$ corresponds to $\mathbf{b}_i^{opt}$ is recalculated using equality of constraints $CD23$.

CASE 2: Consider the case when $in_i(k)$ is assumed a variable. However, in each iteration of distributed approach, a lower bound for this interference is measurable. In this

case, each of the subproblems can be written as,

$$\begin{aligned}
\text{Determine} \quad & [\mathbf{p}_i^T \ \mathbf{in}_i^T \ \mathbf{b}_i^T]^T \\
\text{To Minimize:} \quad & g_i(\mathbf{p}_i,\ \mathbf{in}_i,\ \mathbf{b}_i,\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2) = \sum_{k=1}^{L} \left(\tau_1 + \lambda_1(k) h_{i,m}(k)\right) p_i(k) \\
& - \sum_{k=1}^{L} \left(\tau_2 + \lambda_2(k)\right) b_i(k) \\
\text{subject to} \quad & \\
& CDL21,\ CDL22,\ CDL23,\ CDL24, \\
& CDL25:\ in_i(k) \geq C_i(k),\ \forall\ k,
\end{aligned} \tag{4.21}$$

where, $C_i(k)$ is the lower bound for $in_i(k)$ and equal to $\sum_{j=1,\ j\neq i}^{M} p_j(k) h_{j,i}(k) {\rho_{j,i}}^2$. The corresponding master dual problem is

$$\begin{aligned}
\text{Determine} \quad & [\boldsymbol{\lambda}_1^T \ \boldsymbol{\lambda}_2^T]^T \\
\text{To Minimize:} \quad & D_{j,2}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2) = \sum_{i=1}^{M} g_i(\mathbf{p}_i,\ \mathbf{in}_i,\ \mathbf{b}_i,\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2) \\
& - \sum_{k=1}^{L} \lambda_1(k) I_{th}(k) + \sum_{k=1}^{L} \lambda_2(k) R_{ch}^{u}(k) \\
\text{subject to} \quad & \\
& \lambda_1(k) \geq 0,\ \forall\ k, \\
& \lambda_2(k) \geq -\tau_2,\ \forall\ k.
\end{aligned} \tag{4.22}$$

The pseudo code for user-based distributed joint power and rate allocation algorithm using the gradient projection method (where the primal variables and dual variables are adjusted in the opposite direction to the gradient $\nabla D_{j,2}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$) for CASE 2 can be summarized in Algorithm 4.2. Dual variables $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are initialized. At the beginning of each iteration, each user measures the lower bound of expected interference from other users $\mathbf{C}_i$ (equals to $[C_i(1)\ C_i(2) \cdots C_i(L)]^T$). Then, each user executes the corresponding resource allocation subproblem to compute transmit power and rate for all its channels. The corresponding dual variables are updated. Each user continues to do (1) measure $\mathbf{C}_i$, (2) solve subproblem and

(3) update the dual variables until the desired QoS is achieved and all system constraints (C23 and C24) are satisfied. Here, $\alpha_{j,1}$ and $\alpha_{j,2}$ are as defined before.

The power and rate sequences generated from Algorithm 4.2 converge to the optimal power and rate solution. This implies that $D_{j,2}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ is a Lipschitz function which guarantees the convergence of gradient projection algorithms [90].

Lemma 4.2.2. *(Lipschitz Continuity of $D_{j,2}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$) The function $D_{j,2}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ is differentiable and there exists a constant $K > 0$ such that,*

$$||D_{j,2}(\boldsymbol{\lambda}_{a,1},\ \boldsymbol{\lambda}_{a,2}) - D_{j,2}(\boldsymbol{\lambda}_{b,1},\ \boldsymbol{\lambda}_{b,2}))||_2 \leq K||[\boldsymbol{\lambda}_{a,1}^T\ \boldsymbol{\lambda}_{a,2}^T]^T - [\boldsymbol{\lambda}_{b,1}^T\ \boldsymbol{\lambda}_{b,2}^T]^T||_2$$
$$\forall \boldsymbol{\lambda}_{a,1},\ \boldsymbol{\lambda}_{a,2},\ \boldsymbol{\lambda}_{b,1},\ \boldsymbol{\lambda}_{b,2}\ \in\ \Re_+^L. \tag{4.23}$$

Proposition 4.2.2. *(Convergence of Algorithm 4.2) If $D_{j,2}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ satisfies Lemma 4.2.2 and $0 < \alpha_{j,1},\ \alpha_{j,2} < 2/K$, then starting with any power $\mathbf{0} \leq \boldsymbol{p}_i(\mathbf{0}) \leq \boldsymbol{p}_i^{max}\ \forall\ i$ and dual variables $\boldsymbol{\lambda}_1(\mathbf{0}),\ \boldsymbol{\lambda}_2(\mathbf{0}) \geq \mathbf{0}$, each point $(\boldsymbol{p},\ \boldsymbol{b},\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2)$ of sequence $(\boldsymbol{p}^t,\ \boldsymbol{b}^t,\ \boldsymbol{\lambda}_1^t,\ \boldsymbol{\lambda}_2^t)$ generated by algorithm is converged.*

Finally, each user search in the neighborhood for the optimal discrete valued $b_i(k)$ (denoted as $\mathbf{b}_i^{opt}$) and optimal transmit power $\mathbf{p}_i^{opt}$ corresponds to $\mathbf{b}_i^{opt}$ is recalculated using equality of constraints $CD23$.

Algorithm 4.2: Dual Algorithm to solve (4.11) based on **CASE 2**

Initialization: $\boldsymbol{p}(\mathbf{0}),\ \boldsymbol{\lambda}_1(\mathbf{0}),\ \boldsymbol{\lambda}_2(\mathbf{0})$;
while termination criterion is not true **do**
▷ *% Execute subproblems*
for $i = 1,\ 2, \cdots,\ M$ **do**
Measure $\mathbf{C}_i$;
Solve optimization subproblem (4.21) for $\boldsymbol{p}_i^t$ and for $\boldsymbol{b}_i^t$;
end for
▷ *% Update* $\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2$
for $k = 1,\ 2, \cdots,\ L$ **do**
if $(\sum_{i=1}^{M} p_i(k)h_{i,m}(k) > I_{th}(k))$ **then**
$\lambda_1^{t+1}(k) = [\lambda_1^t(k) - \alpha_{j,1}(-\sum_{i=1}^{M} p_i(k)h_{i,m}(k) + I_{th}(k))]$;
else
$\lambda_1^{t+1}(k) = \lambda_1^t(k)$;
end if
if $(\sum_{i=1}^{M} b_i(k) > R_{ch}^u(k))$ **then**
$\lambda_2^{t+1}(k) = min[-\tau_2,\ \lambda_2^t(k) + \alpha_{j,2}(-\sum_{i=1}^{M} b_i(k) + R_{ch}^u(k))]$;
else
$\lambda_2^{t+1}(k) = \lambda_2^t(k)$;
end if
end for
end while

CASE 3: An alternate formulation can be derived by absorbing the constraint $CD25$ in the objective function. The subproblems for this case can be formulated as

Determine $[\mathbf{p}_i^T\ \mathbf{in}_i^T\ \mathbf{b}_i^T]^T$

To Minimize:

$$g_i(\mathbf{p}_i,\ \mathbf{in}_i,\ \mathbf{b}_i,\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j) = \sum_{k=1}^{L} (\tau_1 + \lambda_1(k)h_{i,m}(k))\, p_i(k)$$
$$+ \sum_{k=1}^{L} \nu_i^j(k) in_i(k) - \sum_{k=1}^{L} (\tau_2 + \lambda_2(k))\, b_i(k)$$

subject to

$$CDL21,\ CDL22,\ CDL23,\ CDL24,\ CDL25. \tag{4.24}$$

The corresponding master dual problem is

$$
\begin{aligned}
\text{Determine} \quad & [\boldsymbol{\lambda}_1^T \ \boldsymbol{\lambda}_2^T \ \boldsymbol{\nu}_1^{j,T} \ \cdots \ \boldsymbol{\nu}_i^{j,T} \ \cdots \ \boldsymbol{\nu}_M^{j,T}]^T \\
\text{To Minimize:} \quad & D_{j,3}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j \ \forall\ i) = \sum_{i=1}^{M} g_i(\mathbf{p}_i,\ \mathbf{in}_i,\ \mathbf{b}_i,\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j) \\
& - \sum_{k=1}^{L} \lambda_1(k) I_{th}(k) + \sum_{k=1}^{L} \lambda_2(k) R_{ch}^u(k) \\
& - \sum_{k=1}^{L} \sum_{i=1}^{M} \nu_i^j(k) C_i(k) \\
\text{subject to} \quad & \\
& \lambda_1(k) \geq 0,\ \forall\ k, \\
& \lambda_2(k) \geq -\tau_2,\ \forall\ k, \\
& \nu_i^j(k) \geq 0,\ \forall\ i,\ k; \qquad (4.25)
\end{aligned}
$$

where,

$$
\boldsymbol{\nu}_i^j = [\nu_i^j(1)\ \nu_i^j(2) \cdots \nu_i^j(L)]^T. \qquad (4.26)
$$

The pseudo code for user-based distributed joint power and rate allocation algorithm using the gradient projection method (where the primal variables and dual variables are adjusted in the opposite direction to the gradient $\nabla D_{j,3}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j \ \forall\ i)$) for CASE 3 can be summarized in Algorithm 4.3. Dual variables $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$ and $\boldsymbol{\nu}_i^j \ \forall\ i$ are initialized. At the beginning of each iteration, each user measures the lower bound of expected interference from other users $\mathbf{C}_i$ (equals to $[C_i(1)\ C_i(2) \cdots C_i(L)]^T$). Then, each user executes the corresponding resource allocation subproblem to compute transmit power and rate for all its channels. The corresponding dual variables are updated. Each user continues to do (1) measure $\mathbf{C}_i$, (2) solve subproblem and (3) update the dual variables until the desired QoS is achieved and all system constraints (C23 and C24) are satisfied. In Algorithm 4.3, $\alpha_{j,1}$ and $\alpha_{j,2}$ are as defined before; β_j is a sufficiently small positive step size, and $[\cdot]^+$ denotes the projection onto nonnegative orthant.

Algorithm 4.3: Dual Algorithm to solve (4.11) based on **CASE 3**

Initialization: $\boldsymbol{p}(\mathbf{0}),\ \boldsymbol{\lambda}_1(\mathbf{0}),\ \boldsymbol{\lambda}_2(\mathbf{0})$;
Initialization: $\boldsymbol{\nu}_1^j(\mathbf{0}), \cdots,\ \boldsymbol{\nu}_i^j(\mathbf{0}), \cdots,\ \boldsymbol{\nu}_M^j(\mathbf{0})$;
Measure $\mathbf{C}_1, \cdots,\ \mathbf{C}_i, \cdots,\ \mathbf{C}_M$;
while termination criterion is not true **do**
▷ *% Execute subproblems*
for $i = 1,\ 2, \cdots,\ M$ **do**
Solve optimization subproblem (4.24) for $\boldsymbol{p}_i^t$ and for $\boldsymbol{b}_i^t$;
end for
▷ *% Update* $\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2$
for $k = 1,\ 2, \cdots,\ L$ **do**
if $(\sum_{i=1}^{M} p_i(k)h_{i,m}(k) > I_{th}(k))$ **then**
$\lambda_1^{t+1}(k) = [\lambda_1^t(k) - \alpha_{j,1}(-\sum_{i=1}^{M} p_i(k)h_{i,m}(k) + I_{th}(k))]$;
else
$\lambda_1^{t+1}(k) = \lambda_1^t(k)$;
end if
if $(\sum_{i=1}^{M} b_i(k) > R_{ch}^u(k))$ **then**
$\lambda_2^{t+1}(k) = min[-\tau_2, \lambda_2^t(k) + \alpha_{j,2}(-\sum_{i=1}^{M} b_i(k) + R_{ch}^u(k))]$;
else
$\lambda_2^{t+1}(k) = \lambda_2^t(k)$;
end if
end for
▷ *% Update* $\boldsymbol{\nu}_1^j, \cdots,\ \boldsymbol{\nu}_i^j, \cdots,\ \boldsymbol{\nu}_M^j$
for $i = 1,\ 2, \cdots,\ M$ **do**
Measure $\mathbf{C}_i$;
for $k = 1,\ 2, \cdots,\ L$ **do**
$\nu_i^{j,t+1}(k) = [\nu_i^{j,t}(k) - \beta_j(-in_i(k) + C_i(k))]^+$;
end for
end for
end while

The power and rate sequences generated from Algorithm 4.3 converge to the optimal power and rate solution. This implies that $D_{j,3}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j\ \forall\ i)$ is a Lipschitz function which guarantees the convergence of gradient projection algorithms [90].

Lemma 4.2.3. *(Lipschitz Continuity of* $D_{j,3}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j\ \forall\ i)$*) The function* $D_{j,3}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j\ \forall\ i)$

is differentiable and there exists a constant $K > 0$ *such that,*

$$||D_{j,3}(\boldsymbol{\lambda}_{a,1},\ \boldsymbol{\lambda}_{a,2},\ \boldsymbol{\nu}_i^j\ \forall\ i) - D_{j,3}(\boldsymbol{\lambda}_{b,1},\ \boldsymbol{\lambda}_{b,2},\ \boldsymbol{\nu}_i^j\ \forall\ i))||_2 \leq$$
$$K||[\boldsymbol{\lambda}_{a,1}^T\ \boldsymbol{\lambda}_{a,2}^T\boldsymbol{\nu}_{a,i}^{j,T}\ \forall\ i]^T - [\boldsymbol{\lambda}_{b,1}^T\ \boldsymbol{\lambda}_{b,2}^T,\ \boldsymbol{\nu}_{b,i}^{j,T}\ \forall\ i]^T||_2$$
$$\forall \boldsymbol{\lambda}_{a,1},\ \boldsymbol{\lambda}_{a,2},\ \boldsymbol{\lambda}_{b,1},\ \boldsymbol{\lambda}_{b,2},\ \boldsymbol{\nu}_{a,i}^j\ \forall\ i,\ \boldsymbol{\nu}_{b,i}^j\ \forall\ i\ \in\ \Re_+^L. \tag{4.27}$$

Proposition 4.2.3. *(Convergence of Algorithm 4.3) If* $D_{j,3}(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j\ \forall\ i)$ *satisfies Lemma 4.2.3 and* $0 < \alpha_{j,1},\ \alpha_{j,2},\ \beta_j < 2/K$*, then starting with any power* $\mathbf{0} \leq \boldsymbol{p}_i(\mathbf{0}) \leq \boldsymbol{p}_i^{max}\ \forall\ i$ *and dual variables* $\boldsymbol{\lambda}_1(\mathbf{0}),\ \boldsymbol{\lambda}_2(\mathbf{0}),\ \boldsymbol{\nu}_i^j(\mathbf{0})\ \forall\ i \geq \mathbf{0}$*, each point* $(\boldsymbol{p},\ \boldsymbol{b},\ \boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \boldsymbol{\nu}_i^j\ \forall\ i)$ *of sequence* $(\boldsymbol{p}^t,\ \boldsymbol{b}^t,\ \boldsymbol{\lambda}_1^t,\ \boldsymbol{\lambda}_2^t,\ \boldsymbol{\nu}_i^{j,t}\ \forall\ i)$ *generated by algorithm is converged.*

Finally, each user search in the neighborhood for the optimal discrete valued $b_i(k)$ (denoted as $\mathbf{b}_i^{opt}$) and optimal transmit power $\mathbf{p}_i^{opt}$ corresponds to $\mathbf{b}_i^{opt}$ is recalculated using equality of constraints $CD23$.

In summary, based on a priori information or ability to measure interference power, we can formulate different versions of distributed implementation of the proposed resource allocation problem. It is also important to note that initializing dual variables and choice of step sizes are critical for convergence speed of the distributed solution [80, 91].

From an exchange of information standpoint, the distributed approach is more attractive than a centralized scheme (depending on the number of iterations, I). This is because, the centralized scheme requires information about all users and channels in the network. The required amount of information exchange in centralized scheme is $O(M^2)$. As a result, it incurs a high communication overhead and poor scalability in CRN with large number of SUs. In the proposed distributed resource allocation framework, every SU requires local information along with knowledge of dual variables, $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. The required amount of information exchange in each of the distributed cases is $O(M)$. Thus, the distributed implementation needs minimal communication overhead making it more attractive than a centralized scheme.

4.3 Numerical Results

In this section, we quantify the performance of the proposed joint resource allocation framework. We assume a CRN with $L = 11$ available channels and a total of $M = 10$ secondary users. We assume a usage pattern as shown in Table 4.1, where a 1 indicates that the corresponding channel is being used by the SU. Table 4.2 provides information on the channel quality for all L channels. Table 4.3 lists the minimum rate requirement for each SU. Finally, Table 4.4 contains all other system parameters that are relevant to our resource allocation framework. Based on all this information, our objective is to find the optimal transmit power and rate that each of the M SUs should employ to guarantee their QoS.

Table 4.1: *Usage pattern across channels.*

Channel, k	1	2	3	4	5	6	7	8	9	10	11
User, 1	1	0	1	0	0	1	1	0	0	0	1
User, 2	1	1	1	0	0	1	1	0	1	0	1
User, 3	0	1	1	0	1	1	0	0	1	0	1
User, 4	0	1	1	0	1	1	1	0	1	1	1
User, 5	0	1	1	1	1	1	1	1	1	1	1
User, 6	0	1	1	1	1	0	1	1	1	1	1
User, 7	0	1	0	1	1	0	1	1	0	1	1
User, 8	1	1	1	0	1	0	1	1	0	1	1
User, 9	1	0	1	0	1	0	1	1	0	1	1
User, 10	1	0	1	0	1	1	0	1	0	1	1

Table 4.2: *Channel quality parameters.*

Channel, k	1	2	3	4	5	6	7	8	9	10	11
$\sigma^2(k)$, $(\times 10^{-3})$	5	4	3	2	2.5	6	4	4	5	3.5	4.5

As the joint resource allocation problem has one non-linear constraint (constraint $C27$), we use "Sequential Quadratic Programming (SQP)" method to solve this problem. SQP method has been briefly discussed in Appendix B. Figure 4.1 presents the transmit power

Table 4.3: *Minimum rate requirement of users.*

User, u	1	2	3	4	5	6	7	8	9	10
R_i^l	3	8	4	12	9	7	14	5	10	8

Table 4.4: *System parameters.*

$p_i^{max}(k)\ \forall\ i,\ k$	5
$b_i^{max}(k)\ \forall\ i,\ k$	6
$p_{e,i}^{th}\ \forall\ i$	10^{-3}
$I_{th}(k)\ \forall\ k$	$200 \times \sigma^2(k)$
$R_{ch}^u(k)\ \forall\ k$	20
$\rho_{j,i}$	0.03125

and rate allocation across channels for users 7 and 10 assuming scalarization parameters τ_1 and τ_2 to 0.5. Here, user 7 operates on channels 2, 4, 5, 7, 8, 10 and 11; user 10 operates on channels 1, 3, 5, 6, 8, 10 and 11. As in the two stage formulation, we can see that "reverse water filling" effect is once again observed as users tend to use more transmit power in poor quality channels. Since, SINR is not a constraint in this formulation, the users do not attempt to maintain a constant SINR in the channels. However, QoS is maintained by adjusting the rate allocated to each channel; e.g., Fig. 4.1 shows that for high SINR channels, more bits/channel are allocated and vice versa. Figure 4.2 shows the allocation of total transmit power and total rate across users. Once again, the resource allocation engine is successful in meeting the rate requirements for all SUs. Finally, Fig. 4.3 illustrates the variation in total power and rate employed by user 1 as the number of interfering users increases. Unlike, the two stage resource allocation approach, the joint resource allocation solution reflects a different optimal strategy for user 1. From Fig. 4.3, it is clear that best strategy for user 1 is to reduce its rate (and therefore, transmit power) up to a point where it can barely satisfy its rate and BER requirement. In a sense, the solution reflects an "accommodating attitude" for all users until they are all functioning at a state where their

bare minimum requirements are met. One can visualize this as a socially optimal solution.

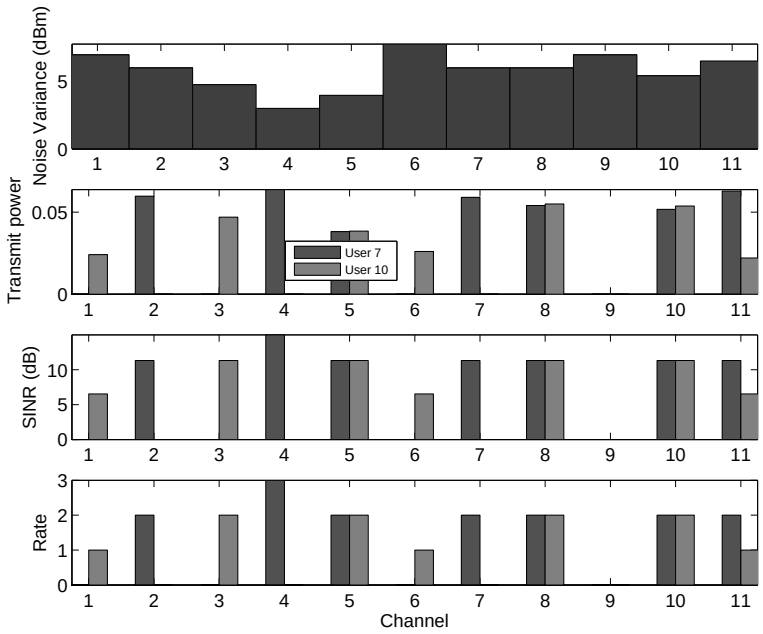

Figure 4.1: *Allocation of transmit power and rate with channel noise variance and SINR for users* 7 *and* 10 *($\tau_2/\tau_1 = 1$).*

Figure 4.4 compares the total transmit power and total rate across users obtained from three formulations of distributed approach (CASE 1, CASE 2 and CASE 3) with a centralized solution (for $\tau_2/\tau_1 = 0.20$). In each case, we initialize dual variables $\lambda_1(k)$ and $\lambda_2(k)$ to 0 for all channels. The step sizes $\alpha_{j,1}$ and $\alpha_{j,2}$ are set to 0.6 and 0.01, respectively. For CASE 3, $\nu_i^j(k)$ is initialized to 0 for all users and channels and the step size β_j is set to 0.1. From Fig. 4.4, we see that all three distributed formulations are successful in meeting minimum rate requirements of all active users. We also observe that the distributed solution obtained from CASE 2 matches that obtained from CASE 3. Whereas the solution from CASE 1 is different from CASE 2 and CASE 3. In terms of total rate, each of distributed formulations is inferior with respect to centralized formulation beyond a certain value of the scalarization ratio τ_2/τ_1. This is due to the impact of scalarization ratio τ_2/τ_1 on the decision variables (rate $b_i(k)$ and transmit power $p_i(k)$) in the objective function.

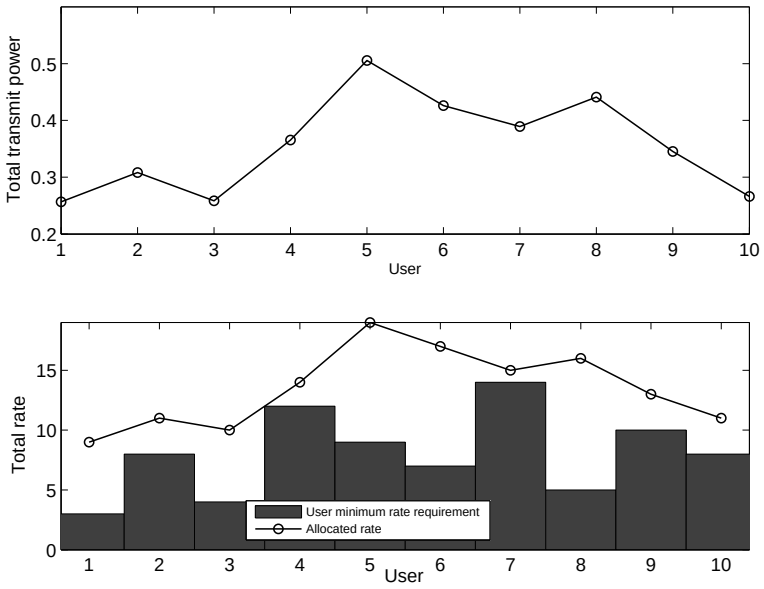

Figure 4.2: *Allocation of total transmit power and total rate across users ($\tau_2/\tau_1 = 1$).*

The centralized solution is obtained with this scalarization ratio as importance on rate and transmit power, respectively in each of the channels and for each user. In the distributed formulations, this scalarization ratio (0.2 in this example) is the starting importance on rate and transmit power, respectively, in intended channels for a user. With iteration, this importance changes and is determined by the evolution of dual variables (see Eqs. (4.19), (4.22) and (4.25); and Algorithms 4.1, 4.2 and 4.3). The solution obtained from the distributed approach depends on this importance factor at the terminating iteration. At the terminating iteration, if the importance factor is same in different distributed formulations then same solution is achieved in different formulations.

Figures 4.5 and 4.6 show the evolution of dual variables i.e., $\lambda_1(4)$ and $\lambda_2(4)$, correspond to measured interference temperature ($\sum_{i=1}^{M} p_i(k)h_{i,m}(k)$) and allocated rate ($\sum_{i=1}^{M} b_i(k)$) for channel 4 with iteration (CASE 2). Figures 4.7 and 4.8 show the evolution of dual variables i.e., $\lambda_1(8)$ and $\lambda_2(8)$, correspond to measured interference temperature ($\sum_{i=1}^{M} p_i(k)h_{i,m}(k)$) and allocated rate ($\sum_{i=1}^{M} b_i(k)$) for channel 8 with iteration (CASE 2). From Figs. 4.5 and 4.7, we see that at some iteration, if measured interference temperature ($\sum_{i=1}^{M} p_i(k)h_{i,m}(k)$)

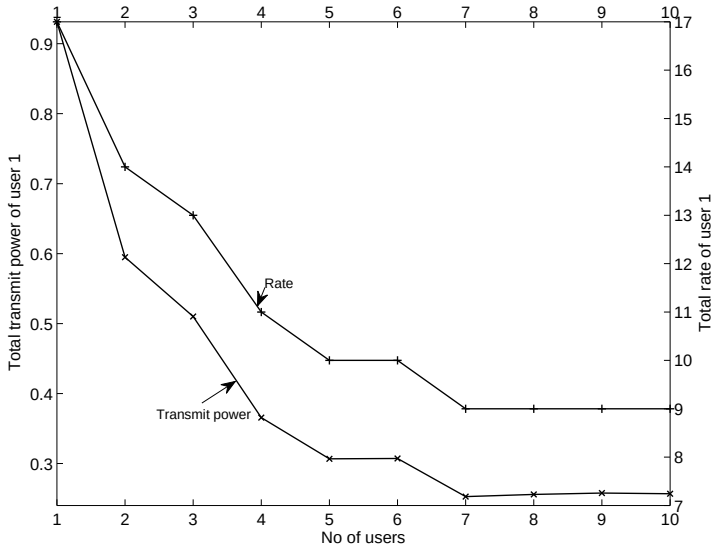

Figure 4.3: *Total transmit power and total rate for user* 1 *with number of users* ($\tau_2/\tau_1 = 1$).

in a channel is above the limit ($I_{th}(k)$) then corresponding dual variable ($\lambda_1(k)$) increases. This in turn causes an increase in importance on transmit power in the objective function and hence, in next iteration, transmit power in that channel is reduced or at best remains the same that results measured interference temperature getting closer to the limit. At some iteration, if measured interference temperature is within the limit, then dual variable is kept to its previous iteration value. From Figs. 4.6 and 4.8, we see that if allocated rate ($\sum_{i=1}^{M} b_i(k)$) in a channel is above the limit ($R_{ch}^{u}(k)$), then corresponding dual variable ($\lambda_2(k)$) decreases. This in turn causes an decrease in importance on rate in the objective function and hence, in next iteration, allocated rate in that channel is reduced or at best remains the same that forces allocated rate getting closer to the limit. If allocated rate is within the limit then dual variable is kept to its previous iteration value. This pattern of evolution of dual variables is consistent across other channels and for three distributed formulations.

Figure 4.9 shows the convergence time for proposed distributed approaches in terms of number of iterations required to satisfy system constraints ($C23$ and $C24$) for the given

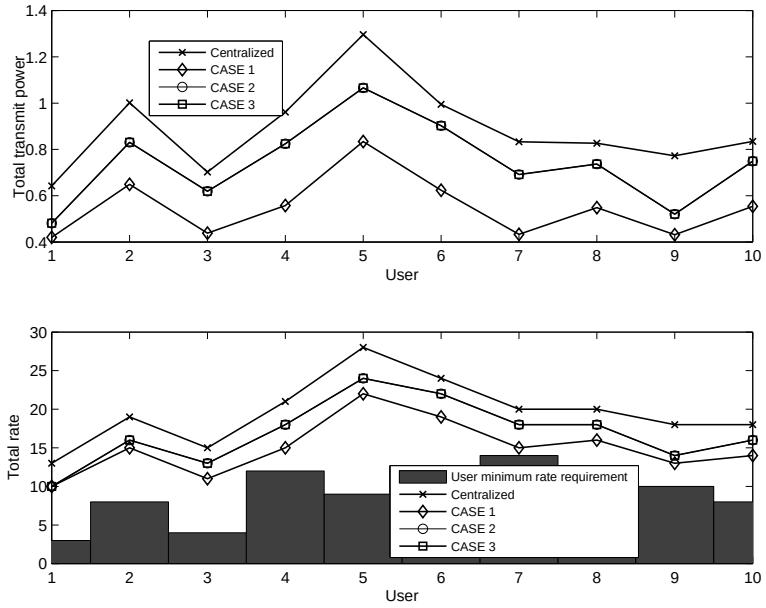

Figure 4.4: *Allocation of total transmit power and total rate across users from different formulations of distributed approach ($\tau_2/\tau_1 = 0.20$).*

simulation set up. From Fig. 4.9, we see CASE 1, CASE 2 or CASE 3 are comparable in convergence time. However, in terms of net transmission cost i.e., the numeric value of $\sum_{k=1}^{L}\sum_{i=1}^{M} p_i^{opt}(k)b_i^{opt}(k)$ (as shown in Table 4.5), CASE 1 is preferable than CASE 2 or CASE 3. The reason can be explained by looking at resource allocation problem formulation. In CASE 1, $in_i(k)$ (corresponding to $\sum_{j=1,\ j\neq i}^{M} p_j(k)h_{j,i}(k){\rho_{j,i}}^2$) is assumed as constant; and in CASE 2 or CASE 3, $in_i(k)$ is treated as a variable and is lower bounded by $C_i(k)$ (constraint $CDL25$). As power allocation is done based on the lower bound in CASE 2 or CASE 3, the optimal power/rate is higher than in CASE 1. From Fig. 4.9 and Table 4.5, we can conclude that CASE 2 and CASE 3 exhibits similar performances. From (4.21) and (4.24), we can see that CASE 2 is a special case of CASE 3 (when $\nu_i^j(k) = 0$). The presence of $\nu_i^j(k)$ into objective function (CASE 3) works as a importance factor for the variable $in_i(k)$. From Algorithm 4.3, we can see that at the end of an iteration, if $in_i(k)$ goes beyond the measured $C_i(k)$, then $\nu_i^j(k)$ goes up. This higher value of $\nu_i^j(k)$ results in a decrease of $in_i(k)$ at next iteration. However, since both CASE 2 and CASE 3 impose the same lower

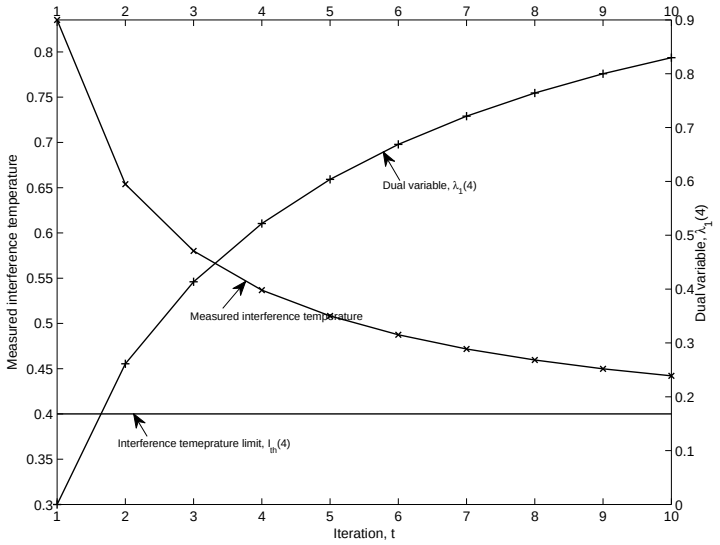

Figure 4.5: *Evolution of dual variable and measured interference temperature with iteration ($\tau_2/\tau_1 = 0.20$, Channel 4).*

Table 4.5: *Net Transmission Cost* $\left(\sum_{k=1}^{L}\sum_{i=1}^{M} p_i^{opt}(k)b_i^{opt}(k)\right)$ *for different cases*

	$\tau_2/\tau_1 = 0.20$	$\tau_2/\tau_1 = 0.05$	$\tau_2/\tau_1 = 0.00$
CASE 1	15.60	7.30	2.43
CASE 2	19.40	8.30	2.53
CASE 3	19.40	8.30	2.53

bound on $in_i(k)$, the resulting transmission cost is similar.

Next, we present a comparison between the two resource allocation frameworks (solved as centralized allocation problem) by evaluating their ability to minimize F_1 and maximize F_2. Table 4.6 demonstrates that while both schemes maximize rate to a comparable level, the total transmit power to maintain QoS is lower when the joint resource allocation approach is employed. This is because, the joint resource allocation approach provides the capability to adapt two variables (power and bits/channel use) simultaneously in order to achieve a certain BER. In the two stage resource allocation framework, only one of these variables is adapted in each stage resulting in a lower degree of freedom.

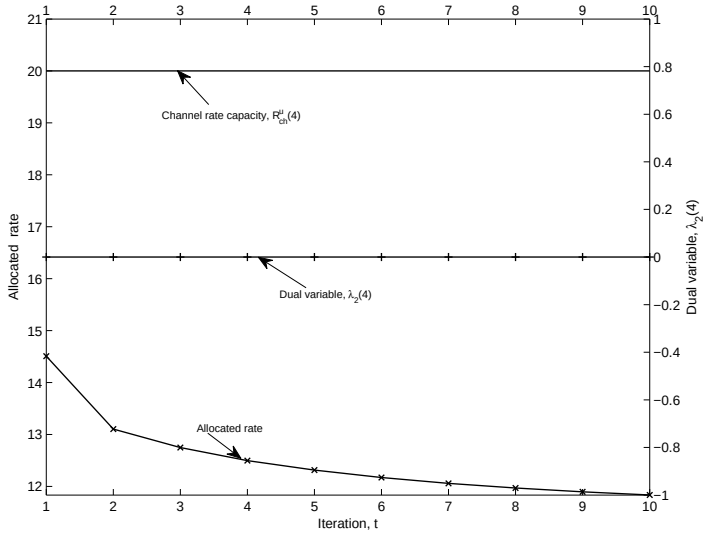

Figure 4.6: *Evolution of dual variable and allocated rate with iteration ($\tau_2/\tau_1 = 0.20$, Channel 4).*

Table 4.6: *Comparison between resource allocation frameworks.*

	Two-stage	Joint
Total power/bit, $\left(\sum_{k=1}^{L}\sum_{i=1}^{\widetilde{N}_s(k)} p_i^{opt}(k)\right)$	5.14	3.56
Total bit/channel use, $\left(\sum_{k=1}^{L}\sum_{i=1}^{\widetilde{N}_s(k)} b_i^{opt}(k)\right)$	148	135
Net Transmission cost, $\left(\sum_{k=1}^{L}\sum_{i=1}^{\widetilde{N}_s(k)} p_i^{opt}(k)b_i^{opt}(k)\right)$	10.27	6.93

4.4 Summary

In this chapter, we propose a joint resource allocation framework that provide the optimal transmit power and rate distribution that each SU needs to employ while maintaining QoS in a multi-channel CRN. We consider that a single channel can be used by multiple SUs and a single secondary user can simultaneously employ multiple channels. Simulation results illustrate that optimal transmit power follows reverse water filling process and optimal rate allocation is proportional to SINR. The solution obtained from proposed user-based distributed approaches follow the centralized solution. We also observe joint resource allo-

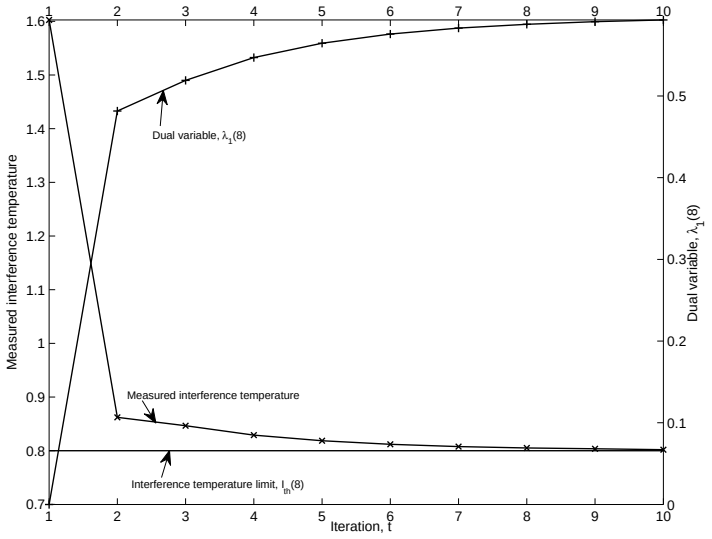

Figure 4.7: *Evolution of dual variable and measured interference temperature with iteration ($\tau_2/\tau_1 = 0.20$, Channel 8).*

cation of power and rate results in a more power efficient solution relative to a two stage resource allocation architecture. In the following chapters, we consider fairness and game theoretic distributed solution approach for the joint resource allocation problem.

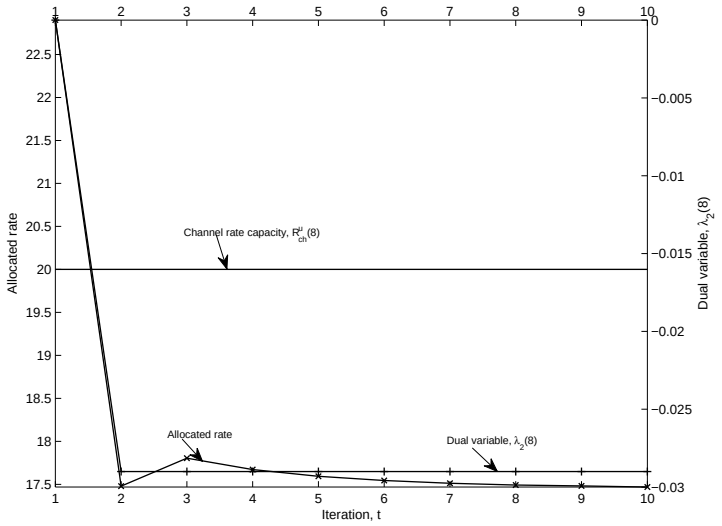

Figure 4.8: *Evolution of dual variable and allocated rate with iteration ($\tau_2/\tau_1 = 0.20$, Channel 8).*

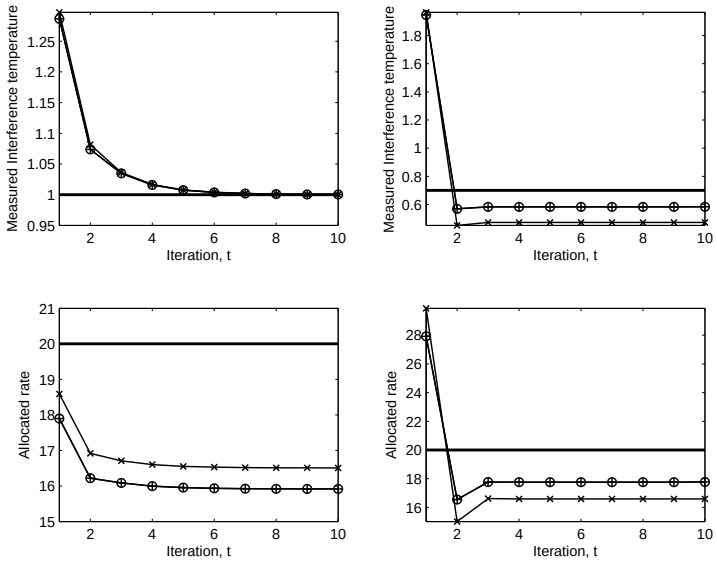

Figure 4.9: *Evolution of measured interference temperature and allocated rate with iteration ($\tau_2/\tau_1 = 0.20$, Channel 1 and 10).*

Chapter 5

Fairness in Resource Allocation

In chapter 4, we proposed a joint resource allocation framework to determine transmit power and rate for secondary users in a multi-channel multiuser competitive CRN. If the underlying optimization problem is convex, resource allocation is optimal. However, users may not be satisfied with optimal allocation of resources based on instantaneous QoS. An example of dissatisfaction among SUs may arise when two SUs with different minimum rate requirements are allocated the same rate. Another example of dissatisfaction among SUs may arise when a user is assigned higher average power per bit relative to other users. Typically dissatisfaction is a feeling that develops over time. Hence, fairness in terms of current and prior history of user satisfaction with respect to QoS in optimal resource allocation is an important consideration. In this chapter, we quantify user experience over time (time index is denoted by n) by introducing dynamic fairness weights for each SU in the resource allocation framework. The dynamics of the weights are governed by social behavioral models. We study the effect of Homo Egualis (HE), Homo Parochius (HP) and Homo Reciprocan (HR) models. We consider Jain system level fairness index [77] as a measure of fairness in resource allocation. We adopt the same system model introduced in chapter 4. Table 5.1 defines all relevant terms (at n-th time instant) used throughout the chapter.

The rest of the chapter is organized as follows. Section 5.1 describes the proposed resource allocation framework. Section 5.2 shows the analogy between the social behavior of human beings and that of SUs in CRN, and introduces the society models of interest to

Table 5.1: *Notations*

$\sigma^2(n,k)$	Noise variance in k-th channel
$\rho_{j,i}(n)$	Orthogonality factor between users j and i
$h_{i,i}(n,k)$	Power gain from i-th transmitter to i-th receiver in k-th channel
$h_{i,m}(n,k)$	Power gain from i-th transmitter at location m in k-th channel
$p_i(n,k)$	Transmit power per bit of i-th user in k-th channel
$p_i^{max}(n,k)$	Maximum transmit power per bit of i-th user in k-th channel
$I_{th}(n,k)$	Interference temperature constraint in k-th channel
$b_i(n,k)$	Rate of i-th user in k-th channel
$b_i^{max}(n,k)$	Maximum rate of i-th user in k-th channel
$R_{ch}^u(n,k)$	Maximum rate supported by k-th channel
$R_i^l(n)$	Minimum required rate for i-th user
$p_{e,i}(n,k)$	BER for i-th user in k-th channel
$p_{e,i}^{th}(n)$	BER threshold at receiver for i-th user in any channel
$\gamma_i(n,k)$	SINR per bit for i-th user in k-th channel

this chapter. Modeling fairness is described in Sec. 5.3. Numerical results are presented in Sec. 5.4. Finally, Sec. 5.5 summarizes the chapter.

5.1 Resource Allocation Framework

The objectives of the resource allocation framework are to (1) minimize the total transmit power, and (2) maximize the total rate while satisfying the QoS requirements and maintaining fairness across all active SUs. The mathematical description of the bi-objective resource

allocation scheme corresponds to:

$$
\begin{aligned}
\text{Determine} \quad & [\mathbf{p}^T(n)\ \mathbf{b}^T(n)]^T \\
\text{where,} \quad & \mathbf{p}(n) = [p_1(n,1)\cdots\ p_M(n,1)\cdots\ p_1(n,L)\cdots\ p_M(n,L)]^T \text{ and} \\
& \mathbf{b}(n) = [b_1(n,1)\cdots\ b_M(n,1)\cdots\ b_1(n,L)\cdots\ b_M(n,L)]^T \\
\text{To Minimize:} \quad & F_1 = \sum_{k=1}^{L}\sum_{i=1}^{M} w_i^p(n) p_i(n,k) \text{ and} \\
\text{Maximize:} \quad & F_2 = \sum_{k=1}^{L}\sum_{i=1}^{M} w_i^b(n) b_i(n,k) \\
\text{subject to} & \\
& C1:\ 0 \leq p_i(n,k) \leq p_i^{max}(n,k),\ \forall\ i,\ k; \\
& C2:\ b_i(n,k) \in [1,\cdots,\ b_i^{max}(n,k)],\ \forall\ i,\ k; \\
& C3:\ \sum_{i=1}^{M} p_i(n,k) h_{i,m}(n,k) \leq I_{th}(n,k),\ \forall\ k; \\
& C4:\ \sum_{i=1}^{M} b_i(n,k) \leq R_{ch}^u(n,k),\ \forall\ k; \\
& C5:\ \sum_{k=1}^{L} b_i(n,k) \geq R_i^l(n),\ \forall\ i; \\
& C6:\ p_{e,i}(n,k) \leq p_{e,i}^{th}(n),\ \forall\ i,\ k.
\end{aligned} \tag{5.1}
$$

where

$$
p_{e,i}(n,k) \leq \frac{4}{b_i(n,k)} Q\left(\sqrt{\frac{3b_i(n,k)\gamma_i(n,k)}{(2^{b_i(n,k)}-1)}}\right),\ \forall\ i,\ k,\ \text{odd } b_i(n,k); \tag{5.2}
$$

$$
p_{e,i}(n,k) = \frac{4}{b_i(n,k)} \left(1 - 2^{-\frac{b_i(n,k)}{2}}\right) Q\left(\sqrt{\frac{3b_i(n,k)\gamma_i(n,k)}{(2^{b_i(n,k)}-1)}}\right),\ \forall\ i,\ k,\ \text{even } b_i(n,k); \tag{5.3}
$$

$$
\gamma_i(n,k) = \frac{p_i(n,k) h_{i,i}(n,k)}{\sum_{j=1,\ j\neq i}^{M} p_j(n,k) h_{j,i}(n,k) \rho_{j,i}^2(n) + \sigma^2(n,k)},\ \forall\ i,\ k. \tag{5.4}
$$

Here, $w_i^p(n)$ and $w_i^b(n)$ are the dynamic fairness weights for user i based on allocation of transmit power and rate till time instant $(n-1)$, respectively; constraints $C1$ and $C2$

indicate limits on transmit power and rate, respectively; $C3$ indicates the interference temperature constraint; $C4$ indicates the total rate supported by a channel; $C5$ represents the required rate of users and finally $C6$ is QoS/BER constraint. Since $b_i(n,k)$ is discrete and constraint $C6$ is nonlinear, the optimization formulation presented above is a constrained multi-objective mixed integer nonlinear programming (multi-objective MINLP) resource allocation scheme, which is NP-hard in general. Relaxing the integer constraint on rate, $b_i(n,k)$ (as assumed in [93]) and assuming $b_i(n,k)$ as continuous variable, the above resource allocation scheme can be restated with $C2$ as:

$$C2: 1 \leq b_i(k) \leq b_i^{max}(k),\ \forall\ i,\ k. \tag{5.5}$$

As in chapter 3, constraint $C6$ can be written as

$$C7: \ -\gamma_i(n,k) \leq -C_{qarg}(n)(2^{b_i(n,k)}-1),\ \forall\ i,\ k; \tag{5.6}$$

where, $C_{qarg}(n)$ is a constant and is also determined following the analysis in chapter 3. The resource allocation scheme with combined single objective can be rewritten as:

$$\begin{aligned} &\text{Minimize } \tau_1 F_1 - \tau_2 F_2 \\ &\text{subject to} \\ &C1,\ C2,\ C3,\ C4,\ C5,\ C7. \end{aligned} \tag{5.7}$$

The parameters τ_1 and τ_2 in the combined objective function are the scalarization factors and can be set following the discussion in [94]. Finally, we use the solution obtained from the convex formulation (Eq. (5.7)) as a starting point to search in the neighborhood for the optimal discrete valued $b_i(n,k)$ (denoted as $\mathbf{b^{opt}}$). Based on the new discrete solution, the optimal transmit power $\mathbf{p^{opt}}$ is recalculated using Eq. (5.6).

In the following section, we describe the analogy between the social behavior of human beings and that of SUs in CRN. We then use the society models to design the evolution models for the fairness weights $w_i^p(n)$ and $w_i^b(n)$.

5.2 Human Society Model and Cognitive Radio Networks

Behavior of human beings in a society can be categorized as individual or group. Self-interest, rationality or irrationality are the notions of individual behavior. Grouping for some public good (survival, fairness) is an example of group behavior. The secondary users in a CRN may behave rationally while competing and cooperating for resources, survival and social efficiency just like human beings in society [95]. Hence, secondary users behavior in CRN can be modeled based on human society model. HE society model, HP society model and HR society model are some examples of society models. In this work, our objective is to define effective instantaneous fairness weight for each user in the network and to develop an evolution model for fairness weight based on present and past user experiences with respect to QoS. We first introduce the society models of interest to this work.

5.2.1 Homo Egualis Society Model

In many decision-making and strategy-settings people do not behave like the self-interested "rational" actor depicted in neoclassical economics and game theory [96]. In a Homo Egualis society, individuals have an inequality aversion. As a result altruists appear in ultimatum and public games. As Gintis states in [96], support for Homo Egualis comes from the anthropological literature describing how Homo Sapiens evolved in small hunter-gatherer groups. Such societies had no centralized structure of governance, so the enforcement of norms depends on the voluntary participation of peers. A Homo Egualis Society can be modeled following [96] where the utility function of player m, u_m in an M-player game is:

$$u_m = z_m - \frac{\alpha_m}{M-1} \sum_{o=1, z_o > z_m}^{M} (z_o - z_m) - \frac{\beta_m}{M-1} \sum_{o=1, z_o < z_m}^{M} (z_m - z_o) \tag{5.8}$$

where $\mathbf{z} = [z_1, \cdots, z_m, \cdots, z_M]^T$ is the pay-off vector of the players and $0 \leq \beta_m < \alpha_m < 1$. In Eq. (5.8), considering β_m less than α_m in the utility model reflects the fact that Homo Egualis exhibits a weak urge to inequality when doing better than the others and a strong

urge to reduce inequality when doing worse that the others. In [96], it is also shown that in this model the salient behaviors in ultimatum and public games, where fairness does matter, can be reproduced.

5.2.2 Homo Parochius Society Model

Homo Parochius is the society that divides the world into insiders and outsiders according to context-dependent and even apparently arbitrary characteristics [96]. They care more for the welfare of insiders than outsiders and partially suppress personal goals in favor of the goals of the group insiders. Race, ethnicity, language and nationality are well-known examples of characteristics that are used to distinguish "insiders" from "outsiders." In a Homo Parochius Society, the utility function u_m of member m in a group of "insiders" of size N can be defined as

$$u_m = z_m - \alpha_m \sum_{o=1, z_m > z_o}^{N} (z_m - z_o), \tag{5.9}$$

where $\mathbf{z} = [z_1, \cdots, z_m, \cdots, z_N]^T$ is the pay-off vector of the members and $0 \leq \alpha_m < 1$. The utility model in Eq. (5.9) captures the fact that when insiders in a group are performing worse than others in the same group, then others suppress their personal goals by allowing a decrease in utility. Similarly, in the cognitive society, the secondary users belonging to the same service provider or association may provide expedient access to other members such as sharing more airtime or offering higher spectrum opportunities [95].

5.2.3 Homo Reciprocan Society Model

Homo reciprocans interact strategically with a propensity to cooperate [96]. They respond to cooperate behavior by maintaining or increasing the level of cooperation and retaliate against offenders that exhibit noncooperative behavior even if this comes at cost. That the retaliatory action could lead to a loss of future personal gains does not matter to the Homo reciprocan. Homo reciprocans are not selfish in that they try to maximize their own payoffs but they are not selfless altruists of Utopian theory either (when other forms of

punishment are not available, homo reciprocan responds to defection with defection, leading to a downward spiral of noncooperation). Gift exchange is a good example of reciprocal behavior where one agent behaves more kindly than required toward another, with the hope and expectation that the other will respond kindly as well [95]. In [97], the author defines utility function of each player incorporating kindness in a two players game. In [97], the utility function of player i is defined as

$$u_i(e_i, d_j, c_i) = \pi_i(e_i, d_j) + \widetilde{f}_j(d_j, c_i)[1 + f_i(e_i, d_j)], \tag{5.10}$$

where, $\pi_i(e_i, d_j)$ is individual i's material payoff given that he takes action e_i and he believes individual j's actions are d_j, c_i is the individual j's belief about individuals i's actions, $f_i(e_i, d_j)$ and $\widetilde{f}_j(d_j, c_i)$ are kindness function of individual i and belief of individual i how kind the other individual to him, respectively. $f_i(e_i, d_j)$ and $\widetilde{f}_j(d_j, c_i)$ are defined as

$$f_i(e_i, d_j) = \frac{\pi_j(d_j, e_i) - \pi_j^e(d_j)}{\pi_j^h(d_j) - \pi_j^{min}(d_j)} \tag{5.11}$$

and

$$\widetilde{f}_j(d_j, c_i) = \frac{\pi_i(c_i, d_j) - \pi_i^e(c_i)}{\pi_i^h(c_i) - \pi_i^{min}(c_i)}, \tag{5.12}$$

where, $\pi_j^e(d_j)$ is what individual i think is the "equitable payoff" for individual j and is defined as $\pi_j^e(d_j) = [\pi_j^h(d_j) + \pi_j^l(d_j)]/2$, $\pi_j^h(d_j)$ is individual j's highest possible payoff and $\pi_j^l(d_j)$ is individual j's lowest possible payoff from all possible Pareto outcomes, and $\pi_j^{min}(d_j)$ is individual j's lowest income. The utility model in Eq. (5.10) captures the individuals desire to be unkind to somebody that has been unkind to them. If individual i believes that individual j is kind to him (the function $\widetilde{f}_j(d_j, c_i)$ is positive), then he would increase his utility be being kind in return (the function $f_i(e_i, d_j)$ is positive). If individual i believes that individual j is unkind to him (the function $\widetilde{f}_j(d_j, c_i)$ is negative), then he would increase his utility be being unkind in return (the function $f_i(e_i, d_j)$ is negative).

In the following section, we develop evolution models for $w_i^p(n)$ and $w_i^b(n)$ following the concepts of HE, HP and HR society models.

5.3 Modeling Fairness

We define two fairness metrics, one based on instantaneous average power per bit $p_i^{opt}(n,k)$ and another based on instantaneous allocated rate $b_i^{opt}(n,k)$ for user i as

$$x_i^p(n) = \frac{1}{L}\sum_{k=1}^{L} p_i^{opt}(n,k) \tag{5.13}$$

and

$$x_i^b(n) = \frac{R_i^l(n)}{\sum_{k=1}^{L} b_i^{opt}(n,k)}. \tag{5.14}$$

Equation (5.13) tells that a lower value of $x_i^p(n)$ means a favorable power allocation from the resource allocation scheme to a user. Equation (5.14) tells that $x_i^b(n)$ can take value between 0 to 1. It is to be noted that $x_i^b(n)$ with value close to 0 indicates a comparatively higher allocated rate to minimum requirement (favorable rate allocation to a user) and $x_i^b(n)$ with value close to 1 indicates a comparatively lower allocated rate to minimum requirement. The fairness weights $w_i^p(n)$ and $w_i^b(n)$ are modeled as a function of metrics $x_i^p(n)$ and $x_i^b(n)$, respectively, following the concepts in human society model. The metrics $x_i^p(n)$ and $x_i^b(n)$ into fairness weights $w_i^p(n)$ and $w_i^b(n)$, respectively, capture the current quality of experiences and the evolution models of $w_i^p(n)$ and $w_i^b(n)$ capture the past quality of experiences.

The system level fairness (as in [77]) at time instant n can be defined as

$$Fairness\ index(n) = \frac{1}{M}\frac{\sum_{i=1}^{M} x_i(n)}{\sum_{i=1}^{M} x_i^2(n)}\sum_{i=1}^{M} x_i(n), \tag{5.15}$$

where, $x_i(n) = x_i^p(n)$ or $x_i^b(n)$. It is also to be noted that system fairness index can take value from 0 to 1. An index 0/1 means that system is totally unfair/fair in allocation. An index close to 1 results when $x_i(n)$ of all users are comparable and close to 1. As an example, for a system with three users ($M = 3$), an allocation scheme that results $x_1(n)$, $x_2(n)$ and $x_3(n)$ as 0.85, 0.75 and 0.60, respectively (system fairness index value is computed as 0.98) is more fair than the allocation scheme that results $x_1(n)$, $x_2(n)$ and $x_3(n)$ as 1.00, 0.90 and 0.30, respectively (system fairness index value is computed as 0.85). This is because 0.85, 0.75 and 0.60 has lower variance than 1.00, 0.90 and 0.30 (same average in both cases).

5.3.1 Weight Evolution based on HE Society Model

The evolution model for the fairness weights $w_i^{p,HE}(n)$ and $w_i^{b,HE}(n)$ based on HE society model are shown in Algorithm 5.1. In Algorithm 5.1, n^{max} represents considered time horizon. In the first time instant, fairness weight vectors, $\mathbf{w}^{p,HE}(n)$ and $\mathbf{w}^{b,HE}(n)$ are initialized to $\mathbf{1}$. In the following time instants, based on relative values of quality of experiences $x_i^p(n)$ and $x_i^b(n)$, weights $w_i^{p,HE}(n)$ and $w_i^{b,HE}(n)$ are updated. A relatively higher value of $x_i^p(n)$ and $x_i^b(n)$ (with respect to other users) result $w_i^{p,HE}(n)$ and $w_i^{b,HE}(n)$ to be a higher value. A higher value of $w_i^{p,HE}(n)$ and $w_i^{b,HE}(n)$ causes more importance on minimizing power and maximizing allocated rate, respectively, in next time instant. For a user i, a smaller value for β_m than α_m indicates a weak urge to reduce $w_i^{p,HE}(n)$ and $w_i^{b,HE}(n)$ (as it reduces importance on minimizing power and maximizing allocated rate, respectively, in next time instant). The criteria $\beta_m < \alpha_m$ captures the Homo Egualis society attitude among the SUs in CRN.

Algorithm 5.1: Evolution model of weights $w_i^{p,HE}(n)$ and $w_i^{b,HE}(n)$

while $n <= n^{max}$ **do**
  Initialization;
  if $(n-1) == 1$ **then**
    for $i = 1,\ 2, \cdots,\ M$ **do**
      $w_i^{p,HE}(0) = 1$
      $w_i^{b,HE}(0) = 1$
    end for
    $\mathbf{w}^{p,HE}(0) = \mathbf{w}^{p,HE}$
    $\mathbf{w}^{b,HE}(0) = \mathbf{w}^{b,HE}$
  end if
  for $i = 1,\ 2, \cdots,\ M$ **do**
    for $j = 1,\ 2, \cdots,\ M$ **do**
      if $j \neq i$ **then**
        if $(x_i^p(n) \geq x_j^p(n))$ **then**
          $w_i^{p,HE}(n) = \max(0, w_i^{p,HE}(0) + \alpha_m(x_i^p(n) - x_j^p(n)))$
        else
          $w_i^{p,HE}(n) = \max(0, w_i^{p,HE}(0) - \beta_m(x_j^p(n) - x_i^p(n)))$
        end if
        if $(x_i^b(n) \geq x_j^b(n))$ **then**
          $w_i^{b,HE}(n) = \max(0, w_i^{b,HE}(0) + \alpha_m(x_i^b(n) - x_j^b(n)))$
        else
          $w_i^{b,HE}(n) = \max(0, w_i^{b,HE}(0) - \beta_m(x_j^b(n) - x_i^b(n)))$
        end if
      end if
    end for
  end for

end while

5.3.2 Weight Evolution based on HP Society Model

The evolution model for the fairness weights $w_i^{p,HP}(n)$ and $w_i^{b,HP}(n)$ based on HP society model are shown in Algorithm 5.2. The number of groups among SUs is assumed to be G. An user i belonging to a group g where $g = 1,\ 2..., G$ is denoted as ϕ_i^g. As in Algorithm 5.1, in Algorithm 5.2, n^{max} represents considered time horizon and the fairness weight vectors,

$\mathbf{w}^{p,HP}(n)$ and $\mathbf{w}^{b,HP}(n)$ are initialized to $\mathbf{1}$. In the following time instants, based on relative values of quality of experiences $x_i^p(n)$ and $x_i^b(n)$, weights $w_i^{p,HP}(n)$ and $w_i^{b,HP}(n)$ are updated. A relatively higher value of $x_i^p(n)$ and $x_i^b(n)$ (with respect to other users in the group) result $w_i^{p,HP}(n)$ and $w_i^{b,HP}(n)$ to be a higher value. A higher value of $w_i^{p,HP}(n)$ and $w_i^{b,HP}(n)$ cause more importance on minimizing power and maximizing allocated rate, respectively, in next time instant. It is to be noted that users in two different groups do not care for each other.

Algorithm 5.2: Evolution model of weights $w_i^{p,HP}(n)$ and $w_i^{b,HP}(n)$

while $n <= n^{max}$ **do**
  Initialization;
  if $(n-1) == 1$ **then**
    for $i = 1,\ 2, \cdots,\ M$ **do**
      $w_i^{p,HP}(0) = 1$
      $w_i^{b,HP}(0) = 1$
    end for
    $\mathbf{w}^{p,HP}(0) = \mathbf{w}^{p,HP}$
    $\mathbf{w}^{b,HP}(0) = \mathbf{w}^{b,HP}$
  end if
  for $i = 1,\ 2, \cdots,\ M$ **do**
    for $j = 1,\ 2, \cdots,\ M$ **do**
      if $j \neq i$ and $\phi_j^g = \phi_i^g$ **then**
        if $(x_i^p(n) \leq x_j^p(n))$ **then**
          $w_i^{p,HP}(n) = \max(0, w_i^{p,HP}(0) - \beta_m(x_j^p(n) - x_i^p(n)))$
        end if
        if $(x_i^b(n) \leq x_j^b(n))$ **then**
          $w_i^{b,HP}(n) = \max(0, w_i^{b,HP}(0) - \beta_m(x_j^b(n) - x_i^b(n)))$
        end if
      end if
    end for
  end for

end while

5.3.3 Weight Evolution based on HR Society Model

The evolution model for the fairness weights $w_i^{p,HR}(n)$ and $w_i^{b,HR}(n)$ based on HR society attitude are shown in Algorithm 5.3. As in Algorithm 5.1 or 5.2, in Algorithm 5.3, n^{max}

represents considered time horizon and the fairness weight vectors, $\mathbf{w}^{p,HR}(n)$ and $\mathbf{w}^{b,HR}(n)$ are initialized to $\mathbf{1}$. In the following time instants, based on kindness or unkindness attitude, weights $w_i^{p,HR}(n)$ and $w_i^{b,HR}(n)$ are updated. In this work, to a user, we consider resource allocation scheme is unkind or kind if the fairness metric based on power and rate, $x_i^p(n)$ and $x_i^b(n)$, respectively are higher or lower than the average of fairness metric of all other users in the network based on power and rate (i.e., $\frac{1}{M}\sum_{j=1,j\neq i}^{M} x_j^p(n)$ and $\frac{1}{M}\sum_{j=1,j\neq i}^{M} x_j^b(n)$), respectively. If an user finds resource allocation as unkind, user increases utility by increasing weights which results in more desire of minimizing power or maximizing rate in next time instant. If an user finds resource allocation as kind, user decreases utility by decreasing weights which results in less desire of minimizing power or maximizing rate in next time instant.

Algorithm 5.3: Evolution model of weights $w_i^{p,HR}(n)$ and $w_i^{b,HR}(n)$

while $n <= n^{max}$ **do**
  Initialization;
  if $(n-1) == 1$ **then**
    for $i = 1,\ 2, \cdots,\ M$ **do**
      $w_i^{p,HR}(0) = 1$
      $w_i^{b,HR}(0) = 1$
    end for
    $\mathbf{w}^{p,HR}(0) = \mathbf{w}^{p,HR}$
    $\mathbf{w}^{b,HR}(0) = \mathbf{w}^{b,HR}$
  end if
  for $i = 1,\ 2, \cdots,\ M$ **do**
    for $j = 1,\ 2, \cdots,\ M$ **do**
      if $j \neq i$ **then**
        if $(x_i^p(n) \geq \frac{1}{M-1}\sum_{j=1,j\neq i}^{M} x_j^p(n))$ **then**
          $w_i^{p,HR}(n) = \max(0, w_i^{p,HR}(0) + \alpha_m(\frac{1}{M-1}\sum_{j=1,j\neq i}^{M} x_j^p(n) - x_i^p(n)))$
        else
          $w_i^{p,HR}(n) = \max(0, w_i^{p,HR}(0) - \beta_m(x_i^p(n) - \frac{1}{M-1}\sum_{j=1,j\neq i}^{M} x_j^p(n)))$
        end if
        if $(x_i^b(n) \geq \frac{1}{M-1}\sum_{j=1,j\neq i}^{M} x_j^b(n))$ **then**
          $w_i^{b,HR}(n) = \max(0, w_i^{b,HR}(0) + \alpha_m(\frac{1}{M-1}\sum_{j=1,j\neq i}^{M} x_j^b(n) - x_i^b(n)))$
        else
          $w_i^{b,HR}(n) = \max(0, w_i^{b,HR}(0) - \beta_m(x_i^b(n) - \frac{1}{M-1}\sum_{j=1,j\neq i}^{M} x_j^b(n)))$
        end if
      end if
    end for
  end for

end while

It is to be noted that in the resource allocation framework (Eq. (5.1)), $w_i^p(n)$ can be $w_i^{p,HE}(n)$, $w_i^{p,HP}(n)$ or $w_i^{p,HR}(n)$ and $w_i^b(n)$ can be $w_i^{b,HE}(n)$, $w_i^{b,HP}(n)$ or $w_i^{b,HR}(n)$.

Table 5.2: *Channel Quality Parameters*

Channel, k	1	2	3	4	5	6	7	8
$\sigma^2(n,k)$, $(\times 10^{-3})$	5	4	3	2	2.5	6	4	4

Table 5.3: *Minimum Rate Requirement of Users*

User, i	1	2	3	4	5	6	7	8	9
$R_i^l(n)$	8	9	10	12	13	10	14	15	10

5.4 Numerical Results

In this section, we evaluate the impact of introducing dynamic fairness weight in the resource allocation framework. We assume a CRN with $L = 8$ available channels and a total of $M = 9$ secondary users. Table 5.2 provides information on the channel quality for all L channels. Table 5.3 lists the minimum rate requirement for each SU. Finally, Table 5.4 contains all other system parameters that are relevant to our resource allocation framework. Our objective is to find the optimal transmit power and rate that each of the M SUs should employ to achieve fairness in quality of experience across them.

Since optimization formulation has one non-linear constraint (constraint $C7$); we use

Table 5.4: *System Parameters*

$p_i^{max}(n,k)\ \forall\ i,\ k$	5
$b_i^{max}(n,k)\ \forall\ i,\ k$	6
$p_{e,i}^{th}(n)\ \forall\ i$	10^{-3}
$I_{th}(n,k)\ \forall\ k$	$200 \times \sigma^2(n,k)$
$R_{ch}^u(n,k)\ \forall\ k$	30
$\rho_{j,i}(n)$	0.03125
n^{max}	50
α_m	0.35
β_m	0.15

"Sequential Quadratic Programming (SQP)" method to solve the problem. In all simulation, we set the scalarization constants τ_1 and τ_2 to 0.5. We first consider the HE based dynamic fairness weight case. It is to be noted that unweighted resource allocation corresponds to the case of assuming $w_i^{p,HE}(n) = 1,\ \forall\ i,\ n$ and $w_i^{b,HE}(n) = 1,\ \forall\ i,\ n$ in the resource allocation scheme described in Sec. 5.1. Figures 5.1(c), 5.1(d), and 5.2(c), 5.2(d) present the short term averaged (averaged over a moving window of size 9) power and rate allocated for users 1 and 5 from weighted and unweighted resource allocation schemes, respectively. Evolution of weights $w_1^{p,HE}(n)$, $w_1^{b,HE}(n)$ and $w_5^{p,HE}(n)$, $w_5^{b,HE}(n)$ are also shown in Figs. 5.1(a), 5.1(b) and 5.2(a), 5.2(b), respectively.

Figures 5.1(d) and 5.2(d) show that a decreasing fairness weight with time (as shown in 5.1(b)) results in smaller allocated rate; whereas an increasing fairness weight with time (as shown in 5.2(b)) results in an higher allocated rate. Figures 5.1(d) and 5.2(d) also show that allocated rate with dynamic fairness weight is smaller for user 1 and higher for user 5 compared to that obtained from unweighted resource allocation. This indicates that HE fairness weight in the resource allocation scheme results in a scenario where the rate of user 1 is sacrificed and rate of user 5 is allowed to increase resulting in a rational allocation. That is, dynamic fairness weights $w_i^{p,HE}(n)$ and $w_i^{b,HE}(n)$ promote cooperative, rational attitude of SUs in CRN like human beings in Homo Egualis Society. Similar impact of the weights are observed on short term averaged rate allocated to other users in the CRN.

Figures 5.1(c) and 5.2(c) depict that allocated power is insensitive to fairness weight (shown in Figs. 5.1(a) and 5.2(a)) for both users. This can be explained as follows. In order to satisfy constraint $C7$, one can increase power, $p_i(n,k)$ or decrease rate, $b_i(n,k)$. Equation (5.6) suggests that varying rate, $b_i(n,k)$ is more effective. This is because $\gamma_i(n,k)$ is linearly related to $p_i(n,k)$ while rate $b_i(n,k)$ is an exponent of denominator in constraint $C7$. Therefore, for a given BER constraint, the optimization engine prefers to vary $b_i(n,k)$ instead of $p_i(n,k)$ to satisfy the constraint. Hence, changing $w_i^{b,HE}(n)$ has stronger impact than $w_i^{p,HE}(n)$ in the resource allocation scheme.

It is to be noted that at any instant of time, resource allocation in a coupled multiuser environment is determined by user requirements, fairness weights and channel conditions. That is, it is not necessary that an increasing fairness weight for rate will always result in higher allocated rate and vice versa. An increasing fairness weight merely indicates a preference for increasing rate, which may or may not be feasible (depending on channel conditions, other SUs weights and rate requirements). Figures 5.3(a) and 5.3(b) show one such example. Here, we see that a higher fairness weight results in smaller allocated rate compared to unweighted scheme for user 2. It is however, true that the allocated rate will be equal to or greater than the minimum rate requirements for all users at all time instants.

Figure 5.4(a) shows the long term averaged (averaged over 50 time instants) transmit power allocated across users. As expected, the transmit power resulting from weighted resource allocation scheme is same as the power allocated based on unweighted allocation scheme. Figure 5.4(b) illustrates the long term averaged rate allocated across users. Here, we see that the unweighted scheme allocates comparable rates across users irrespective of their demand. However, HE weights in the resource allocation scheme cause some users {1, 2, 3, 4, 6, 9} to sacrifice resources and other users {5, 7, 8} to gain resources to maintain a balance in allocation. Long term averaged rate allocated across users also reflect the cooperative, rational attitude of SUs in CRN.

Next, we study the impact of imposing an HR society model into the optimization framework. Figure 5.5 shows the long term averaged allocated rate across users with HR and HE based dynamic fairness weights in the resource allocation scheme. Here, we observe that HR based dynamic fairness weights result in higher allocated rate than HE based dynamic fairness weights in the resource allocation scheme. The reason can be explained as follows. In HE society based evolution, every user's weight is updated based on all other users relative fairness level (Algorithm 5.1). On the other hand, in HR society based evolution, the weight of every user is updated based on the average of all other users fairness level (Algorithm 5.3). Therefore, in the HR model, users are not over benefitted or over

Table 5.5: *Fairness index of weighted (HE and HR model) and unweighted schemes*

Resource allocation scheme	Weighted (HE)	Weighted (HR)	Unweighted
Fairness index	0.9887	0.9887	0.9616

penalized by every action of other users. Instead the weight reacts to an average behavior of all other users. As a result, the urge for equality in HR model is not as strong as the HE model resulting in higher total rate (for the given example, total rate for HR and HE models is 142 and 134, respectively). The long term averaged fairness index for weighted and unweighted resource allocation schemes are shown in Table 5.5. Table 5.5 shows that both HE and HR model based weighted allocation schemes provide comparable fairness index and better than unweighted allocation scheme. The reason of comparable fairness index of HR and HE models can be explained as follows. In HR model, all users experience a proportional increase in allocated rate relative to HE model. As a result, system level fairness index as defined in Eq. (5.15) are comparable.

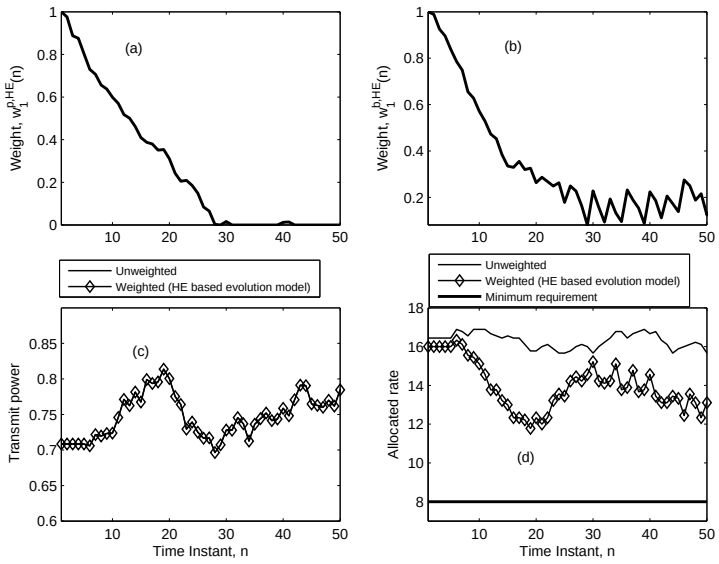

Figure 5.1: *Short term averaged transmit power and rate allocated to user 1 from weighted (HE based evolution model) and unweighted resource allocation schemes.*

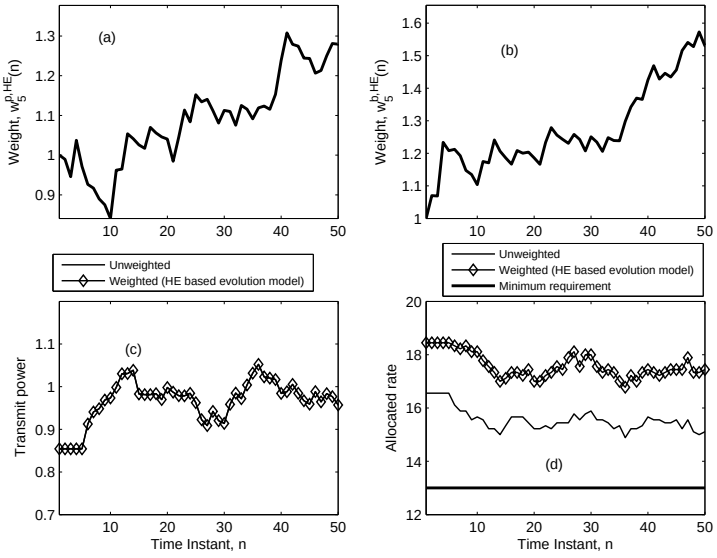

Figure 5.2: *Short term averaged transmit power and rate allocated to user 5 from weighted (HE based evolution model) and unweighted resource allocation schemes.*

Finally, we evaluate the effect of incorporating an HP model in resource allocation framework. Consistent with the HP model, we assume users $\{2,\ 3,\ 4,\ 6,\ 8\}$ form a group (group 1) and users $\{1,\ 5,\ 7,\ 9\}$ do not form any group. We denote this grouping scenario as case 1. The users in the group i.e., $\{2,\ 3,\ 4,\ 6,\ 8\}$ behave with each other as "insiders." The other users i.e., $\{1,\ 5,\ 7,\ 9\}$ are "outsiders" to the "insiders." The impact of group formation is better illustrated by the subsystem/group level fairness index for rates of "insiders" and "outsiders." In Fig. 5.6(a), the short term averaged fairness index is computed based on instantaneous fairness index using Eq. (5.15) and assuming a subsystem consists of users in "insiders" group. We observe from Fig. 5.6(a) that "insiders" has higher fairness index than unweighted allocation scheme. In Fig. 5.6(b), the short term averaged fairness index is computed based on instantaneous fairness index using Eq. (5.15) and assuming a subsystem consists of users in "outsiders" group. From Fig. 5.6(b), we see that "outsiders" has almost same fairness index compared to unweighted allocation scheme. That is, forming group helps to achieve a more fair allocation in the group than not forming group. The

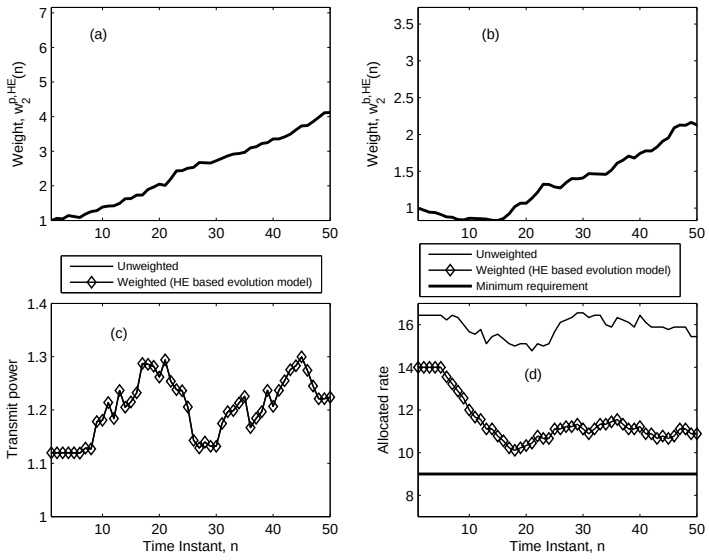

Figure 5.3: *Short term averaged rate allocated to user 2 from weighted (HE based evolution model) and unweighted resource allocation schemes.*

reason behind this is users in a group sacrifice personal goals and collaborate each other to maximize group goal.

Next, we let users $\{2,\ 3,\ 4,\ 6,\ 8\}$ and users $\{1,\ 5,\ 7,\ 9\}$ form two groups (group 1 and 2, respectively). We denote this grouping scenario as case 2. Users within a group behave with each other as "insiders" and to the users of other group behave as "outsiders." Figures 5.7(a) represents the fairness index across the users in the group 1 from HP weighted (case 2) compared to HP weighted (case 1) or unweighted allocation scheme. Here, we see that group level fairness index for case 2 is comparable to case 1 and better than unweighted allocation scheme. Figure 5.7(b) shows the fairness index across the users in the group 2 from case 2 compared to case 1 or unweighted allocation scheme. Here, we observe that forming group (i.e., scenario case 2) improves the fairness index across the users compared to not forming group or unweighted allocation scheme. The long term averaged fairness index of groups 1 and 2 for two cases are shown in Table 5.6. Forming group (case 2) results in achieving a better overall system fairness index than unweighted allocation scheme.

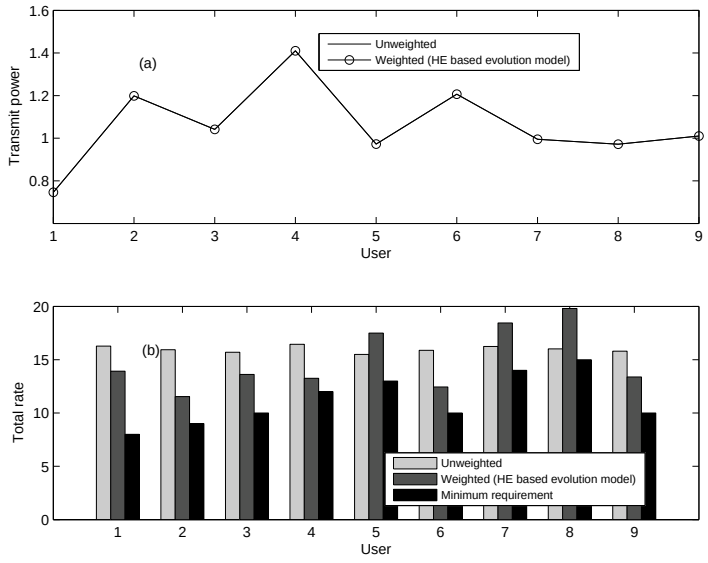

Figure 5.4: *Long term averaged transmit power and rate allocated across users from weighted (HE based evolution model) and unweighted resource allocation schemes.*

Table 5.6: *Fairness index of weighted (HP model) and unweighted schemes*

Resource allocation scheme	Weighted (HP)			Unweighted
	Group 1	Group 2	Overall	
Case 1	0.9835	0.9559	0.9559	0.9573
Case 2	0.9831	0.9804	0.9583	0.9573

5.5 Summary

In this chapter, we determine optimal transmit power and rate distribution that each SU needs to employ in a multi-channel CRN considering current and past history of user experience with respect to QoS. We consider fairness weights for each user that captures current and past history of user experience and study three different evolution models for the fairness weights based on HE society model, HR society model and HP society model. We consider Jain system level fairness index as a measure of fairness in resource allocation scheme. Simulation results illustrate that incorporating dynamic fairness weights in

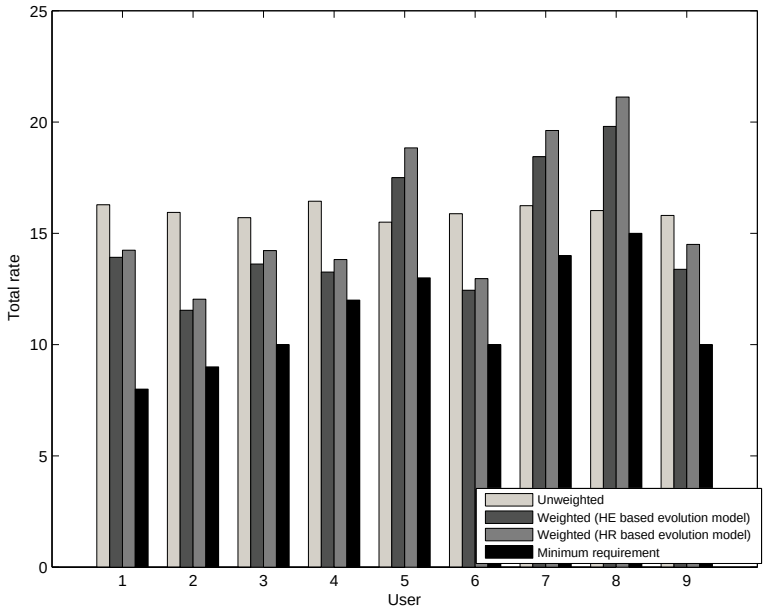

Figure 5.5: *Long term averaged rate allocated across users from weighted (HE, HR based evolution models) and unweighted resource allocation schemes.*

the resource allocation scheme provide better system level fairness index than unweighted scheme.

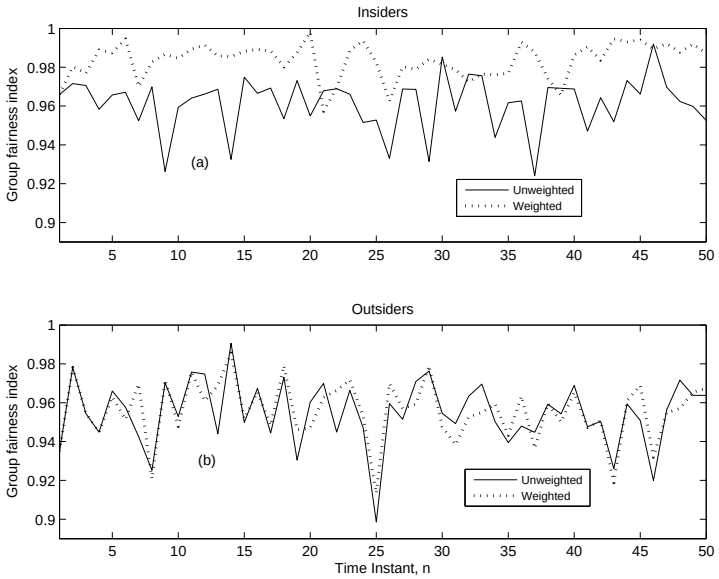

Figure 5.6: *Short term averaged subsystem/group level fairness index ((a) for insiders and (b) for outsiders) for rate from HP based weighted allocation scheme.*

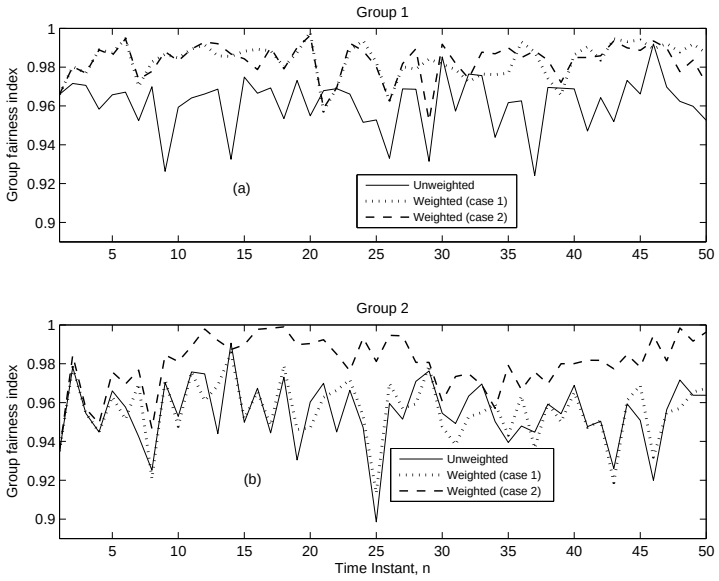

Figure 5.7: *Short term averaged subsystem/group level fairness index ((a) for group 1 and (b) for group 2) for rate from HP based weighted allocation scheme.*

Chapter 6

Game Theory based Distributed Implementation

In chapter 4, we proposed a joint resource allocation framework to determine transmit power and rate for secondary users in a multi-channel multiuser competitive CRN. Firstly, we solved the resource allocation problem in centralized manner and then we developed dual based distributed solution approaches. In this chapter, we apply game theoretic concepts to develop a distributed scheme for the resource allocation framework introduced in chapter 4. Specifically, we concentrate on formulating a game in a "noncooperative" CRN. We adopt the same notations introduced in chapter 3.

The rest of the chapter is organized as follows. Section 6.2 describes the game formulation. Section 6.3 shows the analysis of the game. Numerical results are presented in Sec. 6.4. Other possible game formulation is shown in Sec. 6.5. Finally, Sec. 6.6 summarizes the chapter.

6.1 Game Theory

Game theory analyzes the interactions of rational decision makers in decision-making processes. Game theory requires that each player has an *action space* of possible actions and a *utility function*, which represents the relative desirability of a player's action (chosen from his action space) in combination with actions from the rest of the players (chosen from their

action space) [12]. Players are said to play *rationally* if they try to choose an action that, in conjunction with the other player actions, maximizes their utility function. A strategic noncooperative game Γ is expressed as $\Gamma = \{\Omega,\ A,\ U\}$ and consists of components:

1. Player set $\Omega\ :\ \Omega = 1,\ 2,\ \ldots,\ M$, where M is the number of rational players.

2. Action set $A\ :\ \mathbf{a}\ \in\ A = \prod_{i=1}^{M} A_i = A_1 \times A_2 \times \ldots A_n$, where each component, a_i, of the action vector $\mathbf{a}$ belongs to the set A_i, the action set of player i. Action vector is also denoted as $\mathbf{a} = (a_i, \mathbf{a}_{-i})$, where a_i is player i's action and $\mathbf{a}_{-i}$ denotes the actions of rest $(M-1)$ players. $A_{-i} = \prod_{j=1, j\neq i}^{M} A_j$ is the action set of all players other than player i.

3. Utility U: $U_i:\ A \rightarrow \mathcal{R}$ is the utility (payoff) function of player i, which depends on the strategies of all players and $U = (U_1, \ldots, U_M)\ : A \rightarrow \mathcal{R}^M$ denotes the utility vector of utility functions.

Several properties of action vectors have been identified. The most common is Nash Equilibrium (NE) [12]. The NE is an action vector that corresponds to the mutual best response for all players. In other words, at NE, no individual player can benefit from unilateral deviation.

Theorem 6.1.1. *(Nash Equilibrium) An action vector* $\hat{\mathbf{a}}$ *is a NE if , for every player* i *and every action vector* $\mathbf{a}$

$$U_i(\hat{a}_i, \hat{\mathbf{a}}_{-i}) \geq U_i(a_i, \hat{\mathbf{a}}_{-i}). \tag{6.1}$$

6.2 Game Formulation 1

In a "noncooperative" CRN, to determine power and rate, each SU is interested in minimizing its own power and maximizing its own rate (modulation order) while maintaining QoS. Let $G = \{\Omega,\ \mathcal{P},\ \mathcal{B},\ \{u_i(.)\}\}$ denote the noncooperative power and rate (modulation order) control game (NPRG) corresponding to our proposed joint resource allocation framework (4.8). $\Omega = 1,\ 2,\ \ldots,\ M$ is the set of players corresponding to M secondary users;

$P = \mathcal{P}_1 \times \mathcal{P}_2 \times \ldots$ is the action space for power with $\mathcal{P}_i$ as the action set for power of player i; $B = \mathcal{B}_1 \times \mathcal{B}_2 \times \ldots$ is the action space for rate with $\mathcal{B}_i$ as the action set for rate of player i. Each SU selects a power vector $\mathbf{p_i} \in \mathcal{P}_i$ and a rate vector $\mathbf{b_i} \in \mathcal{B}_i$. For ease in presentation, we define the action for user i as $\mathbf{y}_i = [\mathbf{p}_i^T \ \mathbf{b}_i^T]^T$, where, ($\mathbf{p}_i = [p_i(1) \ p_i(2) \ldots p_i(L)]^T$ and $\mathbf{b}_i = [b_i(1) \ b_i(2) \ldots b_i(L)]^T$). We consider utility function of user i as

$$u_i(\mathbf{y}_i, \mathbf{y}_{-i}) = -\tau_1 \sum_{k=1}^{L} p_i(k) + \tau_2 \sum_{k=1}^{L} b_i(k), \tag{6.2}$$

where, $\mathbf{y}_{-i}$ is the union set of all other users actions and $\mathbf{y}_{-i} \triangleq [\mathbf{y}_1^T \ldots \mathbf{y}_{i-1}^T \ \mathbf{y}_{i+1}^T \ldots \mathbf{y}_M^T]^T$. The "noncooperative" game formulation to determine transmit power and rate can be formally stated as

$$\begin{aligned}
\text{Determine} \quad & \mathbf{y}_i \\
\text{To Maximize} \quad & u_i(\mathbf{y}_i, \mathbf{y}_{-i}) \\
\text{subject to} \quad & \\
& CG1: \ 0 \leq p_i(k) \leq p_i^{gmax}(k) \ \forall \ i, \ k \\
& CG2: \ 1 \leq b_i(k) \leq b_i^{gmax}(k) \ \forall \ i, \ k \\
& CG3: \ \sum_{k=1}^{L} b_i(k) \geq R_i^l \ \forall \ i, \\
& CG4: -\gamma_i(k) \leq -C_{qarg}(2^{b_i(k)} - 1) \ \forall \ i, \ k,
\end{aligned} \tag{6.3}$$

where,

$$\gamma_i(k) = \frac{p_i(k) h_{i,i}(k)}{\sum_{j=1, \ j \neq i}^{\widetilde{N}_s(k)} p_j(k) h_{j,i}(k) {\rho_{j,i}}^2 + \sigma^2(k)}. \tag{6.4}$$

It is important to note how the system constraints $C23$ and $C24$ in our proposed joint resource allocation framework (4.8) are considered in the formulated "noncooperative" game. We assume the total interference (constraint $C23$) caused by all SUs in a channel is divided equally across all SUs in that channel. This approach results in changing maximum transmit power for each SU. In 6.3, this is captured in the constraint $CG1$. Here, $p_i^{gmax}(k)$ is set as

the minimum of $p_i^{max}(k)$ and $I_{th}(k)/(h_{i,m}(k)\widetilde{N}_s(k))$. Similarly, total supported rate in a channel is also divided across all SUs in that channel. This approach results in changing maximum possible rate for each SU. In 6.3, this is captured in the constraint $CG2$. Here, $b_i^{gmax}(k)$ is set as the minimum of $b_i^{max}(k)$ and $R_{ch}^u(k)/(\widetilde{N}_s(k))$. The value of $\widetilde{N}_s(k)$ can be determined using our proposed modeling and forecasting tool presented in chapter 2.

6.3 Analysis of the Game

The solution that is most widely used for game theoretic implementations is the Nash Equilibrium (NE). At a NE point, given the power and rate levels of other users, no user can improve its utility level by making individual changes in its power and rate. The NE concept results in a stable solution of a game where players with conflicting interests compete through self optimization and reach a point where no player wishes to deviate. If there is a solution to the above game, then it would be the one that reaches NE. The following theorem show that a NE solution always exists for the game G in 6.3.

Theorem 6.3.1. *For a given $p_i^{gmax}(k)$, $p_i^{gmax}(k)$, R_i^l and C_{qarg}, there is at least one NE for the game G in 6.3.*

Proof. The game is our setup can be shown to be a concave game if the following two conditions are satisfied:
(1) the action spaces $\mathcal{P}$ and $\mathcal{B}$ are closed and bounded convex set and
(2) the utility function $u_i(\mathbf{y}_i, \mathbf{y}_{-i})$ is concave over its strategy set.
It is very easy to show that the first condition is satisfied by the game G. The utility function $u_i(\mathbf{y}_i, \mathbf{y}_{-i})$ is linear (and hence considered concave) in $p_i(k)$ and $b_i(k)$. As a concave game admits at least one NE [38], the theorem follows immediately. □

Given the existence of NE solution for the game, next we design an algorithm for SUs to reach the NE. The algorithm is shown in Algorithm 6.1. In Algorithm 6.1, t is the iteration counter. At first, each SU measures the interference and noise power

$(i.e., \sum_{j=1,\ j\neq i}^{\widetilde{N}_s(k)} p_j(k)h_{j,i}(k){\rho_{j,i}}^2 + \sigma^2(k))$ across its intended channels. Then, each user executes its own optimization problem (6.3) to determine power and rate optimally. Each user continues to do (1) measure the interference and noise power term and (2) solve own optimization problem until a certain number of iterations (t_{max}) is complete or stopping criteria as shown in the Algorithm 6.1 is reached. Finally, each user searches in the neighborhood for the optimal discrete valued $b_i^t(k)$ (denoted as $\mathbf{b}_i^{opt}$) and optimal transmit power $\mathbf{p}_i^{opt}$ corresponds to $\mathbf{b}_i^{opt}$ is recalculated using Eq. (6.3). Generally, ϵ is set to a reasonable small value.

Algorithm 6.1: Algorithm to reach NE for the game G

Stopping counter, $t = 1$;
while $(t \leq t_{max}$ or $||(\mathbf{p}_i^t - \mathbf{p}_i^{t-1}||/||\mathbf{p}_i^{t-1}|| \leq \epsilon),\ \forall i)$ **do**
▷ *% Execute optimization problem*
for $i = 1,\ 2, \cdots,\ M$ **do**
for $k = 1,\ 2, \cdots,\ L$ **do**
Measure the interference and noise power $(i.e., \sum_{j=1,\ j\neq i}^{\widetilde{N}_s(k)} p_j^{t-1}(k)h_{j,i}(k){\rho_{j,i}}^2 + \sigma^2(k))$ across the intended channels;
end for
Solve optimization problem (6.3) and obtain $\mathbf{p}_i^t$ and $\mathbf{b}_i^t$;
end for
for $i = 1,\ 2, \cdots,\ M$ **do**
Transmit $\mathbf{p}_i^t$;
end for
$t = t + 1$;
end while

The convergence of Algorithm 6.1 is always observed in simulation. However, the convergence condition of Algorithm 6.1 and uniqueness of NE are left as future work.

6.4 Numerical Results

In this section, we quantify the performance of the game theory based distributed implementation of the proposed joint resource allocation framework. We assume the same simulation setup as in chapter 4.

As the optimization problem (6.3) has one non-linear constraint (constraint $CG4$), once again we use "Sequential Quadratic Programming (SQP)" method to solve this problem. Figure 6.1 presents the transmit power and rate allocation across channels for users 4 (assuming scalarization parameters τ_1 and τ_2 to 0.5) from the proposed distributed scheme along with centralized scheme (Eq. 4.8). User 4 operates on channels 2, 3, 5, 6, 7, 9, 10 and 11. As in the centralized formulation, we can see that "reverse water filling" effect is observed for the distributed scheme as user tends to use more transmit power in poor quality channels. As in centralized scheme, SINR is not a constraint in this used-based resource allocation formulation, the user does not attempt to maintain a constant SINR in the channels. QoS is maintained by adjusting the rate allocated to each channel; e.g., Fig. 6.1 shows that for high SINR channels, more bits/channel are allocated and vice versa. The similar pattern on power and rate allocation is also observed for other users.

Figure 6.2(a) shows the allocation of total transmit power across users from both centralized (Eq. 4.8) and distributed schemes, respectively. Figure 6.2(b) shows the allocation of total rate across users from both centralized (Eq. 4.8) and distributed schemes, respectively. We see from Figs. 6.2(a) and 6.2(b) that both the total allocated power and rate across users in distributed case are comparable to centralized scheme. Therefore, our proposed distributed resource allocation scheme is successful in meeting the rate requirements for all SUs. The reason is obvious from the proposed formulation (6.3). A user executes the optimization problem (6.3) after checking the feasibility of the optimization problem. The feasibility is determined by user minimum rate requirement (constraint $CG3$) and the upper bound of rate (constraints $CG2$). For each user, if the optimization problem is feasible in terms of user minimum rate requirement and upper bound of rate, the distributed scheme is guaranteed to be successful in meeting the rate requirements for all SUs.

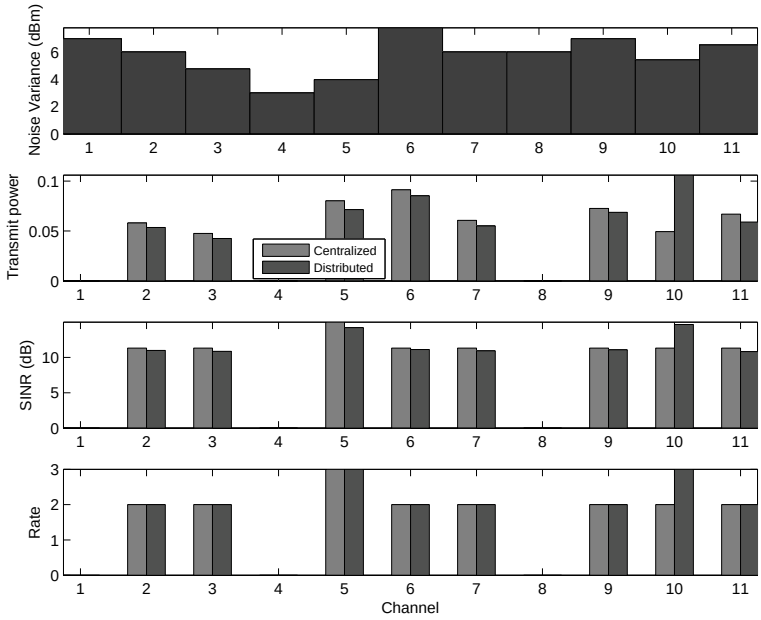

Figure 6.1: *Allocation of transmit power and rate with channel noise variance and SINR for user* 4 *($\tau_2/\tau_1 = 1$).*

6.5 Other Game Formulation

6.5.1 Game Formulation 2: Repeated Game

The game described in this chapter is a single shot/stage game. However, the same set of SUs may compete for resources over a long period of time. As described in section 6.2, each user is required to know the interference power (constraint $CG4$) for determining their optimal actions. In a "noncooperative" CRN, users cannot be forced to share this information. The game described in section 6.2 demands some kind of self enforcing mechanism to share the true information for interference power [98]. In this context, it may be reasonable to model the game in section 6.2 as a repeated or dynamic game where the players play multiple rounds. In repeated or dynamic game, the players decide on transmission parameters remembering the past experience. The utility of player i at time instant $n = N$ can be

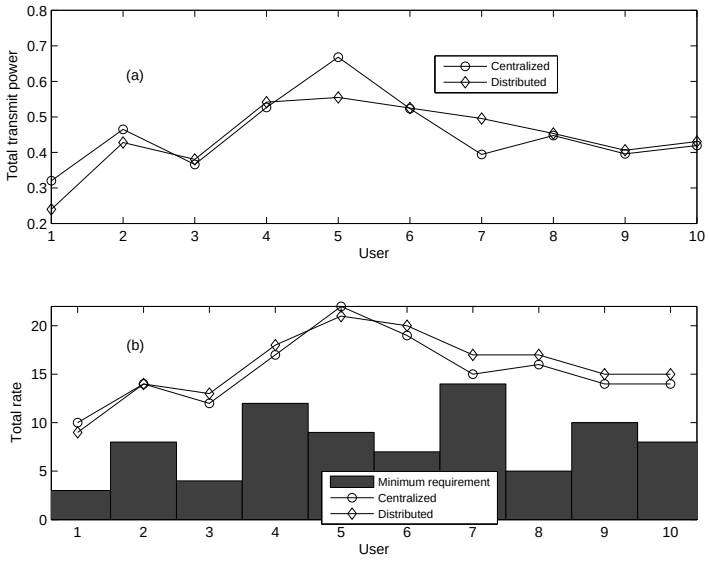

Figure 6.2: *Allocation of total transmit power and total rate across users ($\tau_2/\tau_1 = 1$).*

defined as

$$u_i(\mathbf{y}_i(N), \mathbf{y}_{-i}(N))(N) = -\tau_1 \sum_{k=1}^{L} w_{g,i}^{p}(N,k) p_i(N,k) + \tau_2 \sum_{k=1}^{L} w_{g,i}^{b}(N,k) b_i(N,k),$$
$$w_{g,i}^{p}(N,k) = \text{function of } (\mathbf{y}_i(n), \mathbf{y}_{-i}(n),\ n = 0, \cdots, N-1),$$
$$w_{g,i}^{b}(N,k) = \text{function of } (\mathbf{y}_i(n), \mathbf{y}_{-i}(n),\ n = 0, \cdots, N-1). \qquad (6.5)$$

Table 6.1 defines all relevant terms (at n-th time instant) used to state repeated game. In Eq. (6.5), $\mathbf{y}_i(N)$ is the action vector of player i at time instant N and $\mathbf{y}_i(N) = [\mathbf{p}_i(N)^T\ \mathbf{b}_i(N)^T]^T$, $\mathbf{p}_i(N) = [p_i(N,1) \cdots\ p_i(N,L)]^T$, $\mathbf{b}_i(N) = [b_i(N,1) \cdots\ b_i(N,L)]^T$, $\mathbf{y}_{-i}(N)$ is the action vector of all other players except player i at time instant N and $\mathbf{y}_{-i}(N) \triangleq [\mathbf{y}_1^T(N) \ldots \mathbf{y}_{i-1}^T(N) \ldots \mathbf{y}_M(N)]^T$, $u_i(\mathbf{y}_i(N), \mathbf{y}_{-i}(N))(N)$ is the utility value of player i at time instant N, $w_{g,i}^{p}(N,k)$ and $w_{g,i}^{b}(N,k)$ are the factors that capture experience till $(N-1)$ th instant of player i to decide on transmission parameters at time instant N. The

Table 6.1: *Notations*

$\sigma^2(n,k)$	Noise variance in k-th channel
$\rho_{j,i}(n)$	Orthogonality factor between users j and i
$h_{i,i}(n,k)$	Power gain from i-th transmitter to i-th receiver in k-th channel
$p_i(n,k)$	Transmit power per bit of i-th user in k-th channel
$b_i(n,k)$	Rate of i-th user in k-th channel
$R_i^l(n)$	Minimum required rate for i-th user
$C_{qarg}(n)$	BER threshold at receiver in any channel
$\gamma_i(n,k)$	SINR per bit for i-th user in k-th channel

repeated game formulation to determine transmit power and rate can be formally stated as

$$
\begin{aligned}
\text{Determine} \quad & \mathbf{y}_i(N) \\
\text{To Maximize} \quad & u_i(\mathbf{y}_i(N), \mathbf{y}_{-i}(N))(N) \\
\text{subject to} \quad & \\
& CDG1: \ 0 \leq p_i(N,k) \leq p_i^{gmax}(N,k) \ \forall \ i, \ k, \\
& CDG2: \ 1 \leq b_i(N,k) \leq b_i^{gmax}(N,k) \ \forall \ i, \ k, \\
& CDG3: \ \sum_{k=1}^{L} b_i(N,k) \geq R_i^l(N) \ \forall \ i, \\
& CDG4: -\gamma_i(N,k) \leq -C_{qarg}(N)(2^{b_i(N,k)} - 1), \ \forall \ i, \ k. \qquad (6.6)
\end{aligned}
$$

The design of $w_{g,i}^p(N,k)$ and $w_{g,i}^b(N,k)$, and the analysis of the game are left as future work.

6.6 Summary

In this chapter, we develop game theory based distributed approach to solve our proposed joint resource allocation framework that provide the optimal transmit power and rate distribution that each SU needs to employ while maintaining QoS in a multi-channel CRN. Simulation results illustrate that optimal transmit power follows reverse water filling process and optimal rate allocation is proportional to SINR. The solution obtained from proposed user-based distributed approach follows the centralized solution closely.

Chapter 7

Conclusion

In this chapter, we summarize the contributions of this dissertation and discuss possible directions for future work.

7.1 Summary of Key Contributions

We considered a competitive CRN with multiple channels available for opportunistic use by multiple secondary users. We also assumed that multiple secondary users may coexist in a channel and each secondary user (SU) can use multiple channels to satisfy their rate requirements. In this context, firstly, we presented an integrated spectrum usage model and forecasting strategy for both primary and secondary users in a competitive CRN. The effectiveness of the proposed architecture is demonstrated using experiments on both practically measured as well as simulated data. We observed that our proposed forecasting technique, not only provides a good upper bound prediction for the number of primary and secondary user, it is also robust to model parameter estimation errors. We extended the modeling and forecasting framework to the case when SU traffic is governed by Erlangian process and observed that the proposed forecasting strategy provides robust upper bound predictor for the number of secondary users.

Secondly, we assumed that scheduling is complete and SUs have identified the channels to use, we presented two centralized resource allocation frameworks named as two-stage and joint, respectively, for resource allocation to secondary users in a competitive CRN.

Unlike prior efforts, we transformed the BER constraint in both frameworks into a convex constraint in order to ensure optimality of our resulting solutions. In both frameworks, we observe that optimal transmit power follows reverse water filling process and optimal rate allocation is proportional to SINR. We found that the rate distribution in stage 2 (two-stage resource allocation framework) based on our proposed heuristic is close to optimal graph theoretic solution. We found that in terms of total power (i.e., net transmission cost), the joint resource allocation framework is more economical relative to the two-stage resource allocation framework. This is because, the joint formulation offers more degrees of freedom with the ability to adapt both power and rate simultaneously in order to achieve a certain BER. In the two-stage resource allocation framework, either power or rate is available to adapt to achieve a certain SINR or BER, respectively.

Thirdly, we borrowed ideas from social behavioral models such as Homo Egualis (HE), Homo Parochius (HP) and Homo Reciprocan (HR) models and applied it to the resource management solutions to maintain fairness among SUs in a competitive CRN setting. Specifically, we defined fairness metric for each SU in our proposed joint resource allocation framework. We incorporated dynamic fairness weight into joint resource allocation framework to maintain fairness in allocating power and rate across SUs. The dynamics of the weights are governed by social behavioral models. We observed that considering dynamic fairness weights in the resource allocation scheme provide a better system level fairness index relative to the unweighted allocation scheme.

Finally, we considered the communication overhead associated with centralized solution of proposed resource allocation frameworks and designed distributed approaches (requiring minimal or no communication overhead relative to centralized scheme). In this context, at first, we designed three user based distributed approaches based on dual theory (requiring minimal communication overhead than centralized scheme) for stage 1 of two-stage and joint resource allocation frameworks. Then, we formulated a fully distributed approach based on game theory to solve our proposed joint resource allocation framework. Simulation results

showed that the solution from each distributed implementation for both frameworks follows the centralized solution.

7.2 Future Work

Some possible future work based on the work in this dissertation is provided in this section.

- In our proposed joint resource allocation framework presented in chapter 4, we optimize two transmission parameters- transmit power and rate. As depicted in Fig. 1.3 in chapter 1, the CR has other transmission parameters such as channel, coding gain as "knobs." The BER expression including channel coding gain (γ_i^c) for M-ary QAM is expressed as

$$p_{e,i}(k) = \frac{4}{b_i(k)}\left(1-2^{-\frac{b_i(k)}{2}}\right)Q\left(\sqrt{\frac{3b_i(k)\gamma_i^c\gamma_i(k)}{(2^{b_i(k)}-1)}}\right), \forall\ i,\ k,\ \text{even } b_i(k); \quad (7.1)$$

$$p_{e,i}(k) \leq \frac{4}{b_i(k)}Q\left(\sqrt{\frac{3b_i(k)\gamma_i^c\gamma_i(k)}{(2^{b_i(k)}-1)}}\right),\ \forall\ i,\ k,\ \text{odd } b_i(k). \quad (7.2)$$

  Orthogonality factor ($\rho_{j,i}$) can also be a "knob." Resource allocation framework to determine optimal channel, transmit power ($p_i(k)$), rate ($b_i(k)$), channel coding gain (γ_i^c) and orthogonality factor ($\rho_{j,i}$) can be designed.

- In chapter 5, we presented an evolution model for fairness weights based on HP society model. It will be interesting to observe the impact of group size (i.e., number of members in the group) on group level fairness index for case 1 grouping scenario example presented in section 5.4.

- The uniqueness of NE for game formulation 1 presented in chapter 6 and convergence condition of Algorithm 6.1 can be studied.

- The analysis of game formulation presented in section 6.5 can be studied.

Bibliography

[1] D. Cabric, S. M. Mishra, D. Willkomm, R. W. Brodersen, and A. Wolisz, "A cognitive radio approach for usage of virtual unlicensed spectrum," in *White Paper*, 2004.

[2] M. Vilimpoc and M. McHenry, "Dupont circle spectrum utilization during peak hours," Shared Spectrum Company, Tech. Rep., 2006.

[3] F. C. Commission, "Spectrum policy task force report (et docket no. 02-135)," Federal Communications Commission, Tech. Rep., 2002.

[4] S. S. Company, "Spectrum occupancy report for New York city during the republican convention august 30-september 1, 2004," Shared Spectrum Company, Tech. Rep., 2005.

[5] F. C. Commission, "IEEE 802.22 working group on wireless regional area networks," Federal Communications Commission, Tech. Rep., 2006.

[6] H. Zamat, "Practical implementation of sensing receiver in cognitive radios," Ph.D. dissertation, Kansas State University, 2009.

[7] J. M. III, "Cognitive radio for flexible mobile multimedia communications," *IEEE International Workshop on Mobile Multimedia Communications*, pp. 3–10, Nov. 1999.

[8] T. R. Newman, B. A. Barker, A. M. Wyglinski, A. Agah, J. B. Evans, and G. J. Minden, "Cognitive engine implementation for wireless multicarrier transceivers," *Wiley Journal on Wireless Communications and Mobile Computing*, vol. 7, no. 9, pp. 1129–1142, Nov. 2007.

[9] A. M. Hayar, R. Knopp, and R. Pacalet, "Cognitive radio research and implementation challenges."

[10] D. Cabric and R. W. Brodersen, "Physical layer design issues unique to cognitive radio systems," in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2005, pp. 759–763.

[11] P. Pawelczak, "Technical challenges of cognitive radio-related systems," 2008.

[12] C. R. W. Thomas, "Cognitive networks," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2007.

[13] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 2004, pp. 772–776.

[14] J.-K. Lee, J.-H. Yoon, and J.-U. Kim, "A new spectral correlation approach to spectrum sensing for 802.22 WRAN system," in *International Conference on Intelligent Pervasive Computing*, Oct. 2007, pp. 101–104.

[15] T. Yucek and H. Arslan, "Spectrum characterization for opprtunistic cognitive radio systems," in *IEEE Military Communications Conference*, Oct. 2006, pp. 1–6.

[16] B. Wild and K. Ramchandran, "Detecting primary receivers for cognitive radio applications," in *IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, pp. 124–130.

[17] P. Kaligineedi and V. K. Bhargava, "Distributed detection of primary signals in fading channels for cognitive radio networks," in *IEEE Global Telecommunications Conference*, 2008, pp. 3154–3158.

[18] P. De and Y.-C. Liang, "Blind spectrum sensing algorithms for cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 5, pp. 2834–2842, 2008.

[19] J. Ma and Y. G. Li, "A probability-based spectrum sensing scheme for cognitive radio," in *IEEE International Conference on Communications*, 2008, pp. 3416 – 3420.

[20] Y. Zeng, C. L. Koh, and Y.-C. Liang, "Maximum eigenvalue detection: Theory and application," in *IEEE International Conference on Communications*, 2008, pp. 4160–4164.

[21] C. Song and Q. Zhang, "Sliding-window algorithm for asynchronous cooperative sensing in wireless cognitive networks," in *IEEE International Conference on Communications*, 2008, pp. 3432–3436.

[22] H.-S. Chen, W. Gao, and D. G. Daut, "Spectrum sensing for OFDM systems employing pilot tones and application to DVB-T OFDM," in *IEEE International Conference on Communications*, 2008, pp. 3421–3426.

[23] L. Luo and S. Roy, "A two-stage sensing technique for dynamic spectrum access," in *IEEE International Conference on Communications*, 2008, pp. 4181–4185.

[24] H. Sun, D. I. Laurenson, J. S. Thompson, and C.-X. Wang, "A novel centralized network for sensing spectrum in cognitive radio," in *IEEE International Conference on Communications*, May 2008, pp. 4186–4190.

[25] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio-part I: Two user networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2204–2213, 2007.

[26] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Secure cooperative sensing techniques for cognitive radio systems," in *IEEE International Conference on Communications*, 2008, pp. 3406–3410.

[27] M. Matsui, H. Shiba, K. Akabane, and K. Uehara, "A cooperative sensing technique with weighting based on distance between radio stations," in *14th Asia-Pacific Conference on Communications*, Oct. 2008, pp. 1–4.

[28] W. Zhang, R. K. Mallik, and K. B. Letaif, "Cooperative spectrum sensing optimization in cognitive radio networks," in *IEEE International Conference on Communications*, May 2008, pp. 3411–3415.

[29] H. Zamat and B. Natarajan, "Use of dedicated broadband sensing receiver in cognitive radio," in *IEEE International Conference on Communications*, May 2008, pp. 508–512.

[30] X. Kang, Y.-C. Liang, and A. Nallanathan, "Optimal power allocation for fading channels in cognitive radio networks under transmit and interference power constraints," in *IEEE International Conference on Communications*, May 2008, pp. 3568–3572.

[31] A. Attar, M. R. Nakhai, and A. H. Aghvami, "Cognitive radio game: A framework for efficiency, fairness and qos guarantee," in *IEEE International Conference on Communications*, May 2008, pp. 4170–4174.

[32] L. Qian, X. Li, J. Attia, and Z. Gajic, "Power control for cognitive radio ad hoc networks," in *IEEE Workshop on Local and Metropolitan Area Networks*, Jun. 2007, pp. 7–12.

[33] S. Im, H. Jeon, and H. Lee, "Autonomous distributed power control for cognitive radio networks," in *IEEE Vehicular Technology Conference*, Sep. 2008, pp. 1–5.

[34] Y. Zhu, W. Wang, T. Peng, and W. Wang, "A non-cooperative power control game considering utilization and fairness in cognitive radio network," in *International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, Aug. 2007, pp. 31–34.

[35] J. Huang, R. A. Berry, and M. L. Honig, "Auction-based spectrum sharing," *Mobile Networks and Applications*, vol. 11, pp. 405–418, 2006.

[36] Y. Xing and R. Chandramouli, "Qos constrained secondary spectrum sharing," *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 658–661, Nov. 2005.

[37] F. F. Digham, "Joint power and channel allocation for cognitive radios," in *IEEE Wireless Communications and Networking Conference*, Mar. 2008, pp. 882 – 887.

[38] F. Wang, M. Krunz, and S. Cui, "Spectrum sharing in cognitive radio networks," in *IEEE Conference on Computer Communications*, Jan. 2008, pp. 1885–1893.

[39] ——, "Price-based spectrum management in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, pp. 74–87, 2008.

[40] L. S. Pillutla and V. Krishnamurthy, "Game theoretic rate adaptation for spectrum-overlay cognitive radio networks," in *IEEE Global Telecommunications Conference*, 2008, pp. 4523–4527.

[41] J. W. Huang and V. Krishnamurthy, "Rate adaptation for cognitive radio systems with latency constraints," in *IEEE Global Telecommunications Conference*, 2008, pp. 3024–3028.

[42] Y. Wu and D. H. K. Tsang, "Distributed multichannel power allocation algorithm for spectrum sharing cognitive radio networks," in *IEEE Wireless Communications and Networking Conference*, 2008, pp. 1436–1441.

[43] Y. Xing, C. N. Mathur, M. A. Haleem, R. Chandramouli, and K. P. Subbalakshmi, "Real-time secondary spectrum sharing with QoS provisioning," in *IEEE Consumer Communications and Networking Conference*, 2006, pp. 630–634.

[44] Y. Xing, R. Chandramouli, and C. de M. Cordeiro, "Price dynamics in competitive agile spectrum access markets," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 613–621, 2007.

[45] Y. Xing, C. N. Mathur, M. A. Haleem, R. Chandramouli, and K. P. Subbalakshmi, "Dynamic spectrum access with QoS and interference temperature constraints," *IEEE Transactions on Mobile Computing*, vol. 6, no. 4, pp. 423–433, 2007.

[46] Z. Hasan, E. Hossain, C. L. Despins, and V. K. Bhargava, "Power allocation for cognitive radios based on primary user activity in an OFDM system," in *IEEE Global Telecommunications Conference*, 2008, pp. 4512–1517.

[47] V. Asghari and S. Aïssa, "Resource sharing in cognitive radio systems: Outage capacity and power allocation under soft sensing," in *IEEE Global Telecommunications Conference*, 2008, pp. 4518–4522.

[48] T. Qin and C. Leung, "A cost minimization algorithm for a multiuser OFDM cognitive radio system," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, 2007, pp. 518–521.

[49] P. Cheng, Z. Zhang, H. Huang, and P. Qiu, "A distributed algorithm for optimal resource allocation in cognitive OFDMA systems," in *IEEE International Conference on Communications*, May 2008, pp. 4718 – 4723.

[50] A. Guha and V. Ganapathy, "Power allocation schemes for cognitive radios," in *International Conference on Communication Systems Software and Middleware and Workshops*, 2008, pp. 51 – 56.

[51] E. Del Re, G. Gorni, L. Ronga, and R. Suffritti, "A power allocation strategy using game theory in cognitive radio networks," in *GameNets'09: Proceedings of the First ICST international conference on Game Theory for Networks*, 2009, pp. 117–123.

[52] Y. Song, Y. Fang, and Y. Zhang, "Stochastic channel selection in cognitive radio networks," in *IEEE Global Telecommunications Conference*, Nov. 2007, pp. 4878 – 4882.

[53] B. Atakan and O. B. Akan, "BIOlogically-inspired spectrum sharing in cognitive radio networks," in *IEEE Wireless Communications and Networking Conference*, Mar. 2007, pp. 43 – 48.

[54] Y. Xing, R. Chandramouli, S. Mangold, and S. S. Nandagopalan, "Dynamic spectrum access in open spectrum in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 627–637, Mar. 2006.

[55] Z. Ma, Z. Cao, and W. Chen, "A fair opportunistic spectrum access (FOSA) scheme in distributed cognitive radio networks," in *IEEE International Conference on Communications*, May 2008, pp. 4054–4058.

[56] C. Peng, H. Zheng, and B. Y. Zhao, "Utilization and fairness in spectrum assignment for opportunistic spectrum access," *Mobile Networks and Applications*, vol. 11, no. 4, pp. 555–576, May 2006.

[57] J. Tang, S. Misra, and G. Xue, "Joint spectrum allocation and scheduling for fair spectrum sharing in cognitive radio wireless networks," *Computer Networks*, vol. 52, no. 11, pp. 2148–2158, Apr. 2008.

[58] A. Attar, M. R. Nakhai, and A. H. Aghvami, "Cognitive radio game for secondary spectrum access problem," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 2121–2131, Apr. 2009.

[59] Y. Xing, R. Chandramouli, S. Mangold, and S. S. N, "Analysis and performance evaluation of a fair channel access protocol for open spectrum wireless networks."

[60] A. O. Nasif and B. L. Mark, "Collaborative opportunistic spectrum access in the presence of multiple transmitters," in *IEEE Global Telecommunications Conference*, 2008, pp. 3004–3008.

[61] D. Niyato and E. Hossain, "Competitive pricing for spectrum sharing in cognitive radio networks: Dynamic game, inefficiency of nash equilibrium, and collusion," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 192–202, 2008.

[62] T. Jin, C. Chigan, and Z. Tian, "Game-theoretic distributed spectrum sharing for wireless cognitive networks with heterogeneous QoS," in *IEEE Global Telecommunications Conference*, 2006, pp. 1–6.

[63] C. Zhao, M. Zou, B. Shen, B. Kim, and K. Kwak, "Cooperative spectrum allocation in centralized cognitive networks using bipartite matching," in *IEEE Global Telecommunications Conference*, 2008, pp. 2998–3003.

[64] D. Niyato and E. Hossain, "A game-theoretic approach to competitive spectrum sharing in cognitive radio networks," in *IEEE Wireless Communications and Networking Conference*, 2007, pp. 16–20.

[65] Z. Ma and Z. Cao, "Secondary user cooperation access scheme in opportunistic cognitive radio networks," in *IEEE Military Communications Conference*, Oct. 2007, pp. 4054–4058.

[66] D. Julian, M. Chiang, D. O'Neill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and Ad hoc networks," in *IEEE Conference on Computer Communications*, vol. 2, Jun. 2002, pp. 477–486.

[67] J. M. III, "Cognitive radio policy languages," in *IEEE International Conference on Communications*, 2009, pp. 1–4.

[68] N. Jesuale and B. C. Eydt, "A policy proposal to enable cognitive radio for public safety and industry in the land mobile radio bands," in *IEEE International Symposium on New frontiers in Dynamic Spectrum Access Networks*, 2007, pp. 66–11.

[69] B. Hilburn, T. R. Newman, and T. Bose, "Sector-based policy generation and enforcement for cognitive radios," in *IEEE Military Communications Conference*, 2009, pp. 1–7.

[70] J. Hwang and H. Yoon, "Dynamic spectrum management policy for cognitive radio: An analysis of implementation feasibility issues," in *IEEE International Symposium on New frontiers in Dynamic Spectrum Access Networks*, 2008, pp. 1–9.

[71] G. R. Faulhaber, "Deploying cognitive radio: Economic, legal and policy issues," *International Journal of Communications*, pp. 1114–1124, 2008.

[72] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov. 2005, pp. 269 – 278.

[73] L. Akter, B. Natarajan, and C. Scoglio, *Spectrum Usage Modeling and Forecasting in Cognitive Radio Networks,*.

[74] ——, "Modeling and forecasting secondary users acitivity in cognitive radio networks," in *17th International Conference on Computer Communications and Networks*, Aug. 2008, pp. 1–6.

[75] L. Akter and B. Natarajan, "Modeling and forecasting secondary user activity considering bulk arrival and bulk departure traffic model," in *IEEE Vehicular Technology Conference*, Sep. 2009, pp. 1–5.

[76] ——, "QoS constrained resource allocation to secondary users in cognitive radio networks," *Computer Communications*, vol. 32, no. 18, pp. 1923–1930, Dec. 2009.

[77] R. K. Jain, D. W. Chiu, and W. R. Hawe, "A quantitive measure of fairness and discrimination for resource allocation in shared computer systems," 1984.

[78] L. Akter and B. Natarajan, "Modeling fairness in resource allocation for secondary users in a competitive cognitive radio network," in *Wireless Telecommunications Symposium*, Apr. 2010.

[79] ——, "Fair resource allocation among secondary users in a QoS constrained competitive cognitive radio network," *Under Preparation*.

[80] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, 2006.

[81] L. Akter and B. Natarajan, "A two-stage power and rate allocation strategy for secondary users in cognitive radio networks," *Journal of Communications*, vol. 4, no. 10, pp. 781–789, Nov. 2009.

[82] ——, "Distributed approach for power and rate allocation to secondary users in cognitive radio networks," *Under Review*.

[83] ——, "Game theory based distributed approach for power and rate allocation to secondary users in cognitive radio networks," *Under Preparation*.

[84] P. K. Tang, Y. H. Chew, L. C. Ong, and M. K. Halder, "Performance of secondary radios in spectrum sharing with prioritized primary access," in *IEEE Military Communications Conference*, Oct. 2006, pp. 1–7.

[85] L. Kleinrock, *Queueing Systems*. John Wiley and Sons, 1975.

[86] T. Anjali, C. Scoglio, and G. Uhl, "A new scheme for traffic estimation and resource allocation for bandwidth brokers," *Computer networks*, vol. 41, no. 6, pp. 761–777, Apr. 2003.

[87] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Spectrum sensing measurements of pilot, energy and collaborative detection," in *IEEE Military Communications Conference*, Oct. 2006, pp. 1–7.

[88] M. McHenry, D. McCloskey, D. Roberson, and J. MacDonald, "Chicago spectrum occupancy measurements," Report of WIL of IIT, Tech. Rep., Jan. 2006.

[89] F. C. Commission, "Notice of proposed rules (et docket no. 03-237) on termination of proceeding: Interference temperature operation," Federal Communications Commission, Tech. Rep., May 2007.

[90] S.-W. Han, H. Kim, and Y. Han, "Distributed utility-maximization using a resource pricing power control in uplink DS-CDMA," *IEEE Communications Letters*, vol. 12, pp. 286 – 288, Apr. 2008.

[91] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation Numerical Methods*. Parentice Hall, 1989.

[92] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows: theory, algorithms, and applications*. Prentice Hall, 1993.

[93] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[94] T. Marler, "A study of multiobjective optimization methods for engineering application," Ph.D. dissertation, The University of Iowa, 2005.

[95] Y. Xing and R. Chandramouli, "Human behavior inspired cognitive radio network design," *IEEE Communications Magazine*, vol. 12, no. 6, pp. 122–127, Dec. 2008.

[96] H. Gintis, *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Behaviour*. Princeton Univ. Press, 2000.

[97] M. Rabin, "Incorporating fairness into game theory and economics," *The American Economic Reveiw*, vol. 63, no. 5, pp. 1281–1302, Dec. 1993.

[98] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "Repeated open spectrum sharing game with cheat-proof strategies," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1922–1933, Apr. 2009.

[99] T. Mathworks, "Optimization toolbox 3 users guide," sep 2007.

[100] P. Boggs and J. Tolle, "Sequential quadratic programming," *Acta Numerica*, pp. 1–51, 1995.

Appendix A

Linear Interior Point Solver

Linear programming problem is defined as

$$\begin{aligned} \text{Determine} \quad & \mathbf{x} \\ \text{To Minimize} \quad & \\ & \mathbf{f}^T\mathbf{x} \\ \text{subject to} \quad & \\ & AC1 : A_{eq}\mathbf{x} = \mathbf{b}_{eq} \\ & AC2 : A_{ineq}\mathbf{x} \leq \mathbf{b}_{ineq} \\ & AC3 : \mathbf{LB} \leq \mathbf{x} \leq \mathbf{UB} \end{aligned} \tag{A.1}$$

LIPSOL is a primal-dual interior-point method [99]. In this method, a few preprocessing steps on linear optimization problem are performed before starting actual iterative algorithm begins. The steps are

- All decision variables are bounded below by zero,
- All constraints are equalities,
- Fixed variables with equal upper and lower bounds are removed,
- Rows of all zeros in the constraint matrix are removed,

- Columns of all zeros in the constraint matrix are removed,
- The constraint matrix has full structural rank.

After preprocessing, the problem has the form

$$\begin{aligned} &\text{Determine} && \mathbf{x} \\ &\text{To Minimize} && \\ & && \mathbf{f}^T\mathbf{x} \\ &\text{subject to} && \\ & && A\mathbf{x} = \mathbf{b} \\ & && \mathbf{0} \leq \mathbf{x} \leq \mathbf{UB} \end{aligned} \tag{A.2}$$

In A.2, the upper bound constraints $AC2$ are implicitly included in the constraint matrix A. Introducing primary slack variables $\mathbf{s}$, formulation A.2 becomes

$$\begin{aligned} &\text{Minimize} && \\ & && \mathbf{f}^T\mathbf{x} \\ &\text{subject to} && \\ & && A\mathbf{x} = \mathbf{b} \\ & && \mathbf{x} + \mathbf{s} = \mathbf{UB} \\ & && \mathbf{x} \geq \mathbf{0},\ \mathbf{s} \geq \mathbf{0}. \end{aligned} \tag{A.3}$$

Formulation A.3 is referred to as *primal* problem, where $\mathbf{x}$ are primal variables and $\mathbf{s}$ are primary slack variables. The dual problem is

$$\begin{aligned} &\text{Maximize} && \\ & && \mathbf{b}^T\mathbf{y} - \mathbf{u}^T\mathbf{w} \\ &\text{subject to} && \\ & && \mathbf{A}^T\mathbf{y} - \mathbf{w} + \mathbf{z} = \mathbf{f} \\ & && \mathbf{z} \geq \mathbf{0},\ \mathbf{w} \geq \mathbf{0}. \end{aligned} \tag{A.4}$$

Here, $\mathbf{y}$ and $\mathbf{w}$ are dual variables and $\mathbf{z}$ are dual slacks. The optimality conditions for this linear program i.e., the primal equation and dual equation are

$$F(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{s},\mathbf{w}) \;=\; \begin{pmatrix} A\mathbf{x}-\mathbf{b} \\ \mathbf{x}+\mathbf{s}-\mathbf{u} \\ \mathbf{A}^T\mathbf{y}-\mathbf{w}+\mathbf{z}-\mathbf{f} \\ x_i z_i \\ s_i w_i \end{pmatrix} = \mathbf{0}, \tag{A.5}$$

where, $x_i z_i$ and $s_i w_i$ denote component-wise multiplication. The quadratic equations $x_i z_i = 0$ and $s_i w_i = 0$ are called *complementarity* conditions for the linear program and the other linear conditions are called the *feasibility* conditions. The quantity $\mathbf{x}^T\mathbf{z}+\mathbf{s}^T\mathbf{w}$ is the *duality gap*, which measures the complementarity portion of F when $\mathbf{x},\ \mathbf{z},\ \mathbf{s},\ \mathbf{w} \geq \mathbf{0}$.

The algorithm is called *primal-dual algorithm* as both primal and dual programs are solved simultaneously. It can be considered a Newton-like method, applied to linear-quadratic system $F(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{s},\mathbf{w}) = \mathbf{0}$ in A.5, while at the same time keeping the iterates $\mathbf{x}$, $\mathbf{z}$, $\mathbf{s}$ and $\mathbf{w}$ positive, hence the name interior-point method.

Appendix B

Sequential Quadratic Programming

The nonlinear programming problem (NLP) is defined as

$$\begin{aligned} &\text{Determine} && \mathbf{x} \\ &\text{To Minimize} && \\ & && f(\mathbf{x}) \\ &\text{subject to} && \\ & && g_i(\mathbf{x}) = \mathbf{0},\ i = 1,\ \cdots,\ m_e \\ & && g_i(\mathbf{x}) \leq \mathbf{0},\ i = m_e,\ \cdots,\ m \end{aligned} \tag{B.1}$$

where, $f:\ \mathcal{R}^n \rightarrow \mathcal{R}$ and $\mathbf{g}:\ \mathcal{R}^n \rightarrow \mathcal{R}^m$. We assume that the NLP (B.1) has atleast one nonlinear constraint.

Sequential Quadratic Programming (SQP) has become the most successful method for solving nonlinearly constrained optimization problems [99, 100]. The basic idea of SQP is to model NLP at a given approximate solution, say, $\mathbf{x}^k$, by a quadratic programming (QP) subproblem and then to use the solution to this subproblem to construct a better approximation $\mathbf{x}^{k+1}$. The process is repeated to create a sequence of approximations until the approximation converges to a solution $\mathbf{x}^*$. The QP subproblem is formed based on a quadratic approximation of the Lagrangian function of NLP if it has nonlinear constraints. At each major iteration, an approximation is made of the Hessian of the Lagrangian function using a quasi-Newton updating method. This is then used to generate a QP subproblem

whose solution is used to form a search direction for a line search procedure.

The Lagrangian of NLP in B.1 is

$$\mathcal{L}_{NLP} = f(\mathbf{x}) + \boldsymbol{\omega}^T \mathbf{g}(\mathbf{x}). \tag{B.2}$$

The QP subproblem at a given approximate solution $\mathbf{x}^k$ and positive definite approximation of the Hessian matrix H_k (of the Lagrangian function $\mathcal{L}_{NLP}$) is

$$\begin{aligned} \text{Determine} \quad & \mathbf{d} \\ \text{To Minimize} \quad & \\ & \frac{1}{2}\mathbf{d}^T H_k \mathbf{d} + \nabla f(\mathbf{x}_k)^T \mathbf{d} \\ \text{subject to} \quad & \\ & \nabla g_i(\mathbf{x}_k)^T \mathbf{d} + g_i(\mathbf{x}_k) = 0 \; i = 1, \; \cdots, \; m_e \\ & \nabla g_i(\mathbf{x}_k)^T \mathbf{d} + g_i(\mathbf{x}_k) \leq 0 \; i = m_e, \; \cdots, \; m. \end{aligned} \tag{B.3}$$

This subproblem can be solved using any QP algorithm. The solution is used to form a new approximation

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \tag{B.4}$$

where, α_k is step length parameter and is determined by an appropriate line search procedure so that a sufficient decrease in a merit function is obtained. H_k can be updated by any of the quasi-Newton methods. With the above background, the SQP algorithm is summarized in Algorithm B.1.

Algorithm B.1: SQP algorithm

Initialization: $\boldsymbol{x}(\mathbf{0})$, $\boldsymbol{\omega}(\mathbf{0})$, $H(0)$;
Choose merit function $\psi(\mathbf{x})$ and compute $\psi(\mathbf{x}(0))$;
while termination criterion is not true **do**
    Solve QP subproblem (B.3) to obtain $\mathbf{d}_k$;
    Choose step length α_k so that $\psi(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < \psi(\mathbf{x}_k)$;
    Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$;
    Compute H_k;
end while

In our simulation, we used (i) active set strategy to solve QP subproblem, (ii) Broyden-Fletcher-Goldfarb-Shanno (BFGS) method to update Hessian of Lagrangian, and (iii) the merit function

$$\psi(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m_e} r_i g_i(\mathbf{x}) + \sum_{i=m_e+1}^{m} r_i \; max\{0, g_i(\mathbf{x})\}, \tag{B.5}$$

where, r_i is penalty parameter and

$$r_i = r_{k+1,i} = max_i \left\{ \omega_i \frac{1}{2}(r_{k,i} + \omega_i) \right\}, \; \forall \; i. \tag{B.6}$$

This choice of r_i allow positive contribution from constraints that has just become active. r_i is initially set to

$$r_i = \frac{\nabla f(\mathbf{x})}{\nabla g_i(\mathbf{x})}, \tag{B.7}$$

where, $||.||$ represents Euclidean norm.