UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

RESOURCE ALLOCATION OPTIMISATION IN HETEROGENEOUS COGNITIVE RADIO NETWORKS

by

Babatunde Seun Awoyemi

Submitted in partial fulfillment of the requirements for the degree
Philosophiae Doctor (Electronic Engineering)

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment and Information Technology

UNIVERSITY OF PRETORIA

January 2017

SUMMARY

RESOURCE ALLOCATION OPTIMISATION IN HETEROGENEOUS COGNITIVE RADIO NETWORKS

by

Babatunde Seun Awoyemi

Promoter:	Prof. B. T. Maharaj
Co-promoter:	Prof. A. S. Alfa
Department:	Electrical, Electronic and Computer Engineering
University:	University of Pretoria
Degree:	Philosophiae Doctor (Electronic Engineering)
Keywords:	Buffered systems, cognitive radio networks, cooperative diversity, heterogeneous networks, heuristics, linear and non-linear programming, meta-heuristics, optimisation, orthogonal frequency division multiple access, resource allocation, queueing theory

Cognitive radio networks (CRN) have been tipped as one of the most promising paradigms for next generation wireless communication, due primarily to its huge promise of mitigating the spectrum scarcity challenge. To help achieve this promise, CRN develop mechanisms that permit spectrum spaces to be allocated to, and used by more than one user, either simultaneously or opportunistically, under certain preconditions. However, because of various limitations associated with CRN, spectrum and other resources available for use in CRN are usually very scarce. Developing appropriate models that can efficiently utilise the scarce resources in a manner that is fair, among its numerous and diverse users, is required in order to achieve the utmost for CRN. 'Resource allocation (RA) in CRN' describes how such models can be developed and analysed.

In developing appropriate RA models for CRN, factors that can limit the realisation of optimal solutions have to be identified and addressed; otherwise, the promised improvement in spectrum/resource utilisation would be seriously undermined. In this thesis, by a careful examination of relevant literature,

the most critical limitations to RA optimisation in CRN are identified and studied, and appropriate solution models that address such limitations are investigated and proffered.

One such problem, identified as a potential limitation to achieving optimality in its RA solutions, is the problem of heterogeneity in CRN. Although it is indeed the more realistic consideration, introducing heterogeneity into RA in CRN exacerbates the complex nature of RA problems. In the study, three broad classifications of heterogeneity, applicable to CRN, are identified; heterogeneous networks, channels and users. RA models that incorporate these heterogeneous considerations are then developed and analysed. By studying their structures, the complex RA problems are smartly reformulated as integer linear programming problems and solved using classical optimisation. This smart move makes it possible to achieve optimality in the RA solutions for heterogeneous CRN.

Another serious limitation to achieving optimality in RA for CRN is the strictness in the level of permissible interference to the primary users (PUs) due to the activities of the secondary users (SUs). To mitigate this problem, the concept of cooperative diversity is investigated and employed. In the cooperative model, the SUs, by assisting each other in relaying their data, reduce their level of interference to PUs significantly, thus achieving greater results in the RA solutions. Furthermore, an iterative-based heuristic is developed that solves the RA optimisation problem timeously and efficiently, thereby minimising network complexity. Although results obtained from the heuristic are only suboptimal, the gains in terms of reduction in computations and time make the idea worthwhile, especially when considering large networks.

The final problem identified and addressed is the limiting effect of long waiting time (delay) on the RA and overall productivity of CRN. To address this problem, queueing theory is investigated and employed. The queueing model developed and analysed helps to improve both the blocking probability as well as the system throughput, thus achieving significant improvement in the RA solutions for CRN.

Since RA is an essential pivot on which the CRN's productivity revolves, this thesis, by providing viable solutions to the most debilitating problems in RA for CRN, stands out as an indispensable contribution to helping CRN realise its much-proclaimed promises.

LIST OF ABBREVIATIONS

AMC	Adaptive modulation and coding
BnB	Branch-and-bound
BS	Base station
BPSK	Binary phase shift keying
CR; CRs	Cognitive radio; Cognitive radios
CRN	Cognitive radio networks
CSU	Cooperative secondary user
D	Destination terminal
DSA	Dynamic spectrum access
FCC	Federal communication commission
HetNet	Heterogeneous networks
ILP	Integer linear programming
LMI	Linear matrix inequalities
LP	Linear programming
LTE	Long term evolution
MARA	Margin adaptive resource allocation
MATLAB	Matrix laboratory
MIMO	Multiple-input multiple output
MINLP	Mixed integer non-linear programming
NLP	Non-linear programming
NP	Non-deterministic polynomial-time
NRT	Non-real time
OfCom	Office of communications
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
PU; PUs	Primary user; Primary users
QAM	Quadrature amplitude modulation
QoS	Quality of service
RA	Resource allocation
RT	Real time
RARA	Rate adaptive resource allocation
SINR	Signal-to-interference-and-noise ratio
SNR	Signal-to-noise

SSU	Source secondary user
SU; SUs	Secondary user; Secondary users
SUBS	Secondary user base station
xG	Next generation
YALMIP	Yet another LMI parser

LIST OF SYMBOL NOTATIONS

$K;k$	Total number of heterogeneous secondary users; k is used to identify a particular user
K_1	Number of category one secondary users; number of category two users is $K-K_1$ (or K_2)
$N;n$	Number of available OFDMA subchannels; n is used to identify a particular subchannel
L	Number of primary users
$H^c_{k,n}$	Channel gain between SUBS and SU at the kth SU over the nth subchannel
$H^s_{k,n}$	Channel gain from SSU to CSU at the kth SU over the nth subchannel
$H^r_{k,n}$	Channel gain from CSU to D at the kth SU over the nth subchannel
$P^s_{k,n}$	Transmit power from SSU to CSU at the kth SU over the nth subchannel
$P^r_{k,n}$	Power from CSU to D at the kth SU over the nth subchannel
$c_{k,n}$	Data rate at the kth SU over the nth subchannel
$c^s_{k,n}$	Data rate from SSU to CSU at the kth SU over the nth subchannel
$c^r_{k,n}$	Data rate from CSU to D at the kth SU over the nth subchannel
$c_{k,n,D}$	Data rate at the kth SU over the nth subchannel for direct transmission
$c_{k,n,C}$	Data rate at the kth SU over the nth subchannel when cooperation is employed
$P_{k,n,D}$	Transmit power at the kth SU over the nth subchannel for direct transmission
$P_{k,n,C}$	Transmit power at the kth SU over the nth subchannel when cooperation is employed
$P_{\max}$	Total transmit power at SUBS
$\boldsymbol{x}$	Bit allocation vector
$\boldsymbol{x}_I;\boldsymbol{x}_{II}$	Bit allocation vector for a category one SU; bit allocation vector for a category two SU
$\boldsymbol{b}$	Modulation order vector
$\boldsymbol{b}_I;\boldsymbol{b}_{II}$	Modulation order vector for category one SU; modulation order vector for a category two SU
$\boldsymbol{p}$	Power transmission vector
$\boldsymbol{p_D}$	Power transmission vector for direct communication
$\boldsymbol{p}_C$	Power transmission vector for cooperative communication
R_K	Minimum rate demand of a category one SU
γ_K	Proportional rate constraint for a category two SU

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION **1**

1.1 PROBLEM STATEMENT 4

1.1.1 Context of the problem 4

1.1.2 Research gap 5

1.2 RESEARCH OBJECTIVES 6

1.3 HYPOTHESIS AND APPROACH 7

1.4 RESEARCH CONTRIBUTIONS AND OUTPUTS 8

1.5 DEFINITION OF TERMS 10

1.6 OVERVIEW OF THESIS 11

CHAPTER 2 A SURVEY OF RESOURCE ALLOCATION IN COGNITIVE RADIO NETWORKS **13**

2.1 CHAPTER OBJECTIVES 13

2.2 AN OVERVIEW OF COGNITIVE RADIO NETWORKS 13

2.3 ARCHITECTURE OF COGNITIVE RADIO NETWORKS 15

2.4 RESOURCE ALLOCATION IN COGNITIVE RADIO NETWORKS 18

2.5 RESOURCE ALLOCATION PROBLEM FORMULATION IN COGNITIVE RADIO NETWORKS 21

2.6 CLASSIFICATION OF RESOURCE ALLOCATION SOLUTION APPROACHES FOR COGNITIVE RADIO NETWORKS 25

2.6.1 Solutions using classical optimisation 25

2.6.2 Solutions by studying problem structure 26

2.6.3 Solutions by heuristics or meta-heuristics 30

2.6.4 Solutions by multi-objective optimisation 33

2.6.5 Solutions through soft computing 34

2.7 OBSERVATION AND OPEN-ENDED PROBLEMS IN RESOURCE ALLOCATION FOR COGNITIVE RADIO NETWORKS 34
2.8 CONCLUSION 37

CHAPTER 3 RESOURCE ALLOCATION SOLUTION MODELS FOR HETEROGENEOUS COGNITIVE RADIO NETWORKS 39

3.1 CHAPTER OVERVIEW 39
3.2 BACKGROUND 40
3.3 HETEROGENEITY IN COGNITIVE RADIO NETWORKS 40
3.4 RELATED LITERATURE ON HETEROGENEITY IN COGNITIVE RADIO NETWORKS 43
3.5 SYSTEM MODEL 45
3.5.1 General representation of the resource allocation formulation for heterogeneous cognitive radio networks 47
3.5.2 Classification based on minimum rate requirement 48
3.5.3 Classification based on user priority or sensitivity 53
3.5.4 Classification based on delay tolerance 56
3.6 RESULTS AND DISCUSSION 58
3.6.1 Results based on minimum data rate classification 59
3.6.2 Results based on priority and sensitivity classifications 63
3.6.3 Results based on delay tolerance classification 67
3.6.4 Effects of weight on resource allocation in heterogeneous cognitive radio networks 69
3.7 CONCLUSION 71

CHAPTER 4 RESOURCE ALLOCATION IN HETEROGENEOUS COOPERATIVE COGNITIVE RADIO NETWORKS 73

4.1 CHAPTER OVERVIEW 73
4.2 BACKGROUND 74
4.3 RELATED LITERATURE ON COOPERATIVE DIVERSITY IN COGNITIVE RADIO NETWORKS 75
4.4 SYSTEM MODEL 78
4.5 RESOURCE ALLOCATION PROBLEM FORMULATION 81
4.6 REFORMULATION AS AN INTEGER LINEAR PROGRAMMING PROBLEM 83

4.7 ITERATIVE-BASED HEURISTIC 87
4.7.1 Subchannel allocation 87
4.7.2 Iterative bit and power allocation 88
4.8 RESULTS AND DISCUSSION 92
4.9 CONCLUSION 99

CHAPTER 5 RESOURCE ALLOCATION SOLUTION FOR HETEROGENEOUS BUFFERED COGNITIVE RADIO NETWORKS 101
5.1 CHAPTER OVERVIEW 101
5.2 BACKGROUND 102
5.3 RELATED LITERATURE ON BUFFERING IN RESOURCE ALLOCATION FOR COGNITIVE RADIO NETWORKS 103
5.4 SYSTEM MODEL 105
5.4.1 Queueing model 105
5.4.2 Analysis of model 107
5.5 RESULTS AND DISCUSSION 115
5.6 CONCLUSION 120

CHAPTER 6 CONCLUSION 121
6.1 SUMMARY 121
6.2 RECOMMENDATIONS FOR FUTURE WORK 123
6.2.1 Recommendations based on heterogeneous considerations 123
6.2.2 Recommendations based on the use of cooperative diversity for mitigating interference to PUs 124
6.2.3 Recommendations based on the use of queueing models for addressing delay problems 125

REFERENCES 126

ADDENDUM A STATE REDUCTION ALGORITHM 141
A.1 THE GRASSMAN-TAKSAR-HEYMAN (GTH) ALGORITHM 141

CHAPTER 1 INTRODUCTION

Wireless communication has over the years gained global acceptance, becoming the most integral part of modern telecommunication [1]. Its tremendous achievements in terms of ubiquity, mobility, massive 'unlimited' coverage capability, sustained reduction in component size (portability) and improved service costs (affordability), and a host of other positives have continued to make it a preferred choice over wired communication. As a result, demands for wireless applications and usage have been on an explosive exponential rise. It is estimated that by year 2019, mobile communication through phones alone would have reached a staggering 5.07 billion of the world population! [2]. The statistics about wireless and mobile network penetration in Africa have been equally impressive. The expectations are that, in the nearest future, the likelihood of a higher growth in capacities will result in an even wider coverage or reach. This will invariably imply an increase in wireless and mobile broadband demands in many countries of Africa and other parts of the world [3]. In other words, the recent but steady proliferation of 'wireless', if sustained, is set to result in some immense, almost insatiable 'outbreak' in wireless communication operations worldwide.

While this terrific increase in wireless capabilities and demands, in itself, seems harmless (maybe it is even delightful because of its amazing promise of helping to shrink the digital divide), an undeniable and worrisome aftermath of it is a corresponding steady increase in spectrum demand to accommodate the rise. The spectrum, being a fixed and limited resource, has continued to experience a rising demand for more and more portions of it by the numerous interest groups seeking frequency allocations to execute their wireless operations. These demands for spectrum availability have generated a kind of 'overcrowding' of the spectrum space over which wireless communications are usually designed to operate. There is now globally, as it were, a seeming scarceness of radio-frequency spectrum to accommodate these ever-growing wireless communication needs [4]. There are usually consequences and there is always a price to pay for embracing any good, innovative or inventive conception/consideration after all!

The spectrum scarcity, gradually becoming a serious limitation, is threatening to hamper the possibilities and promises of wireless communication. Its threat has therefore triggered meaningful investigations into the present spectrum allotment and usage designs. Regulatory bodies such as the United States' Federal Communication Commission (FCC), United Kingdom's Independent Regulator and Competition Authority, known as Office of Communications (OfCom), and others have been compelled to carry out extensive research on the general usage patterns of the occupied spectrum. The outcome of the investigations clearly revealed that the current, static arrangement of exclusive allocation of certain frequency bands to specific, statutory occupiers is oddly inefficient. By that kind of arrangement, the spectrum space only looked fully occupied but in reality, the utilisation level is quite unimpressive. Large portions of the spectrum, which hitherto have been assumed to be meaningfully engaged by the rightful owners to which they have been allocated, are actually underutilised, either in entirety or at certain time durations [5–7]. The spectrum problem, it seems therefore, may not necessarily be a scarcity problem after all but an inefficient utilisation problem, as these investigations portend.

In light of the aforementioned, it has become imperative to find solutions to the spectrum scarcity/underutilisation problem in order to maintain the drive towards attaining and/or sustaining the goal of achieving ubiquitous global networking and connectivity through wireless communication. Several ideas and propositions have continued to emerge, but the most considerably consistent and which is also currently in the forefront for immediate incorporation, is the call for the introduction of a new spectrum management paradigm - a dynamic spectrum access (DSA) arrangement [8]. With this DSA, spectrum allocations would no longer be static or be an exclusive property of a single occupier, but by some mutual agreement, one or more authorised owners can equally access and use the same spectrum space for their own communication. There will of necessity be rules and regulations guiding this kind of arrangement, and prioritising 'ownership' and 'accessibility rights' will have to be unequivocal, but as recent empirical results have demonstrated, this new arrangement can expressively revamp the spectrum utilisation quagmire [9, 10].

The emergence and deservedly growing acceptance of the DSA has brought with it some new and promising wireless communication paradigms. Of utmost significance, and which has attracted keen interest of various stakeholders, both in academia and industry, is the cognitive radio network (CRN).

CRN ride on the possibilities of the DSA in developing new communication models in which multiple users can communicate separately or alongside each other, over two or more distinct networks, on the same frequency space. Pioneering works on CRN have generally been accredited to Mitola [11]. In his original work, he described a new kind of radios, called software defined radios (and later referred to as cognitive or Mitola radios) with the capability to learn from their environment and intelligently and dynamically adjust their operating parameters, based on what has been learned, to achieve better communication [12]. In other words, a cognitive radio (CR) should be able to, among other things, dynamically adjust its frequency spectrum of operation to access/use a new frequency space to suit its new environment or to meet its new demands. The functionality of CRs therefore greatly depends on the advent and eventual implementation of the DSA. Although CRs, in their ultimate design, would be capable of achieving much more than just dynamic access and the ability to employ free or underutilised spectrum spaces for their communication, the important note at this point is that the institution and embrace of DSA has been the major key-player in the development of CRs and CRN [13].

The CRN's huge promises, particularly that of providing improved wireless capabilities by optimising the use of the scarce spectrum resource, has made it gain immense attention and recognition as the likely, most-reliable paradigm for accomplishing next-generation (xG) wireless communication possibilities [14]. However, despite its enormous promises, several issues about its effectiveness and concerns about possible limitations to its productivity have equally begun to emerge. Of the numerous issues raised by researchers and stakeholders, the most significant, it seems, is the concern that CRN, which by design is predominantly an opportunistic network, may never be able to achieve much, due to the fact that its spectrum resource (alongside other resources) available for its use are very limited and non-guaranteed. It therefore means that, unless mechanisms that can efficiently utilise CRN resources are developed, and the possible limitations to its resourcefulness identified and addressed, it will be extremely difficult, almost impossible, for CRN to achieve its ends. That is the main motivation for the research work carried out and presented in this thesis. The key focus of the thesis is therefore to identify the significant limitations to resource availability and usability in CRN and to address those limitations. The solution models developed in the thesis to address the limitations in the resourcefulness of CRN stands out as invaluable contributions in the field of CRN, and in making it worth its while.

1.1 PROBLEM STATEMENT

Although significant advancements have been made in exploring and even experimentally deploying some prototypes of CRN, there are still a number of open-ended problems that require adequate investigation, if the promises of CRN are ever to materialise. One such problem, of high significance, is in designing methods for achieving the utmost in the allocation of the limited resources on which CRN usually have to build communication. It has already been well established that the amount of resources available for use in CRN is generally limited and that, the demands of users in CRN are usually large and diverse. Hence, unless adequate methods for efficiently utilising the resources of CRN are devised and the limiting problems addressed, it would be very difficult for CRN to achieve meaningful results. The important research question that is sought to be addressed in this thesis is underscored thus: 'how can the limited or scarce resources available to CRN be best administered so as to meet the varying demands of different users in order to achieve maximum utility and productivity of the overall network?' The problem statement is therefore 'to identify and proffer solutions to problems associated with the allocation of the limited resources of CRN in order to meet the diverse needs of the various users in the network, so as to optimise the overall CRN productivity'. The term, 'resource allocation (RA) in heterogeneous CRN', which essentially defines and describes this problem, is thus the focus of the thesis.

1.1.1 Context of the problem

While CRN has gained attention because of its much-proclaimed promise of providing a new and improved way of maximising the use of the spectrum space, a potentially crippling barrier to its effectiveness is the apprehension over its productivity due to the scarceness in the available resources on which it relies for its communication. Generally, in almost all of the wireless communication paradigms, resources such as bandwidth, transmit power, modulation schemes, data rate and frequency spectrum, used in actualising their aims are scarce and limited. Worse still, because CRN is an opportunistic network design and on which several constraints are imposed, this makes the limitations in its resource availability very debilitating. Hence, developing and analysing mechanisms for achieving the utmost from the limited resources of CRN is a necessity, if the promises and possibilities it bears are ever to be realised. That is the whole essence of studying RA in CRN.

Although attempts at developing appropriate RA models for most other conventional wireless communication paradigms may have made bold progress and materials/methods of them already fairly abundant, the case of CRN is somewhat different, due to its being a new technology, and also because of its numerous peculiarities. It is therefore noted that RA models in conventional wireless networks, even though available, may not be directly adoptable by CRN. However, it can be expressed with some optimism that, by careful modifications and adaptations, the already established RA models for some wireless communication systems, like the orthogonal frequency division multiple access (OFDMA), may indeed be useful in addressing RA problems in CRN as well, particularly if certain peculiar limitations of CRN are adequately taken care of in such adaptations. Invariably, since it is necessary to develop RA models that can optimise the utilisation of CRN resources, it is therefore imperative to understudy models developed for solving similar RA problems in older but fairly related wireless networks and to seek how to adapt them in solving RA problems in CRN, taking into consideration the differences and/or intricacies associated with CRN.

In carrying out investigations and developing models for RA in CRN, achieving optimality demands solving the important task of identifying what limitations exist, due to the peculiarities of CRN, that can hamper its resourcefulness. Essentially, such limitations, once spotted, have to be incorporated into their RA models and appropriate solutions investigated. In this thesis, three most significant limitations to RA optimisation in CRN are identified and studied, and efficient solution models that address these limitations are developed and analysed. The three limitations identified are: the fact that CRN is indeed heterogeneous in nature, the stringent level of permissible interference to primary users (PUs) of the network and, the problem of time delay and its associated data queues in the secondary users (SUs) network. Obtaining solutions to the identified problems, thus providing RA models that achieve optimality in CRN's productivity despite the crippling limitations, form the nucleus of this thesis.

1.1.2 Research gap

While there seems to be a fairly considerable amount of work in the literature on RA in CRN already, a comprehensive survey of recent literature, as carried out in the course of this research, reveals that most RA models developed and studied by researchers within the field have omitted or neglected the most significant limiting factors that can make the realisation of optimality in resource availability and usability in CRN improbable. Even in some works where any or some of these limiting problems (like CRN's heterogeneity and PUs' stringent interference limitations) have been identified, not much has been suggested or investigated in mitigating the effects of these limitations. Hence, a significant research gap is identified in this regard. Since CRN, because of the scarceness in its resources, could only meaningfully realise its promises when it is possible to achieve the utmost in the utilisation of its available resources, identifying significant limiting factors to RA optimisation and developing solution models that address such limitations is therefore a key enabler to CRN's eventual acceptance and recognition as the ideal prototype for future wireless communication.

1.2 RESEARCH OBJECTIVES

The following are the objectives achieved in the course of the research:

- Identifying and investigating the critical limiting problems associated with RA in CRN. The problems identified are indeed the factors that limit or hinder the realisation of optimality in the RA solutions for CRN.

- Developing system models for RA in CRN that incorporate the identified limiting factors and studying solution models that address them. The ability to develop system models that address those limitations makes it possible to investigate viable solutions to RA problems in CRN.

- Exploring various optimisation techniques for solving the RA problems in CRN while considering and incorporating the identified limitations in the models. By this exploration, various optimisation techniques/approaches that exploit problem structures and capitalise on such to achieve solutions are developed, making it possible to obtain optimal solutions. Solutions for the developed RA models provided through simulations and analysis are validated by obtaining and comparing with comparative results from other related works in the literature.

- Developing an appropriate iterative-based heuristic to solve the RA problems in CRN. The heuristic gives solutions with much reduced complexity and computational demand while still achieving near-optimality, when its performance is compared to the solutions provided using classical optimisation, as initially realised.

- Developing models that explore concepts of cooperative diversity and queueing theory to address the critical limitations in achieving optimality in RA for CRN. The concepts, when introduced and explored, helped in ameliorating the debilitating effects of the limiting factors identified, thus making it possible to obtain optimal solutions and to significantly improve the resourcefulness and productivity of CRN.

1.3 HYPOTHESIS AND APPROACH

The hypothesis for this research work is stated thus, 'the promises of CRN can be better realised if potential limiting factors in achieving optimality in its RA are identified, and viable solutions provided to address such limitations'. The null hypothesis would then be that 'identifying and addressing the limiting factors in achieving optimal results in RA will not necessarily improve the CRN's performance and/or thereby making it better equipped towards achieving its promises'. The hypothesis being true, as the results presented in subsequent chapters of the thesis suggests, is the basis for the proposition that 'by developing solution models that recognise and address the various limitations of RA in CRN, significant improvement in its productivity can be realised'. This is certainly a positive advancement in the field of CRN.

The research work has followed the well-established pattern for carrying out technical research in the field of electronic engineering and telecommunications, particularly in wireless communication. The various stages in which this research work was carried out is thus highlighted:

- **Literature survey**: The first part of the research work was dedicated to exploring in-depth the concept of CRN, studying RA problems in CRN in comparison with other wireless communication paradigms to identify peculiarities and opportunities, and studying optimisation and other important and relevant bodies of knowledge such as cooperative diversity and queueing theory. The desired outcome at that stage was not only to have a well-grounded foundation of the subject matter but also to identify research gaps and define a focus and direction for the research work carried out. The findings from the investigations at that stage formed a considerable amount of the information provided in chapter two of this thesis.

- **System modelling**: Network and system models are important in solving engineering problems.

In this research work, various system models were developed for achieving RA in CRN. The models incorporated several identified factors that limit the realisation of optimal solutions in the RA formulations. The solution models developed employ concepts of cooperative diversity and queueing theory in solving specific, identified, limiting problems associated with RA in CRN. These models, capturing and addressing each of these limiting challenges, were thoroughly analysed and results obtained are presented and discussed in the various chapters of the thesis.

- **Simulation and numerical analysis**: The system models developed were simulated using MATLAB software, while an optimisation toolbox called 'yet another linear matrix inequalities parser (YALMIP)', developed in [15], was used in solving the optimisation problems. Numerical analyses of the performance of the RA models were carried out. Also, a heuristic for solving RA optimisation problems in heterogeneous cooperative CRN was developed and the results obtained compared with numerical results by using classical optimisation. The concept of queueing theory was further employed in addressing the aspects of the heterogeneous CRN where buffering and delay considerations are a major limitation to achieving optimal RA solutions.

- **Verification and validation of results**: Results obtained through simulations were validated by numerical analyses. Also, verification and validation of results were carried out by obtaining and comparing comparative results from related works in the literature.

- **Thesis write-up:** The research work was concluded with a detailed documentation of the various findings of the research and presented in this thesis.

1.4 RESEARCH CONTRIBUTIONS AND OUTPUTS

The following contributions have been made to the body of knowledge on RA in CRN in the course of the research work:

- A detailed study on RA in CRN exposed the most significant limitations to its resourcefulness and productivity. The identified problems are indeed critical in that, if not addressed, they stand as potential barriers to CRN achieving its promise of providing a viable solution to the spectrum challenge.

- In the research, not only were the the limitations to achieving optimality in RA for CRN exposed, but also, adequate solutions for mitigating their effects on CRN's productivity are proffered. The solution models developed stand out as great contributions to the body of knowledge on CRN, and in helping it achieve its ends.

- In addressing the limiting problem of heterogeneity in RA for CRN, a comprehensive analysis of several techniques for addressing different RA problems with heterogeneous considerations in CRN resulted in the development of a solution approach that can be very adaptable to all kinds of CRN heterogeneity, thereby providing a strong basis for establishing a general, all-inclusive solution approach for RA in heterogeneous CRN.

- The debilitating problem of strictness in the amount of permissible interference to the PUs, resulting in poor productivity of RA in CRN, was addressed in this research. The solution approach provided uses the concept of cooperative diversity in achieving the desired goal of mitigating the PUs interference limitation, thus optimising RA in heterogeneous CRN.

- The limiting problem of delay and buffering of data, which makes it very difficult to achieve optimality in RA for CRN, was addressed by employing queueing theory.

The contributions from this research work have been presented and/or published (or are currently under review for publication) as full-length papers in reputable international conference proceedings and articles in peer-reviewed journals in the field of telecommunications and electronic engineering. A list of the publications from the research work is given below:

Full-length papers in peer-reviewed conference proceedings:

1. B. S. Awoyemi, B. T. Maharaj and A. S. Alfa, "Resource allocation for heterogeneous cognitive radio networks," *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, New Orleans, LA, 2015, pp. 1759-1763.

2. B. S. Awoyemi, B. T. Maharaj and A. S. Alfa, "QoS provisioning in heterogeneous cognitive radio networks through dynamic resource allocation," *AFRICON, 2015 IEEE*, Addis Ababa,

2015, pp. 1-6.

3. B. S. Awoyemi, B. T. Maharaj and A. S. Alfa, "Solving resource allocation problems in heterogeneous cognitive radio networks," *Southern African Telecommunication Networks and Applications Conference (SATNAC), 2015*, Hermanus, 2015, pp. 1-5.

Peer-reviewed, ISI rated journal articles:

1. B. S. Awoyemi, B. T. Maharaj and A. S. Alfa, "Solving resource allocation problems in cognitive radio networks: a survey," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 176-190, July 2016.

2. B. S. Awoyemi, B. T. Maharaj and A. S. Alfa, "Optimal resource allocation solutions for heterogeneous cognitive radio networks," *Digital Communications and Networks*, vol. 2016, no. 1, pp. 1-14 , Nov. 2016.

3. B. S. Awoyemi, B. T. Maharaj and A. S. Alfa, "Resource allocation in heterogeneous cooperative cognitive radio networks," *International Journal of Communication Systems*, vol. 2016, no. 1, pp. 3247-3261, Nov. 2016.

4. B. S. Awoyemi, B. T. Maharaj and A. S. Alfa, "Resource allocation in heterogeneous buffered cognitive radio networks," *Journal of Communications and Networks*, **(under review).**

1.5 DEFINITION OF TERMS

Several terms are used in the thesis to describe various concepts and aspects of RA in CRN being considered. These terms are hereby defined:

- **Productivity**: This is the measure of the efficiency or total output (yield) of the CRN.

- **Utility**: This refers to the degree of usefulness that the CRN achieves.

- **Resourcefulness**: This refers to the ability of CRN to find quick and smart ways to overcome its various limitations.

- **Capacity**: This is the maximum achievable output of the CRN.

- **Throughput**: This is the total amount of data per unit time (total data rate) that is successfully transmitted by the network, and is usually measured in bits per second (bps).

- **Optimality**: This refers to the best or most effective result(s) obtainable, based on current or prevalent conditions (constraints) under which the CRN operates. In the context the thesis, optimality of CRN is achieved when the best performance (measured from the performance metrics of average data rates, throughput, outage probability, etc.) is realised, given the prevailing network conditions (that is, the available resources and the various constraints being considered).

1.6 OVERVIEW OF THESIS

The remainder of the thesis is structured as follows:

Chapter two focuses on presenting detailed background knowledge on the subject matter through a well-thought-out survey on RA in CRN. In the chapter, relevant aspects of, and recent works on, RA in CRN in the literature, are critically examined. The survey establishes the aspects of RA in CRN that have been properly addressed, but also identifies limiting factors to the optimal resourcefulness of the various RA solution models. The analysis of solutions to the identified limitations in RA optimisation for CRN forms the important direction for the research. It also helped shape the structure of the thesis, as the various studies conducted are logically presented in subsequent chapters of the thesis.

In Chapter three, models that capture and address the heterogeneity problem in RA for CRN are developed and analysed. While the heterogeneous systems presented may not be completely exhaustive, the main goal is to present relevant models that succinctly but sufficiently capture the most important categorisations of heterogeneity, as applicable to CRN. Furthermore, by providing solutions to all the identified heterogeneous system classifications presented using very similar optimisation concepts, it is argued that the solution approaches provided are indeed adoptable to and/or adaptable for solving various kinds of RA problems in heterogeneous CRN. This implies that the solution approach can indeed be generalised for solving RA problems with heterogeneous considerations in CRN.

Chapter four introduces cooperative diversity as a means of addressing a fundamental problem limiting the capacity of the RA model in heterogeneous CRN - the problem of interference caused to the PUs when the SUs transmit. This problem is believed to be the most debilitating problem with RA in CRN. In this chapter, by the use of cooperative diversity, the interference problem is effectively mitigated. Also, in the chapter, a heuristic for solving the RA problem is developed. The heuristic is important in that it helps in obtaining timeous but near-optimal solutions for the RA problems in heterogeneous CRN.

In Chapter five, the concept of queueing theory is employed in solving yet another problem encountered in RA for heterogeneous CRN - the problem of delay and buffering that usually results in data queues. When users in the network have data to transmit and there are not enough resources to transmit them in the interim, they usually have to keep the excess data in a buffer and wait for available resources. As it almost happens all the time, the various categories of heterogeneous users in CRN have differing delay tolerances. If the delay time is not adequately managed, the RA solutions proffered may have become obsolete by the time the outcomes are being effected. Therefore, in the chapter, an appropriate model is developed and solution methods are provided for solving RA problems in buffered heterogeneous CRN, thereby mitigating the delay limitation while achieving optimal solutions for RA in CRN.

The conclusions are presented in Chapter six. Recommendations on several aspects of the research for possible further (future) considerations are also highlighted in that concluding chapter.

CHAPTER 2 A SURVEY OF RESOURCE ALLOCATION IN COGNITIVE RADIO NETWORKS

2.1 CHAPTER OBJECTIVES

In its quest to become the preferred xG wireless communication paradigm, CRN will depend heavily on its ability to efficiently manage the limited resources at its disposal in meeting the numerous demands of its users and driving its operations. As a result, a considerable amount of research work has recently been dedicated to investigating and developing RA models that capture the essentials of CRN. The various ideas put forward by researchers to address RA problems in CRN have been somewhat diverse and somehow there seems to be no links that bring cohesion and clarity of purpose and/or ideas. To address this problem and bridge the gap, in this chapter, a comprehensive study of the prevalent techniques developed for addressing RA problems in CRN is carried out, with an intent to put some structure, relevance and meaning to the various solution approaches. The solution models are therefore grouped and/or classified based on certain outstanding criteria, and their strengths and weaknesses highlighted. Open-ended problems, which could also potentially limit the RA productivity in CRN, are identified and suggestions for improving solution models are given. The study therefore gives good directions for further investigations on developing RA solutions in CRN.

2.2 AN OVERVIEW OF COGNITIVE RADIO NETWORKS

CRN is no longer an entirely new concept in the wireless communication space. Actually, since it started gaining attention about a decade or so ago, a plethora of technical reports in the form of books, chapters in books, scholarly articles etc. have already been published on CRN. Consequently, a meaningless repetition of details about its ideas and ideals in this chapter would simply be unnecessary. However, while it is sufficiently safe to assume that the majority of the readers of this thesis are well-grounded in the fundamentals of CRN, it is equally understood that there may be a few who are not adequately familiar with the concept. Such readers and enthusiasts who are hitherto not amply informed but would like to have either an in-depth study, or rather, a quick but detailed survey on CRN, are hereby referred to the following references for the needed help: [13, 16–20]. Nonetheless, for necessary completeness and to provide a sufficient preliminary platform for the further discourses in the thesis, a very brief overview of CRN is provided.

The current surge in interest and drive for CRN can be traced to a number of recent developments in the field of wireless communication. First is the fact that the demand for wireless, mobile, always-on communication and/or connectivity has been growing sporadically in the last few decades, with no indication of a decline, whether in the immediate or near future. This continuous increase in wireless communication demands generally requires an equivalent increase in the use of the radio-frequency spectrum to meet this need. The spectrum is, however, a non-expanding, non-ubiquitous resource. Consequently, the increasing demand for spectrum has resulted in a kind of ‘spectrum scarcity’ problem, making it difficult to accommodate the rising wireless communication expectations. This spectrum scarcity has necessitated a review into the current patterns and principles of allocations and applications of the limited spectrum resource. Interestingly, the investigations revealed that while indeed the spectrum is a limited resource, the current problem is, in fact, not that of an insufficient spectrum but rather of poor/inefficient utilisation of the already allotted spectrum by the networks currently occupying them. An important solution to this spectrum scarcity/underutilisation problem was then suggested in the form of establishing a new allocation strategy called DSA. With DSA, spectrum can now be dynamically allocated, and co-use and/or re-use of a spectrum space by more than one ‘owner’ becomes a possibility. Subsequently, CRN emerged as the most potent driving force for the realisation of this new DSA paradigm.

In essence therefore, CRN, by depending on DSA, will be capable of delivering new and improved ways of managing the spectrum. DSA centres on sharing spectrum between original owners or PUs and opportunistic owners or SUs of the spectrum. In earlier descriptions and applications of DSA in CRN, SUs are designed such that they must be able to detect free spectrum spaces or holes, configure themselves to transmit in those frequencies, detect the return of PUs and immediately cease transmitting in those spectrum frequencies. Then, they must look for other free spectrum spaces, reconfigure themselves and resume transmission and be ready to vacate again should a PU return, all of these happening as seamlessly as possible. In more recent considerations of DSA in CRN, SUs may be enabled to transmit alongside PUs at the same time too, depending on the agreement between them. Usually when that is the case, the SUs transmit at low power over a wide bandwidth (e.g., ultra wide band) to minimise possible interference to the PUs. Further developments have however revealed that CRN are far and above just the ability to better manage or administer spectrum. In her small but rich book on the essentials of CRN, Doyle conceptually described CRN thus: “the CRN must be a self-organising system - it understands the context it finds itself in and can configure itself in response to a given set of requirements in an autonomous fashion. The configuration won’t be on frequency or dynamic spectrum alone but on other features too like power, beam pattern, routing algorithm, coding techniques, filtering techniques etc. From the user point of view, the CRN will offer the benefit of personalising users’ experiences so as to provide services tailored to the specific needs of individual users” [13]. It is therefore safe to say that, if these ideals of CRN, as predicted and promised, are eventually realised, the usefulness of CRN can be far and wide, and applications may cover a wide domain including areas such as the military, public safety, academia, health, commerce etc. The enormous promises of CRN therefore make it a technology in which several aspects of human communication life may eventually rely upon, and thus, it is an important field to study and develop.

2.3 ARCHITECTURE OF COGNITIVE RADIO NETWORKS

The basic components that form CRN are the PUs, the base station (BS), which coordinates the activities of the PUs, the SUs, and in most cases an access point or a secondary user base station (SUBS), which coordinates the activities of the SUs [21]. These basic components of CRN combine to form the different kinds of network architectures in CRN. The most common architectural categorisation described in the literature classifies CRN as either centralised (or infrastructure-based), distributed (or ad-hoc based) or mesh architectures [22–25]. The centralised architecture operates in a manner that the access point or SUBS of the SUs controls and coordinates the transmission activities of the SUs. In the distributed architecture, there is no such infrastructural support, rather, the SUs communicate directly with each other in an ad-hoc manner and information is shared between the SUs that fall within the communication range, usually without a central controller. The mesh infrastructure kind of fuses the two architectures together to obtain the best possible performance.The diagrams in Fig. 2.1 and Fig. 2.2 are respective pictorial representations of the centralised and distributed architectures for CRN.

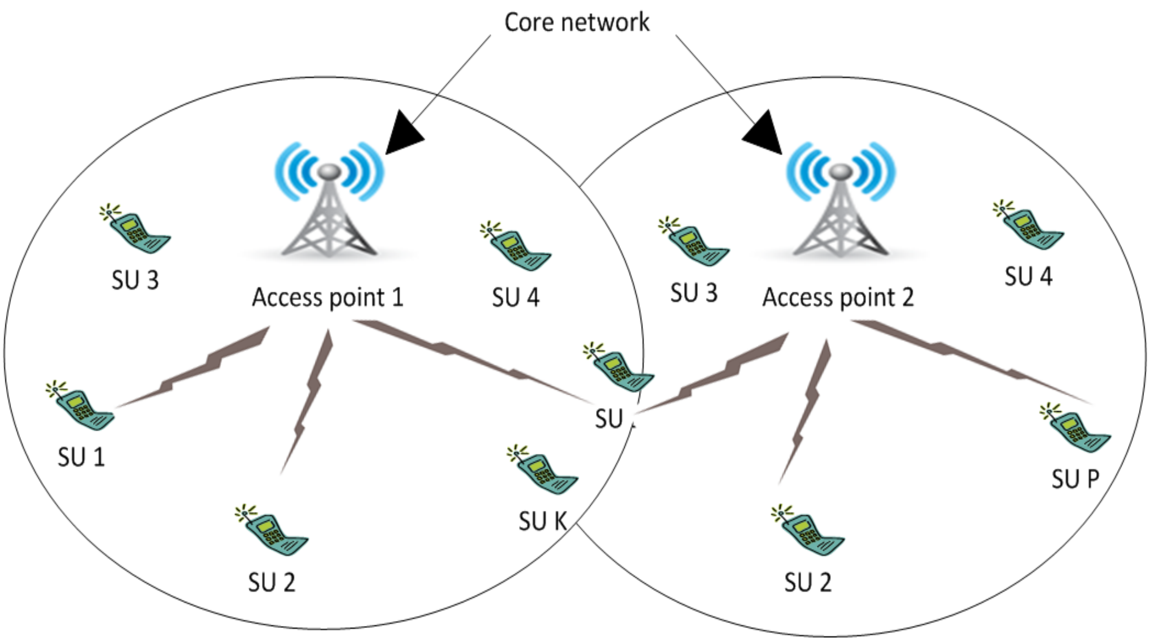

Figure 2.1. Description of the centralised architecture for CRN

Another important description of the primary-secondary network architecture in CRN is based on the interference agreements between the two networks. This network architecture is described by the terms underlay, overlay and hybrid networks. In the underlay architecture, PUs still take priority in the usage of spectrum but SUs are permitted to use the entire frequency space, just as long as the interference they cause to the PUs as a result of their transmission is within a specific tolerance limit [26]. This architecture has the advantage of the possibility of large bandwidth and service provision all of the time, but the limitation on the interference to PUs is always a crippling constraint, especially when the permissible interference temperature limit is very low. In the overlay architecture, SUs have right of spectrum usage only when PUs are not transmitting on that frequency [27]. This implies that the SUs can transmit at maximum power and with high transmission rates during those periods. The SUs however have to vacate the spectrum space immediately the PUs resume their transmission, giving rise

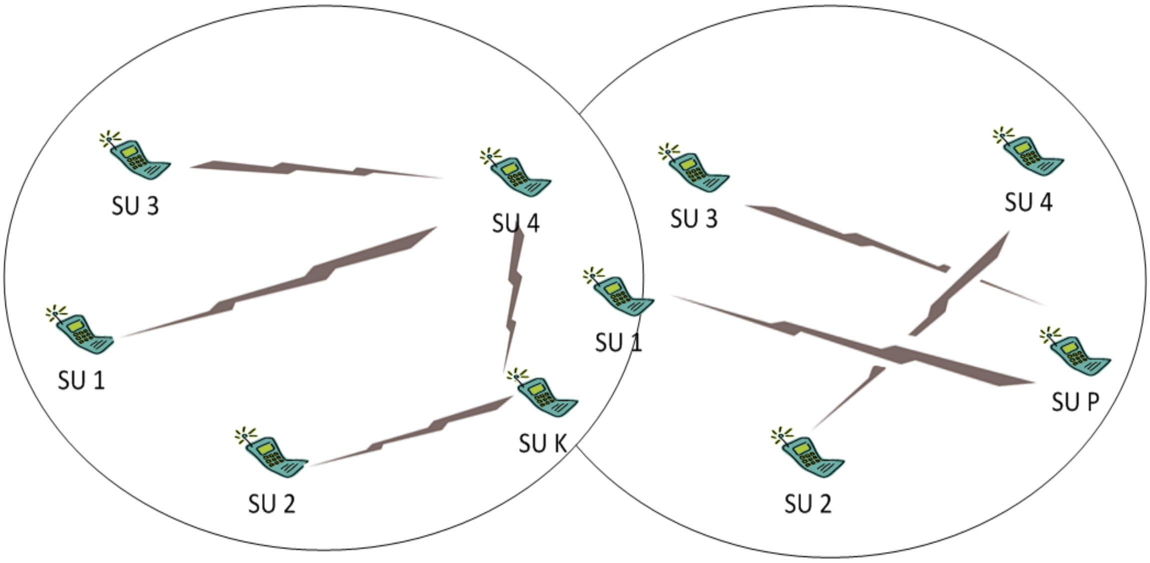

Figure 2.2. Description of the distributed architecture for CRN

to a possibility of service disruptions. There are also timing issues, as well as the problems of missed detections and false alarms frequently associated with the overlay network [28]. With missed detection, the SU network has wrongly judged that the PU is not available to use the spectrum and has instructed the SU to transmit, causing unacceptable interference to the PU. With false alarm, the SU network has again misjudged that the PU is present and informed the SU not to transmit, whereas the PU is actually not present and the spectrum is vacant for use at that moment. The possibility of making these wrong judgements has to be factored into the design for overlay. The hybrid architecture seeks to combine the advantages of the underlay and overlay architectures to provide even better results [29]. In the hybrid design, SUs transmit at maximum rate with full power when the PUs are absent but revert to low transmission immediately the PUs return. However, the network complexity of the hybrid system is much higher than in the other two.

A third architecture commonly associated with CRN is cooperative and non-cooperative architectures. For the cooperative design, the SUs work together to make decisions on such things as spectrum sensing, so that their decisions are usually multilateral and centrally controlled [30, 31]. The non-cooperative architecture works in exactly the opposite way, with each SU making unilateral decisions about its sensing, data transmission etc. Furthermore, cooperation can be between the PU network and the SU network. Certain cooperative architectures describe the PUs and SUs systems as working together in such a way that the SUs transmit some of the PUs' data in exchange for spectrum or some other benefits [32, 33]. The PUs are thus willing to part with a fraction of their frequency band for the SUs to transmit with, as long as they will aid the PUs in completing their own transmission or increasing their capacity. There are a few other cooperative descriptions that have been postulated and demonstrated, such as cooperative beamforming [34], thus making this an equally important architecture for CRN as well.

The various architectures described above form the major classes into which CRN have been mostly divided, and for which studies have been undertaken. Some of these architectures are employed in describing and analysing the models later developed in this thesis. Of major interest in this chapter, however, is the important aspect of resource availability for CRN, and how the available resources can be fairly shared and expeditiously used for CRN to achieve its objective of becoming the preferred model for xG wireless communications.

2.4 RESOURCE ALLOCATION IN COGNITIVE RADIO NETWORKS

Resources used up in wireless communication systems such as power, bandwidth and spectrum have always formed the backbone on which the operations of such systems depend. These resources being generally non-ubiquitous, the various wireless communication models, as developed, have had to factor into their design the mechanisms by which their scarce resources are to be allocated or administered in order to achieve the utmost in their operations. The concept of RA, which seeks to address that need, has therefore been an important aspect of all wireless communication designs. In fact, in several conventional wireless communication systems such as the OFDMA-based wireless networks, RA has been a rather active research topic. For example, a few of the works that have addressed RA problems in OFDMA communication systems can be found in references [35–41]. In general, RA problems in wireless communication essentially define how to optimise the limited resources in the communication network. RA problems are not therefore new and/or characteristic to CRN. Particularly for CRN, RA seeks to jointly address the challenge of allocating its scarce resources, viz. spectrum (frequency band, subchannels and time slots), power, bit, bandwidth, modulation schemes, data rates and other such resources in a manner that is fair to all users (primary and secondary) of the network. Studying RA in CRN is crucial, essentially more crucial that in most other wireless communication paradigms because both the primary and the secondary networks have to rely on the limited resources availability in order to carry out their communication. This exacerbates the challenge of resource scarcity in CRN than in other conventional wireless communications. In other words, the fact that both the original and the opportunistic networks in CRN have to make us of the limited available resources makes resources in CRN to be characteristically scarce. The diagrams in Fig. 2.3 and Fig. 2.4 provide general descriptions of the allocation and usage patterns of resources in underlay and overlay CRN respectively.

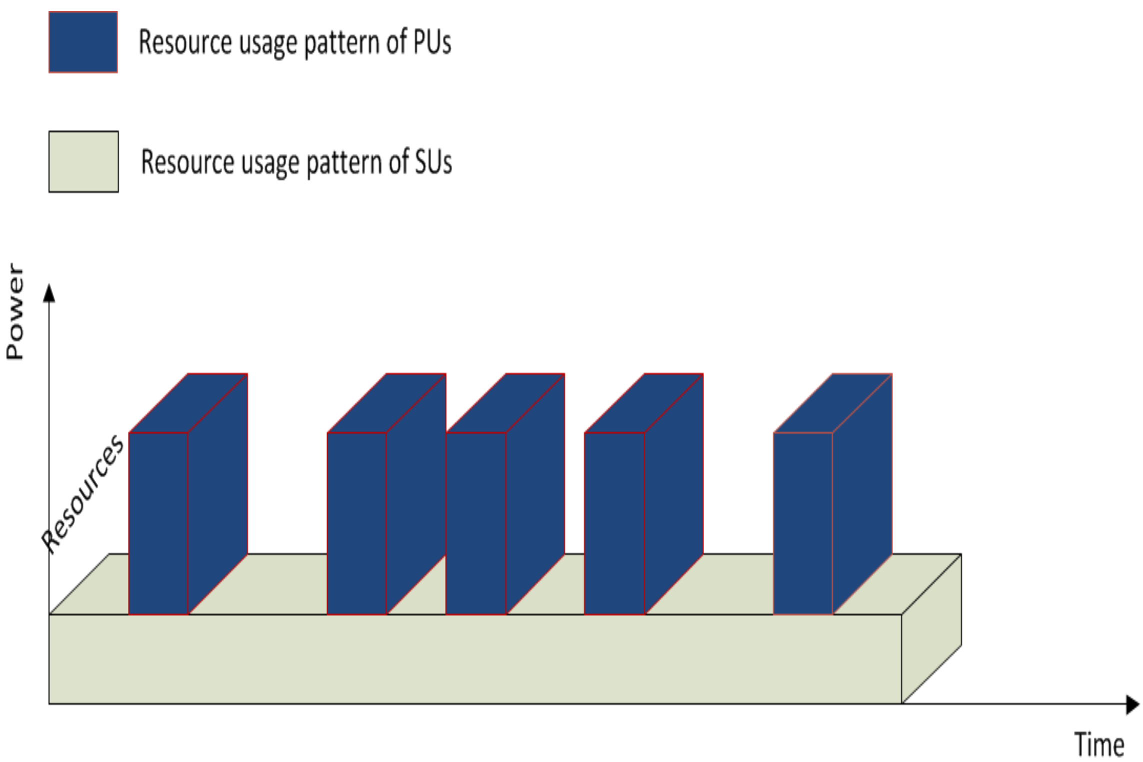

Figure 2.3. Resource allocation and usage pattern for underlay CRN

There are generally two well-developed approaches that have been actively adopted for addressing or describing RA problems in most wireless communication systems like the OFDMA-based networks. The approaches are referred to as rate adaptive resource allocation (RARA) and margin adaptive resource allocation (MARA) models [38]. In RARA, the goal is usually to maximise a given function of the transmission rates, total capacity, fairness etc. of users under a total power constraint at the BS.

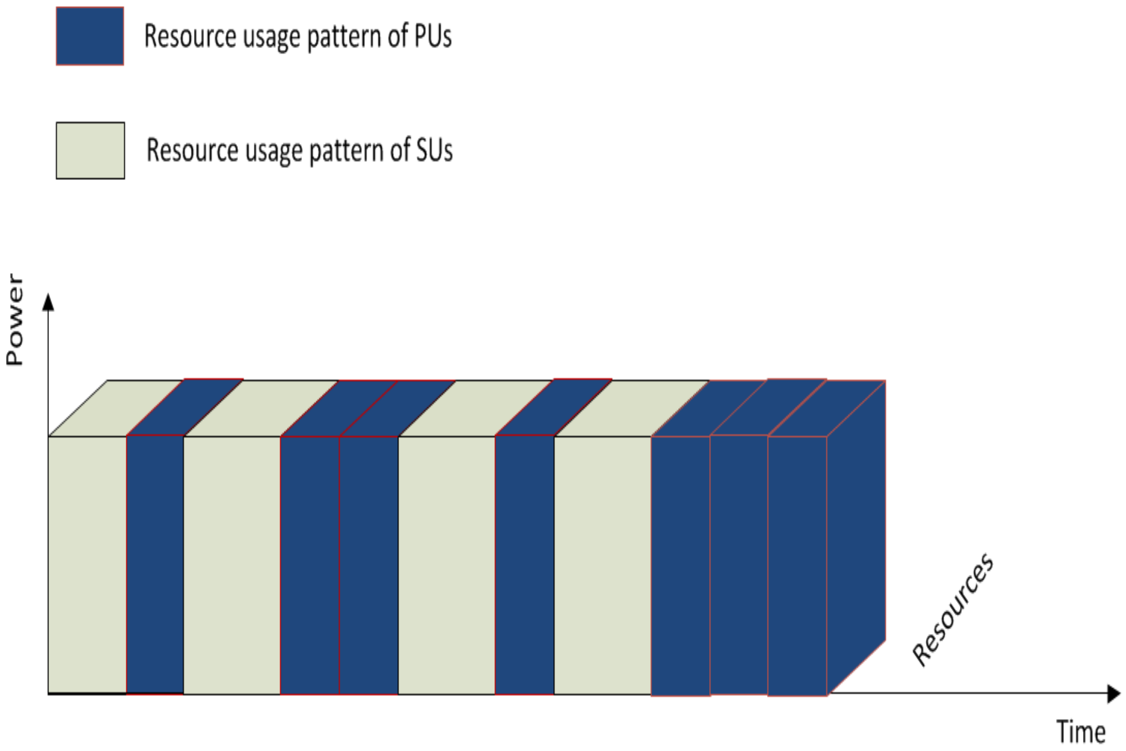

Figure 2.4. Resource allocation and usage pattern for Overlay CRN

Examples of RA problems developed and investigated as RARA problems can be found in references [42–44]. Models that adopt the MARA method seek to minimise the total transmission power used up by the network while ensuring that the required transmission rates for all users are met. References [45–47] give examples of RA problems developed and investigated as MARA problems.

Recent investigations have suggested that the methods developed for addressing RA problems in wireless communications (particularly the OFDM/OFDMA and its variants) are actually very adaptable to the RA problems in CRN as well [48]. It means therefore that, in general, RA problems in CRN can similarly be broadly classified as either RARA or MARA. However, it is important to note that RA problems in CRN do pose a much higher level of challenge or difficulty than in other conventional wireless networks for several reasons. One important reason is the possible fluctuations in the available spectrum and hence, the frequency and bandwidth of operation in CRN [49]. Another critical reason is the difficulty associated with, and the limiting effects of considering CRN as a heterogeneous network. The heterogeneity of CRN would imply that in the design of CRN, the wireless network communication infrastructure must be capable of servicing a heterogeneous, probably incompatible set of wireless consumer devices [50]. One other crucial factor that makes RA problems in CRN very challenging is the limitation in networking and productivity of CRN due to the level of permissible interference to either the PUs, or even among the SUs themselves. The limitation in SUs' transmission due to the level of permissible interference to PUs is probably the most crippling of constraints in achieving great resourcefulness and optimal utility in CRN. The above reasons make it imperative to carry out detailed investigations on the basis and principles for adopting/adapting the developed methods for RA in other wireless communications to CRN. Such investigations will not only ascertain their suitability of application or purpose, but will also help to describe and analyse their workability. A considerable amount of work has already been carried out in this regard, as this literature study reveals, but much more work is still required in order to bridge the research gaps.

2.5 RESOURCE ALLOCATION PROBLEM FORMULATION IN COGNITIVE RADIO NETWORKS

There is already a sizeable amount of research work on solving RA problems in CRN. The various investigations have shown that, in almost all cases, RA problems in CRN are fully demonstrated to be optimisation problems. The knowledge of optimisation is therefore crucial to the understanding of, and in developing solution models to RA problems in CRN. In essence, optimisation can be explored and employed as a vital tool for solving RA problems in CRN. Optimisation, in itself, is a well-developed analytical tool for solving a host of scientific-related problems and is therefore used broadly in different fields of science such as mathematics, operations research, business and financial management, economics, engineering etc. In optimisation, there is usually an objective (there could be more than one objective too) to be achieved, either that of maximising or minimising an entity or a number of entities, and this is always captured in the objective function. Then, there are certain limiting constraints that must be taken into consideration while seeking to achieve the objective. In solving, the constraints cannot be violated, otherwise the solutions to such problems, if ever obtained, become void. The final components of all optimisation problems are the decision variables. These variables are the parameters to be obtained while solving, in order to arrive at (optimal or suboptimal) solutions. Due to space limitations and also to help keep focus, the preliminaries on optimisation are not discussed in this thesis. The following materials are recommended in providing some fundamental knowledge on optimisation, should a reader require such: references [51–54].

Moving forward in the thesis, a general form of RA optimisation problem formulation in CRN is next provided. The general formulation gives a description of what the objective functions usually are, as well as the constraints and the decision variables, and the interplay between them. Let $\mathbf{p}$ and $\mathbf{q}$ be two vectors of dimensions a and b respectively. Also let the set of positive integers $I = \{0, 1, 2, ...\}$. Assume we need to obtain the values of $\mathbf{p}$ and $\mathbf{q}$ for which a function $f(\mathbf{p}, \mathbf{q})$ is maximum, given that there are a set of constraints $g_i(\mathbf{p}, \mathbf{q}) \leq n_i$, $i = 1, 2, ..., r$, and that each variable is non-negative. The above formulation can be written mathematically as:

$$\max z = f(\mathbf{p}, \mathbf{q}) \tag{2.1}$$

subject to

$$g_i(\mathbf{p}, \mathbf{q}) \leq n_i, i = 1, 2, ..., r, \tag{2.2}$$

$$p_j \geq 0, \ j = 1, 2, ..., a, \tag{2.3}$$

$$q_k \in I, \ k = 1, 2, ..., b. \tag{2.4}$$

Equation (2.2) is more simply written as:

$$\mathbf{g}(\mathbf{p}, \mathbf{q}) \leq \mathbf{n},$$

where

$$\mathbf{g}(\mathbf{p}, \mathbf{q}) = \begin{bmatrix} g_1(\mathbf{p}, \mathbf{q}) \\ g_2(\mathbf{p}, \mathbf{q}) \\ \vdots \\ g_r(\mathbf{p}, \mathbf{q}) \end{bmatrix},$$

and $\mathbf{n} = [n_1, n_2, ..., n_r]^T$. If the problem was a minimisation problem, the function $z = f(\mathbf{p}, \mathbf{q})$ could be easily transformed to a form of maximisation function by simply negating the objective function, i.e., $\max w = -f(\mathbf{p}, \mathbf{q})$. From the general formulation given above, equation (2.1) is the objective function, equations (2.2)-(2.4) are the constraints, while p_j and q_k are the decision variables. As an example, equation (2.1) could be a maximisation of the total network capacity, vector $\mathbf{p}$ could be a set of transmission power for users, vector $\mathbf{q}$ could be subcarrier allocation, which would usually take integer values of 0 or 1, and equation (2.2) could be the interference limit constraint or the power constraint.

Table 2.1 presents some examples of works/related works on RA problems in CRN that have either been addressed or sought to be addressed by various authors. The table highlights the objective function, constraints and decision variables employed by these authors in achieving or seeking to achieve their goal. Although it is by no way exhaustive, the intent is to provide an idea on the different formulations developed by authors in achieving optimal or suboptimal RA for CRN.

Table 2.1. Description of RA problem formulations in CRN.

References	Problem definition	Objective function	Main constraints	Decision variables
[55–58]	Optimal RA in MIMO-based CRN.	Maximising the achievable data rate (or total capacity) of SUs.	Transmit power limit of SUs, interference limit to PUs, total transmission time of SUs must be equal to the time slot duration.	Number of SUs served.
[59–61]	Efficient RA for CRN with cooperation.	Maximising the sum rate of all SUs.	Transmission power budget of the SUs and the relays, interference to PUs within its tolerable threshold, each subchannel can only be allocated to one SU.	Achievable rate over a subchannel, power allocated to each subchannel, integer variables of time slot and a binary allocation indicator
[62–64]	Energy-efficient RA for CRN with imperfect sensing and/or femtocells.	Maximising bandwidth capacity for SUs.	Power constraint on SUs network, minimum rate guarantee for some SUs and best effort service for the remaining SUs, each subchannel can only be allocated to one SU.	Transmit power for each SU, the binary channel allocation indicator.
[65–68]	Optimal RA in MIMO-OFDMA based CRN.	Maximising sum throughput of the SUs.	Interference leakage to PUs always below a threshold, each SU must achieve the minimum required data rate, total transmit power of all SUs must be below the available power at base station, no more than one SU is allocated to each subchannel.	Data rate on each subchannel.
[69]	RA for CRN with opportunistic access	Minimising the symbol error rate of the SUs' network transmission.	Constraint on the maximum individual power of each SU, a minimum number of symbols must be sent within a time frame, constraint on the minimum acceptable throughput of the network, interference power to PUs must be below a certain threshold.	Total power available to the system, transmitted symbol time,

From the general formulation of RA in CRN provided in equations (2.1) - (2.4) (which represents succinctly most formulations on RA for CRN in the literature), it can be observed that the RA problems in CRN are best described as complex, non-deterministic polynomial-time hard (NP-hard) optimisation problems. By way of a simple definition, NP-hard problems are problems which may be solvable in polynomial time, but then, only by a non-deterministic algorithm. Determinism in optimisation means that there is a finite and manageable set of actions to be considered, and that everything that happens is determined by a necessary chain of causation [13]. The 'non-determinism' of NP-hard problems therefore implies that there are usually multiple choices of actions for investigating solutions, and the actual choice made when the algorithm runs is not determined by the input or certain values in its register or even its current state. Instead, such algorithm makes an arbitrary choice among the several possibilities in each run or solution attempt. Thus, it is not impossible to have multiple runs of the same algorithm on the same input resulting in different outputs. The whole point of using non-deterministic algorithms in NP-hard problems is that, such algorithms are enabled to make certain guesses at certain points during their computation which are crucial to their ability to obtain solutions [70, 71]. The algorithms are designed so that, if they make the right guesses at all the choice points, then they can solve the problem at hand. The polynomial time part of an NP-hard problem means that, if the non-deterministic algorithm makes all the right guesses, then, the amount of time it takes to obtain solutions to such problems is usually bounded by a polynomial.

In essence, obtaining solutions to NP-hard optimisation problems into which category RA problems in CRN fall, though possible, can be very difficult. More so, by reason of its non-deterministic nature, there is usually an uncertainty in the time duration for arriving at such solutions. Invariably, obtaining solutions could require vastly more time than it takes to describe such problems. Presumably, solutions to wireless communication problems, especially CRN, have to be timeous for them to be meaningful and useful. If the solutions take too long to be reached, premises and prevailing conditions upon which the original problems were designed may have changed considerably, thus rendering the purported solutions unusable. One of the major issues with CRN, still open-ended, is in developing generalised RA solution models with minimal time requirements and low computational complexities. Finding meaningful and useful (applicable) methods for arriving at solutions to the RA problems in CRN is therefore an exciting research focus.

Having established the important premise that RA problems in CRN are indeed optimisation problems and that due to the class of optimisation into which they fall, obtaining solutions for them is generally difficult, investigating and developing methods by which viable solutions can be obtained, taking into consideration the peculiarities and limitations of CRN, is one major goal of the work presented this thesis. To achieve the goal, an investigation into the general approaches to solving optimisation problems of RA in CRN is first carried out. From the investigation, key limiting elements of CRN that have been either ignored or oversimplified in problem analyses are identified, as well as how such omissions and/or commissions affect the overall solutions provided. Further still, possible strengths and weaknesses of past solution methods/models are exposed. Finally, novel and/or improved models, which are almost-certainly all-inclusive in that they capture the most essential aspects/needs of CRN and address its limitations, are proposed and developed. The benefits of the newly developed models become apparent in that, comparative analyses with pre-existing models show marked improvements in overall performance and productivity of CRN when these new ideals are incorporated.

2.6 CLASSIFICATION OF RESOURCE ALLOCATION SOLUTION APPROACHES FOR COGNITIVE RADIO NETWORKS

There are a number of approaches developed for solving the complex NP-hard RA problems in CRN that have been proposed and promulgated. In this section, the solution approaches are classified and critically examined. Comparisons and contrasts are made, and useful inferences are drawn based on the comparative studies. Also, research gaps are identified which then sets the pace for the further research carried out in subsequent chapters of this thesis. For clarification and ease of reference, the various approaches to solving RA problems in CRN are classified into these broad perspectives:

- obtaining solutions through classical optimisation,
- obtaining solutions by a careful study of problem structure,
- obtaining solutions by the use of heuristics or meta-heuristics (global optimisation),
- obtaining solutions by applying game theory (multi-objective optimisation),
- obtaining solutions through soft computing-based optimisation.

These categories are further discussed.

2.6.1 Solutions using classical optimisation

RA problems in CRN that fall into any of the well-developed classical optimisation methods can be solved optimally using the class of optimisation into which they fall. For instance, if a developed RA problem happens to be a linear programming (LP) problem, several established methods for solving such problems exist. Examples of methods for solving LP problems are simplex and interior point methods. In [72], the authors developed their frequency-time allocation problem in cognitive radio wireless mesh network as an LP problem and then employed the simplex method to obtain optimal solutions. In [73], the problem of optimally allocating PU bands to SUs was addressed and the optimisation problem used to obtain the stability region's envelope was shown to be, and solved as, an LP. Interior point method was used in [74] to address the problem of joint transmit beamforming and power control of SUs when they are allowed to transmit simultaneously with PUs. Furthermore, even when a RA problem is non-linear but if its convexity can be established, there are several known methods for solving convex optimisation problems that can be employed to solve such problems. One example of a method for solving convex optimisation problems is by using the Lagrangian duality method, usually with the application of the Karush-Kuhn-Tucker (KKT) conditions [75].

Classical optimisation approaches employ well-established tools in obtaining optimal solutions to developed RA problems for CRN when such problems nicely fit into well-known optimisation structures. In general, classical optimisation tools used for solving well-defined linear or convex programming RA problems are mostly offshoots of either the simplex or the interior point methods. Some of the most common methods and the corresponding references where they have been employed in obtaining solutions to RA problems in CRN are as follows: branch-and-bound (BnB) [35, 67], branch-and-cut (BnC) [76], lift-and-shift (LnS) [77], iterative and double-loop iterative methods [37, 39], dual decomposition [37, 78], Lagrangian duality [62, 78], barrier method [59, 79], gradient decent approach [80], column generation [81, 82], etc. Again, these methods, because of their advantage of obtaining optimal solutions, are highly significant. Their utmost importance lies in the fact that, solutions provided, because they are optimal, can act as bounds for the suboptimal solutions obtained by the use of other approaches or methods. The major disadvantages of these methods are that; firstly, most RA problems in CRN do not usually fit nicely into any standard optimisation model, and secondly, proving convexity for most non-linear programming problems can be herculean, if not impossible to achieve. Also, obtaining solutions with this approach usually requires high complexities and computational time and resources.

2.6.2 Solutions by studying problem structure

As earlier mentioned, most RA problems in CRN do not usually fit into any standard optimisation model, and as such, directly applying classical optimisation to obtain solutions, in most cases, is highly improbable. However, a number of other techniques have been exploited and employed in seeking solutions. One important technique is by carefully studying the structure of such problems to see if there are any special feature(s) that can be exploited to either make such problems easier to solve, or to fit them into some classical optimisation modelled problems. Usually, this approach will either give optimal or suboptimal solutions, depending on how close the restructured or reconstituted problems are to the original problem formulation. Some known approaches based on the study of problem structure are examined below.

2.6.2.1 Solution by separation or decomposition

Certain RA problems can be split into two (or more) simpler problems without significantly affecting the overall import of such problems. In other words, by a careful study of the problem structure, an original RA problem can be separated or decomposed into two or more simpler sub-problems and each solved individually, usually with a lot less difficulty. The solutions are later combined to give the exact (or close to exact) final response to the initial problem. There are several methods of decomposition that have been used in solving RA problems in CRN. One such decomposition method is the Dantzig-Wolfe decomposition [83]. Examples of RA problems in the CRN that have employed decomposition in arriving at solutions can be found in references [37, 57]. In [57], the authors obtained optimal solution to their RA problem by using a primal-dual decomposition method whereby, the overall problem is decomposed into individual power allocation sub-problems and solved for every decision variable pair. Authors in [37] divided their RA problem (joint spectrum and power allocation for multiband CRN) into two stages and used an iterative dual decomposition method to solve it. In [84], the authors developed a CRN duality technique that decomposed their utility maximisation problem into three sub-problems - optimising signal-to-interference-and-noise ratio (SINR) assignment, optimising power and optimising interference temperature. Similarly, the work in [82] used a decomposition approach to jointly address the problem of spectrum sensing, channel assignment and power allocation in cellular CRN. The initial problem, which was a mixed integer non-linear programming (MINLP) problem, was decomposed into two sub-problems - optimal spectrum sensing and optimal channel assignment and power allocation. This was achieved without sacrificing optimality of the entire network. The advantage of this solution technique is the possibility of realising optimal solutions with reduced computational complexity. The major bottlenecks are that not all problems are decomposable, and some problems loose a significant part of their imports when attempted to be decomposed into smaller sub-problems.

2.6.2.2 Solution by linearisation

In almost all RA problems in CRN, the original problems, as developed, are usually non-linear in nature. Either the objective function is not a linear function or one or more of the constraints is/are not linear. Once the linearity of either the objective function or any/some of the constraints cannot be established, the problem is best considered as a non-linear optimisation problem. A useful method for obtaining solutions to non-linear RA problems in CRN is by seeking to linearise the non-linear expressions/constraints of the problem. If/once this can be achieved, obtaining solutions to the linear optimisation counterpart of the problem through classical optimisation becomes straightforward. The linearised expressions may indeed be approximates of the original, but if the values obtained are close estimates or within certain acceptable limits or bounds, the solutions provided by the new problem can be useful and meaningful, even though suboptimal. Examples of RA problems in which linearisation has been employed as a useful tool for obtaining solutions can be found in [36, 85, 86]. In [86], a combination of linearisation, relaxation and reformulation techniques were employed in solving their RA problem. For the linearisation part, a constraint, which was non-linear due to the combination of multiplication and division operations, was transformed into a linear form by the use of the logarithm function. The problem's equivalency was maintained due to the monotonicity property of the logarithm function. The major advantage of this technique is the ease with which LP problems are solved as compared to non-LP, once the linearisation can be achieved. The major challenge with the technique is that certain functions or expressions which commonly appear, either in the objective function or constraints of RA problems in CRN, are very difficult to find equivalent linear expressions for.

2.6.2.3 Solution by relaxation

Some RA problems in CRN are complex and difficult, mainly because of an integer constraint. Indeed, many problems that deal with channel (or subchannel) allocation are binary in nature, whereby a channel is either allocated to a user (assigned the value 1) or it is not (assigned the value 0). These kinds of problems can be solved a lot more easily by relaxing the integer constraint, i.e., by allowing the decision variable to take any value between 0 and 1, rather than imposing it as either 0 or 1. By rounding up or down, approximate solutions to those problems can be more easily obtained. An example of where relaxation has been used in obtaining solution to a RA problem is in [87]. The RA problem was developed as a MINLP problem but by relaxing the integer constraint, the problem became a LP and was then solved. The work in [86] also employed integer relaxation in developing its solution model. The major issue with the relaxation approach is that only suboptimal solutions can be obtained and, in some instances, the gap between solutions obtained after relaxation and the optimal can be significantly wide apart.

2.6.2.4 Solution by approximation

An important method for obtaining useful solutions to RA optimisation problems in CRN is through approximation. Certain functions, appearing in the objective function or the constraints, could be all that render an almost-linear problem non-linear or a should-be convex problem non-convex thereby making the entire problem difficult to solve. If an approximate substitute to such functions can be obtained, the entire problem could become linear or convex, and obtaining solutions could be a lot easier. The substitute to such functions must of necessity be a close approximation of the initial functions before this method can be meaningfully employed. Again, only a suboptimal solution to the original problem can be achieved, but should the approximate substitute of the functions be good enough, the suboptimal solutions can be very close to the optimal and therefore extremely useful. Importantly, the complexities in computations, problem analyses and time to arrive at solutions can be significantly reduced due to the approximation of such functions. There are examples in the literature of the use of approximation in obtaining solutions to RA problems in the OFDMA-based networks, as well as CRN. In [40] for instance, the authors, in order to maximise total network utility of their heterogeneous OFDMA system, approximated their best-effort user utility function as a piece-wise linear function and proposed an LP-based cluster allocation algorithm for solving. The major setbacks with this approach are that, the approximate representation of the original function could contain a number of extra variables, leading to an increase in the number of the decision variables of the entire problem and solutions obtained through approximate substitutes are usually suboptimal rather than optimal.

2.6.2.5 Solution by reformulation

Another important approach used extensively to obtain solutions to NP-hard RA optimisation problems in CRN is through reformulation. By careful consideration of the structure of a RA problem, certain distinct properties of the problem, once identified, can be exploited in arriving at a reformulation or regeneration of the original problem, and that without losing it imports or details. The new or reformulated problem is, in most cases, an easier version of the original problem, and such that classical optimisation tools may be employed in arriving at viable solutions. This method has been applied in a number of RA problems in CRN, examples are in references [38, 65–68, 84, 88]. In [84], the authors, in an attempt to solve the utility maximisation problem for spectrum sharing in CRN, due to its non-convexity and tight coupling between power and interference had to reformulate. The reformulated problem was an optimisation problem involving spectral radius constraint sets and optimal solutions were obtained by using a tuning-free geometrically fast convergent algorithm. Authors in [88] developed algorithms for decision making to optimise radio resource usage in heterogeneous cognitive wireless networks. An important part of the solution was in the reformulation of the heterogeneous BS selection problem to a minimum cost-flow problem, which was then solved as a directional graph with low computational complexity. The works in [65–68] have all followed a similar pattern of reformulating RA problems which were originally non-linear, non-convex NP-hard, into integer linear programming (ILP) problems, and then solving optimally using the brand-and-bound (BnB) optimisation technique. The main advantage of the reformulation approach is that optimal solutions can be obtained to seemingly difficult problems, and sometimes even with much less computational complexity, once that ‘special structure’ has been found and exploited in achieving the reformulation. The sole drawback of this method is the difficulty in finding the special structure that can be exploited in certain RA problems.

2.6.3 Solutions by heuristics or meta-heuristics

A very popular approach, used on many occasions to obtain solutions to RA problems in CRN, is through the development of problem-specific heuristic(s). Certain problems will be almost certainly impossible to solve through classical optimisation, no matter what ‘trick’ is sought for or employed to try to make the problem solvable. Even in situations where any of the already-discussed methods (such as linearisation or approximation) succeed and the RA problems have become solvable, in most cases, it is still unlikely that such optimal or suboptimal solutions provided would be obtainable in a reasonably feasible time frame for practical purposes or real-life scenarios. The complexities of the problem would, in all probability, make the solutions impracticable, most especially for large networks. It therefore means that methods for obtaining much faster solutions with less computational complexities must of necessity be devised. In most cases therefore, a heuristic is always developed alongside the solution provided through any of the aforementioned methods, so as to achieve this goal of obtaining a solution in a reasonable time frame that is good enough for solving the problem at hand.

In developing heuristics, logical reasoning and not necessarily analytical or numerical derivations on how to solve a particular problem, is pursued. The solutions to the RA problems in CRN using heuristics are thus problem-specific, almost certainly non-transferable to solve other RA problems, and they usually only provide suboptimal solutions. The advantages with heuristics are that problems that may not be solvable by classical optimisation may be solved by developing a heuristic for them, and that such solutions are usually obtained at a much more reduced time frame, even with large networks. Examples of heuristic methods that have been developed and employed in solving RA problems in CRN are given below:

- Greedy algorithms: In greedy algorithms, the heuristic is developed in such a way that it selects whatever is currently or immediately the best next step, regardless of whether or not there could be some better steps later. Variants of the greedy algorithm are selective greedy and distributed greedy algorithms. References [27, 35, 36, 89, 90] have all employed greedy algorithms in obtaining solutions to their RA problems in CRN. Solutions provided using this technique are not usually optimal but they can be obtained in a reasonably good time frame.

- Water-filling schemes: Several water-filling heuristics (and their variants) have been developed to solve RA problems in CRN. The water-filling schemes developed from the idea of the popular water jug problem. Examples where these schemes have been employed in solving RA problems in CRN can be found in references [55, 56, 69, 91–93]. The methods are simple to develop and they give very close-to-optimal solutions with reduced complexities.

- Preassignment-and-reassignment algorithms: In preassignment, a certain amount of resources, subchannel or power for instance, are initially pre-allocated as base resources to some or all users before the other resources are optimally shared among the remaining users. As the algorithm runs, more resources are allocated to all or a category of users to achieve a higher overall capacity or productivity. After one or more runs, the algorithm may check that the constraints are not violated and should there still be some residual resources, a reallocation (or reassignment) of resources is again carried out to seek to improve the overall utility of the network. Examples of the use of this heuristic method can be found in references [35, 94].

- Recursive-based and/or iterative-based heuristics: These methods carry out allocation of resources either recursively or iteratively to all users in the network. While iteration uses a repetition structure, recursion uses a selection structure. Importantly, both methods steadily increase utility until further iteration or recursion results in a negligible amount of improvement and thus, a termination is evoked. References [65, 79] have applied these techniques in developing their heuristics to solve their RA problems.

Meta-heuristics are developed for solving computationally demanding RA problems. They are generally wide-ranged, and are employed more for problems that have the possibility of obtaining a number of local 'optimal' solutions, or such problems for which there is no satisfactory problem-specific algorithm to solve them. A meta-heuristic is thus an algorithm designed to solve approximately a wide range of hard optimisation problems, without having to adapt deeply to each problem [95]. Meta-heuristics involve using tricks so that the algorithm does not get stuck around a local minima or maxima, whereas a better optimal solution could still have been realised. Some examples of meta-heuristics that have been used for RA in CRN are given below:

- Genetic algorithms: Genetic algorithms are used by defining resources in the form of chromosomes and genes and the users' QoS requirements are given as input to the algorithm procedure. An example is found in reference [96] where genetic algorithm was used in optimising spectrum allocation in CRN. Genetic algorithm was also used in reference [97] for optimising spectrum utilisation while providing a fairness guarantee between users in CRN.

- Simulated annealing: In this technique, by the process of iterative controlled 'heating' and 'cooling' of the search space, an optimal 'temperature' is found which corresponds to the optimal utility for the system. References [98, 99] have used this technique in solving allocation or utility maximisation problems in CRN.

- Evolutionary algorithms: These are algorithms that have some inclination towards simulating the evolution of individual structures via processes of selection, recombination and mutation reproduction, thereby producing better solutions. Examples are coco search, ant colony, particle swarm optimisation, bee colony etc. In reference [100], authors used the particle swarm optimisation in realising power allocation for users in CRN. Authors in reference [101] used the bee colony idea in achieving relay assignment with power control for users in CRN.

- Tabu searches: These algorithms explicitly use the history of their searches, both to escape from local minima and to implement an explorative strategy. The main characteristic of this approach is indeed based on the use of mechanisms inspired by the human memory. An example of its use in solving RA problems in CRN is found in reference [102] where the method was applied to achieve an optimised channel allocation for all users of the network.

From the explanations given above, it can be seen that heuristics and meta-heuristics are indeed powerful tools for obtaining solutions, especially for large, practical networks. The major limitations with these methods are the deficiencies in analytical/numerical representations of the problems, and the non-transferability of the knowledge acquired in solving a problem to help solve other problems.

2.6.4 Solutions by multi-objective optimisation

An important approach to solving RA problems in CRN, especially problems that are multi-objective in nature, is the use of game theory. Actually, some developed RA problems in CRN are multi-objective optimisation problems in that they require a process of simultaneously optimising two or more conflicting objectives, subject to certain constraints. One method that has been employed in addressing multi-objective optimisation problems is converting them to single-objective optimisation problems by using techniques such as reducing dimension, Min-Max method, the ideal point method, the weighted sum of squares method, the virtual target method, sequencing method, feasible direction method, the centre method, interactive programming method and a few others [103]. However, good as these techniques are, there are instances where conventional optimisation models may not be adequate in addressing such multi-objective problems, hence, the use of other multi-objective solution techniques such as game theory. Several game models exist, and some of them have been employed in solving multi-objective RA problems in CRN. Some examples, and the corresponding references where they have been applied are: cooperative game [55, 56], non-cooperative game [104], Nash bargaining (Pareto optimisation) [32, 33] and Stackelberg game [63, 64].

2.6.5 Solutions through soft computing

A very new/recent approach to solving RA optimisation problems in CRN is through soft computing-based optimisation. In this approach, software/computer-based programming is used in allocating resources to users within the network. The developed programmes use intelligent techniques such as artificial intelligence, neural networks, Q-learning, fuzzy systems, etc. in driving the optimisation processes [105]. In reference [106] for instance, the authors used a special type of Q-learning, called multi-agent reinforcement learning, in achieving RA for multi-user CRN. During the learning process, each SU sees the channel and other secondary users as its environment, updates its Q-values, and takes the best action based on the prevalent situation. Authors in reference [107] used an artificial intelligence technique in developing a decision-making tool for allocating resource in CRN. In the developed model, cognitive radio learning inference and decision-making engine based on Bayesian network was proposed to obtain the optimum configuration rules to adapt to the variation of the environment with the learning and inference algorithm of Bayesian network. In [108], the authors proposed a fuzzy neural system for spectrum allocation in CRN. In the model, parameters such as spectrum utilisation efficiency, degree of mobility and distance to the PUs of CRN are given as inputs to the fuzzy logic decision making process, while the output of that process gives the spectrum access decision, based on linguistic knowledge of some predetermined rules. The major challenge with this approach is that the soft computing techniques, such as artificial intelligence and neural networks, are very difficult and complex to develop, analyse and apply in real life scenarios.

In summary, there is an ample number of methods that have been developed for solving RA problems in CRN and these methods are usually exploited by researchers in obtaining solutions to their formulated RA problems. In this chapter, the most critical ones have been grouped and their workability explained. Both the strong points of these methods, as well as their weak areas, have been highlighted and discussed. Table 2.2 contains the summary of the solution models presented in this chapter.

2.7 OBSERVATION AND OPEN-ENDED PROBLEMS IN RESOURCE ALLOCATION FOR COGNITIVE RADIO NETWORKS

The author's important observations and/or opinions on these solution approaches/schemes are given below:

Table 2.2. Summary of solution approaches to RA problems in CRN.

S/N	Solution approaches	Solution methods and/or models	Features	Drawbacks
1.	Classical optimisation e.g. LP, convex optimisation etc.	Simplex and its variants (BnB, BnC, LnS, implicit enumeration etc.); interior point method and its variants (barrier method, Newton's method etc.); Lagrangian duality; knapsack; travelling salesman problem etc.	Approach gives optimal solutions; solutions act as bounds (upper or lower) to other solution models.	Usually, most RA problems do not fit into any class of classical optimisation; proving convexity can be very challenging; obtaining solutions can be rather computationally complex and time consuming.
2.	Studying problem structure	Decomposition; linearisation; relaxation; approximation; reformulation.	Solutions can be optimal or very close to optimal; computational complexity is significantly lowered.	Special features might be unavailable or difficult to find; transformed problem may be a far cry from the original; new problem may generate more decision variables than in the original one; solutions are mostly suboptimal.
3.	Heuristics	Greedy algorithms; water-filling algorithms; pre-assignment and reassignment algorithms; iterative-based and recursive-based algorithms.	Solutions are quick to find; less computational complexity; requires little or no numerical analysis; solutions are usually suboptimal but could be close to optimal; approach is suitable for large and practical networks.	Solutions are problem-specific and most times are not transferable; solutions cannot be numerically analysed; solutions are always suboptimal.
4.	Meta-heuristics	Genetic algorithms; simulated annealing; evolutionary algorithms; tabu searches.	Algorithms are mostly nature-inspired; they make use of stochastic components (e.g. random variables); they are good with large, practical and/or computationally demanding problems that have large search spaces; they use 'tricks' so as not to get stuck at a local optimal but to try obtain a global optimal solution.	Solutions are not transferable; solutions cannot be analysed numerically.
5.	Multi-objective optimisation (using game theory)	Cooperative game; non-cooperative game; Nash bargaining (Pareto optimisation); Stackelberg game.	They are good with problems that have multiple objectives; they employ ideas from game theory to solve optimisation problems; they are useful for large, practical networks with large search spaces.	Solution models can be complex; they are not transferable; there may be difficulty in achieving analytical modelling of solutions.
6.	Soft computing-based optimisation	Artificial intelligence; neural networks; Q-learning; fuzzy systems etc.	Software/computer-based programming is used in allocating resources to users within the network; the developed programmes use intelligent and very powerful/sophisticated techniques.	They are very difficult and complex to develop, analyse and apply in real life scenarios.

- Generally, there seems to be a kind of disjointedness in RA problem development, as well as in solution models developed by the various authors. The objective functions for even seemingly similar problems are usually diverse, and so are the constraints and decision variables employed. It therefore seems difficult to find any form of coordination or focal point in the problem definitions. Similarly, the ideas put forth for investigating solutions lack any proper order or a particular standard.

- Sequel to the point raised above, there is therefore no general or one-fits-all solution model or approach for all RA problems in CRN that has been established.

- It is observed that most RA models have neglected some important considerations and/or limiting factors of CRN that should have made the problem more realistic and close to practicality. For instance, the issue of heterogeneity in CRN, which would have created more practical scenarios, has been largely ignored by most authors in their RA problem development and solution investigations.

The reasons that can be construed for the issues raised above are the following realities still currently associated with CRN:

- There is a general difficulty in establishing, explaining and capturing all the details of CRN in one single model. As a result of the numerous and divergent architectures that have been postulated for CRN (as earlier explained in section 2.3), it would be very tedious, almost impossible, to develop RA models that could capture all of the important details in one spell or shot. Several small models that address specific areas of interest, while making reasonable and practical assumptions on other details, are thus (or seem to be) the only currently meaningful approach to developing useful research models on the subject matter.

- There are no well-established standards in place yet for CRN as it is still generally a work-in-progress. Though there have been attempts at defining and describing some form of standards (such as the IEEE 802.22 working group, which was set up to develop a standard for wireless regional area networks (WRAN) that would make use of, on a non-interfering basis, TV white spaces [109]), the fact remains that no standard has been fully established and accepted by all stakeholders for CRN to operate by.

- Optimisation, the main tool used in solving RA problems in CRN is, in itself, a diverse and dynamic problem-solving tool with multiple dimensions of interpretation and application for obtaining solutions to problems. Hence, arriving at a single, generalised solution model for solving RA problems in CRN using optimisation is not very likely.

From the exposures and explanations thus far presented on RA in CRN, some open-ended problems that could potentially limit the productivity of CRN in its RA models, and therefore still require further investigations, have been identified. In this section, the most important ones are mentioned and discussed briefly. Investigations on practicable solutions to these problems then form the basis for the work done and presented in subsequent chapters of the thesis. The open-ended problems, which could limit RA optimisation in CRN, and suggested ideas for solutions, are discussed:

- Network heterogeneity: In all probability, CRN would almost certainly be a type of heterogeneous network (HetNet) or, at the least, it would bear a certain semblance or cut across HetNet in some way. Therefore, proper classification and study of HetNet, and how it applies to CRN, would give the needed ideas on addressing the heterogeneity problem in CRN. Inclusion of heterogeneous classifications into the RA problems of CRN and studying it as such would therefore be a step in the right direction, as this would most likely bring the models closer to practicality.

- Limitations due to the level of permissible interference to PUs: In almost all the works studied on RA in CRN, the most prominent denominator, cutting across all kinds of architecture and RA problem definition/formulation, is the fact that the interference to PUs is a limiting constraint, probably the most limiting. The effects of this limiting constraint seem to be what hamper the progress and possibilities of CRN the most. Unfortunately, almost all of the works reviewed have only mentioned this problem, and of course, the authors have included it as one of the constraints in the optimisation problem, but not much has been done towards mitigating its effect on the overall productivity of CRN. If CRN is ever to achieve its ends, the problem of limitation due to interference to PUs must be adequately addressed. As a suggestion going forward, cooperative diversity, not just being applied for spectrum sensing but for RA, if properly investigated and employed, could be a promising solution to the interference limitations in RA for CRN.

- Data buffering in CRN: The possibility of delay in data transmission has seldom been factored into the RA problems of CRN. Almost all works reviewed have equally neglected this concept in their problem definition. In reality, for heterogeneous CRN particularly, delay tolerance of different users might differ significantly, and there might be need for queue considerations. To analyse RA models that capture such possibilities, the use of queueing theory could help in addressing the delay issues. Hence, RA problems in CRN that factor this into their designs, especially when heterogeneous considerations are also involved, would be a good research focus.

2.8 CONCLUSION

The main objective of this chapter has been to provide a critical review of the various approaches to RA in CRN. The review identified what the important challenges with RA in wireless communication are, and the specific peculiarities of CRN that render such problems even more exacerbating, thus making it very difficult to investigate solutions. Thereafter, the chapter revealed the various ideas, methods and reasoning that have been employed by numerous researchers within the field in seeking viable solutions to the RA problems in CRN. From the review, open-ended problems that could limit RA solutions in CRN were identified and some ideas for possible investigation were then postulated. The ideas put forth in this review chapter form the basis for the direction, considerations and investigations further carried out in the remaining parts of the thesis.

CHAPTER 3 RESOURCE ALLOCATION SOLUTION MODELS FOR HETEROGENEOUS COGNITIVE RADIO NETWORKS

3.1 CHAPTER OVERVIEW

It is evident that CRNs are currently gaining immense recognition as the most-likely xG wireless communication paradigm, because of their enticing promise of mitigating the spectrum scarcity and/or underutilisation challenge. Indisputably, for this promise to ever materialise, CRN must of necessity devise appropriate mechanisms to judiciously allocate their rather scarce or limited resources (spectrum and others) among their numerous users. 'RA in CRN', which essentially describes mechanisms that can effectively and optimally carry out such allocation, so as to achieve the utmost for the network, has therefore recently become an important research focus. However, in most research works on RA in CRN, a highly significant factor that describes a more realistic and practical consideration of CRN has been ignored (or only partially explored), that is, the aspect of the heterogeneity of CRN. To address this important aspect, in this chapter, RA models that incorporates the most essential concepts of heterogeneity, as applicable to CRN, are developed and the imports of such inclusion in the overall networking are investigated. Furthermore, to fully explore the relevance and implications of the various heterogeneous classifications to the RA formulations, weights are attached to the different classes and their effects on the network performance are studied. In solving the developed complex RA problems for heterogeneous CRN, a solution approach that examines and exploits the structure of the problem in achieving a less-complex reformulation, is extensively employed. This approach, as the results presented show, makes is possible to obtain optimal solutions to the rather difficult RA problems of heterogeneous CRN.

3.2 BACKGROUND

It has been earlier established that CRN, with DSA and usage capabilities, can significantly help in mitigating the spectrum scarcity and/or underutilisation challenge [13, 16]. The preliminaries on CRN, as well as a fairly comprehensive overview has been provided in the previous chapter, and references [17, 18, 20] are equally an impressive read. Importantly, the detailed literature study presented in that chapter identified RA as a key enabler for the realisation of the potentials and promises of CRN. As a result, a sizeable amount of work is currently being carried out in this regard [19]. However, there are still a few challenges with RA in CRN that are yet to be extensively addressed, and one such is the necessity of developing and studying RA problems in CRN with the more-realistic consideration of it being a heterogeneous system. In all fairness, introducing heterogeneity into CRN surely portends some intricacies in the RA problem formulations, either with the objectives to be realised or the constraints to be considered. These intricacies associated with such inclusion have made most authors, in their works on RA for CRN, to simply ignore or only partially explore the consideration of heterogeneity. However, because of its significance, it is imperative to study and develop RA models for CRN that incorporate relevant heterogeneous concepts, as well as to investigate the imports of such inclusion in the overall network realisation. This chapter addresses that need. To achieve the goal, in the chapter several associated heterogeneous considerations applicable to CRN are investigated and analysed, while an important approach for obtaining optimal solutions to the developed RA problems is established.

3.3 HETEROGENEITY IN COGNITIVE RADIO NETWORKS

As earlier observed, most works on RA in CRN have been carried out with the assumption that CRN are homogeneous systems. However, the practical and realistic CRN, in almost all certainty, would be heterogeneous in nature. Therefore, in system modelling, describing CRN as heterogeneous is germane to achieving the desired near-accuracy in its interpretations and applications. This is because heterogeneity, when incorporated with CRN, certainly describes more appropriately the most realistic and very practical CRN scenarios. Investigating heterogeneity in CRN, especially in its RA designs, is therefore an imperative for achieving the desired level of network efficiency and productivity. Heterogeneity, as applicable to CRN, can generally be considered from three broad perspectives - heterogeneous networks, heterogeneous users (or user demands) and heterogeneous channels. Detailed explanations on these classifications are next provided.

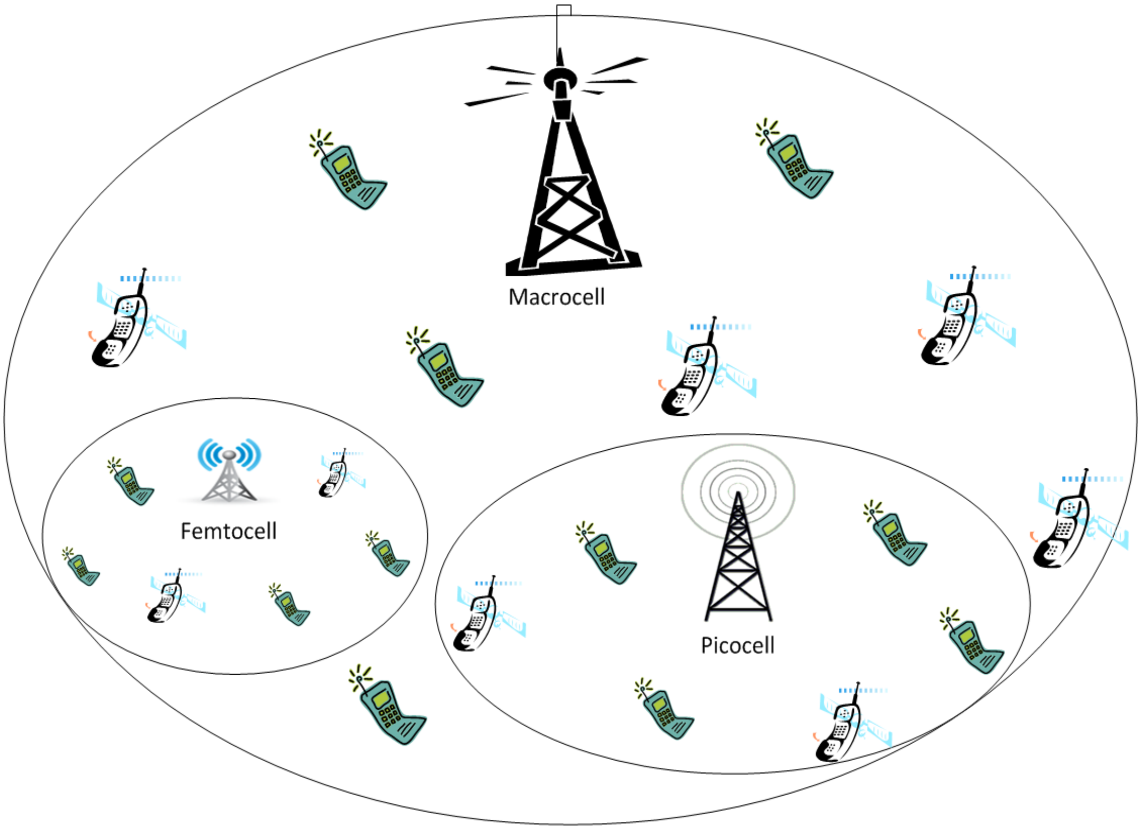

Figure 3.1. An example of heterogeneous network (HetNet)

The concept of **heterogeneous networks**, commonly referred to as HetNet, has gained attention in the research domain in recent times. As a result, several ideas on HetNet are currently being actively investigated. In simple terms, HetNet explain that near-future wireless communication paradigms must be built to work in such a way that they can accommodate simultaneously two or more network configurations, standards, radio access technologies, architectures, transmission solutions, base stations, user demands, etc., in order to expand the mobile network capacity [110]. A very good example of HetNet in recent wireless standards is the femtocells and/or picocells working alongside the more traditional macrocells, as currently being employed and deployed in the long term evolution (LTE)-Advanced technologies. Fig. 3.1 provides a pictorial description of HetNet. The authors in [111] have given a very coordinated analysis on both the concept as well as the major technical challenges associated with HetNet architecture. Importantly, a CRN needs to incorporate the relevant elements of HetNet into its RA problem formulation, so as to achieve a high level of accuracy in its design.

Heterogeneous users or user demands, as applicable to CRN, implies that different users may have different requirements or demands and each user or group of users must be treated based on such considerations [60, 62]. Heterogeneous users can be further classified using the following yardsticks:

- **QoS requirements**: This classification is based on users' minimum rate that will guarantee acceptable QoS. Users that do not have any rate requirements can be treated as best effort service users. An example of the use of this classification can be found in reference [62].

- **Service type or traffic demands**: This classification is based on the type of service being offered by the users e.g. voice call, live-streaming, web surfing, background services like downloading etc. This kind of classification was employed in reference [112].

- **Service availability**: This classification is based on whether the demands are real-time (RT) or non-real-time (NRT). For example, authors in [59, 94] classified their heterogeneous users as either RT or NRT, with RT users being given a higher priority of service provisioning over NRT users.

- **Waiting-time sensitivity**: This classification is based on whether the users are delay-sensitive (DS) or delay-tolerant (DT). An instance of the use of this classification is found in reference [37] where the SUs were classified as either DS or DT, with DS users having a very short waiting time requirement while DT users have a longer waiting time demand.

Heterogeneous channels and/or subchannels, as a class of heterogeneity in CRN, is also very important. Actually, in practical CRN, channels will most likely be located on widely separated slices of frequency bands, and these different channels may have different properties. This implies that a CRN user should be capable of communicating with a heterogeneous set of neighbour users using different channels or channel combinations [113]. This description is further expatiated in reference [114] where, the authors explained that the channels in CRN may not all be identical; different channels would possibly have different propagation characteristics and may support different sets of transmission rates. The very high possibility of having multiple channels for SUs in CRN therefore requires that the devices be capable of using heterogeneous radios. The heterogeneity of channels and radios in CRN introduces a number of issues with their design that must be properly considered and studied. Classifying channels as heterogeneous in CRN and developing and analysing models that incorporate

Table 3.1. Concepts/classification of heterogeneity, as applicable to CRN.

S/N	Heterogeneous categorisation	Basis for classifying users	Basis for classifying users	Basis for classifying users	Basis for classifying users
1.	Heterogeneous networks	Different standards - GSM, EDGE, 3G, LTE, LTE-Advanced etc.	Different cell sizes - macrocells, microcells, femtocells, picocells, etc.	Cooperative networking possibilities - direct communication, cooperative communication, relaying techniques, etc.	Communication technologies - wired, wireless, circuit-switched, packet-switched, etc.
2.	Heterogeneous users and/or user demands (or services)	QoS or rate demands - different minimum rates, different service rates, etc.	Priority - high priority (HP), low priority (LP) users, priority class (PC), best efforts (BE) users, etc.	Sensitivity - sensitive users (XU), general users (GU), etc.	Delay profile - delay sensitive (DS), delay insensitive (DI), delay tolerant (DT) users, etc.
3.	Heterogeneous channels and/or subchannels	Different channel bands - channels and/or subchannels on different slices of frequency bands.	Different channel properties - different channels and/or subchannels may have different properties.	Channel usage designs - a single user should be able to use different channels and/or subchannels simultaneously.	Channel usage examples - OFDM/OFDMA is a classic example of how heterogeneous channels can be applied in CRN.

such inclusion is imperative for the desired near-accurate representation of CRN. In order to cater for the possibility of frequency hopping and mobility, multi-carrier transmission techniques such as the OFDM/OFDMA and their variants have been tipped as the most likely technologies for CRN.

The above classifications of heterogeneity are the most prominent in the field and thus, the most applicable to CRN. Table 3.1 gives a summary of the classifications of heterogeneity applicable to CRN.

3.4 RELATED LITERATURE ON HETEROGENEITY IN COGNITIVE RADIO NETWORKS

There is already in the literature a fairly sizeable number of studies undertaken on RA in CRN. However, only a few of such works have incorporated heterogeneity in their problem formulation or analysis. Even among the few works that have developed their RA problems in CRN with a consideration for some form of heterogeneity, most authors have either obtained suboptimal solutions or simply resorted to developing heuristic(s) to solve their formulated problem(s). The reason for this is because of the extreme difficulty in developing and analysing formulations that can be solved for optimal solutions when heterogeneity in incorporated into CRN. The few works on RA in CRN with some form of heterogeneous considerations, as obtained in the course of this research, are briefly reviewed.

RA in heterogeneous CRN with imperfect spectrum sensing was studied in [60]. In the work, to reduce the complexity of the optimisation problem developed, the authors proposed a subchannel allocation scheme that removes the integer constraint in the channel allocation to the SUs, thus achieving suboptimal solutions. The authors in [62] developed a RA scheme for heterogeneous CRN on the assumption that only the estimates of the channel quality information of the network are available to the SUs. The SUBS carried out its RA to the SUs based on this imperfect channel information. The complexity in computation was reduced by first assuming that subchannel allocations were already known, and on that basis, power was optimally allocated to each SU. An algorithm based on the aggressive discrete stochastic approximation was also proposed to carry out both power and channel allocations for the SUs. Similarly, RA for heterogeneous CRN was studied in [87] while including a guaranteed QoS constraint in the optimisation problem. The complex problem developed was first relaxed and then a low-complexity suboptimal solution method, which separates the RA into two steps - subchannel assignment and power allocation - was employed to obtain solution. In general, the above-mentioned works and probably the few other similar ones in the literature have all identified that the optimisation problems in RA for heterogeneous CRN are extremely complex and difficult to solve. The use of heuristics such as greedy algorithms can help to reduce the complexity and obtain suboptimal solutions [27], and most authors have rather just resorted to using that approach. However, considering that heterogeneity of channels is also a reality in CRN, heuristics that employ suboptimal greedy algorithms may not be well suited for spectrum and channel-sharing networks such as the OFDMA-based CRN because of the multiple constraints on transmit power, interference leakage and individual user data rates [65]. Again, with heuristics, it might be difficult to know how close (or distant) the solutions obtained are to the optimal. More so, obtaining solutions through heuristics alone make it improbable to know what the trade-off between optimality and complexity are. With these points raised, it can be implied that heuristics (alone) might not be the best bet for solving RA problems in heterogeneous CRN. It is therefore imperative to first seek to investigate possible means of obtaining optimal solutions that are both relevant and realistic, even if solutions from heuristics are to be later sought and employed.

In this chapter, the various heterogeneous considerations earlier presented are incorporated into the RA problems and the resulting formulations analysed.In solving the developed complex RA problems for heterogeneous CRN, a solution approach that examines and exploits the structure of a problem in achieving a less-complex reformulation is extensively employed. With this approach, the RA problems, even though NP-hard in their original formulations, are smartly reformulated as integer linear programming (ILP) problems and optimal solutions are obtained for them. A clue is taken from the work in reference [66] and exploited in achieving the reformulation. Thereafter, a special

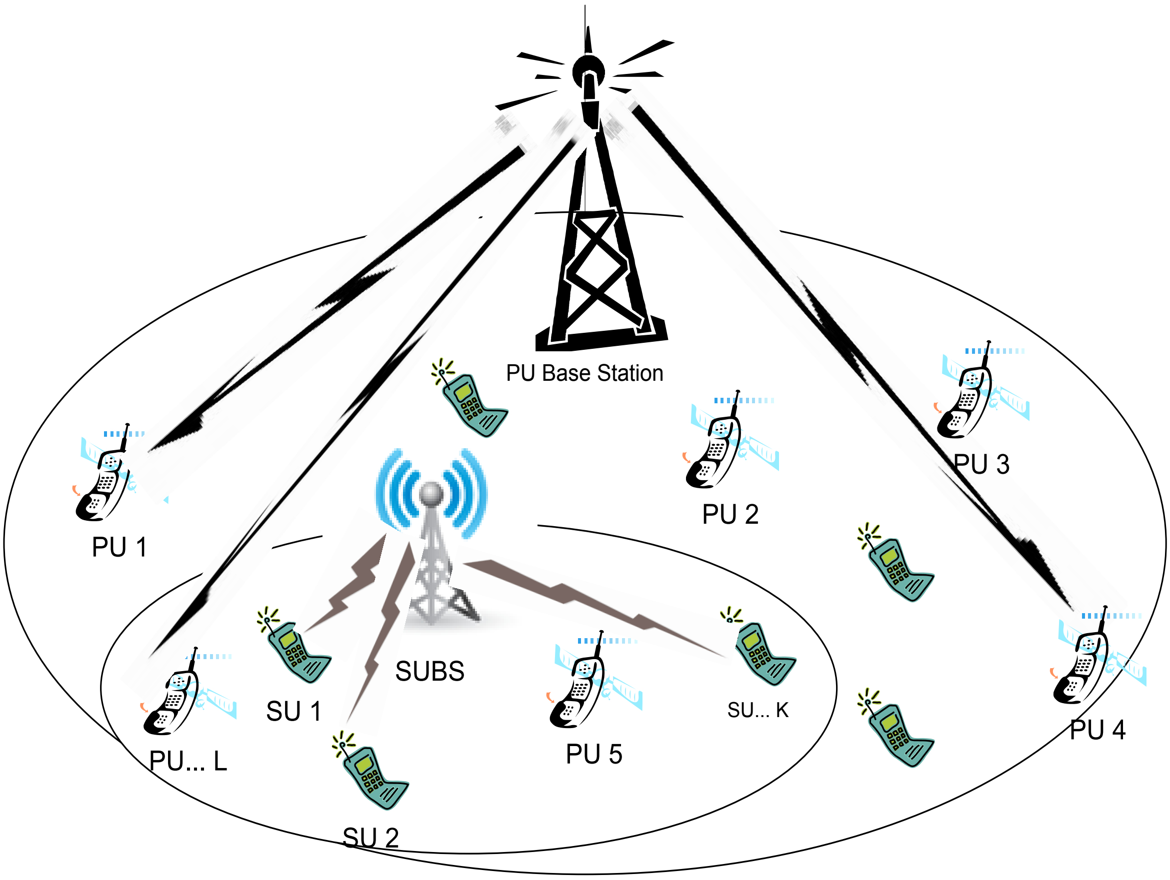

Figure 3.2. System model for heterogeneous CRN

branch-and-bound (BnB) technique, called implicit enumeration [51], is employed to solve the ILP problems. Finally, the chapter investigates the impacts that assigning weights to the various categories of SUs can have on the overall performance of the network.

3.5 SYSTEM MODEL

The system model shown in Fig. 3.2 is applicable to all different kinds of heterogeneous classifications considered in this thesis. In other words, all the different considerations of heterogeneity applicable to CRN, as discussed in the previous section, are incorporated in the model. Network heterogeneity is captured by separating the SUs network from that of the PUs, with each being controlled by its own BS. More so, each network is capable of operating using different configurations of modulation schemes, power levels, interference etc. Channel heterogeneity is taken care of by the use of the OFDMA platform, making it possible to use different slices of the frequency band for different users at the same time. User heterogeneity is incorporated in that the various users are classified and serviced based on some predetermined criteria.

The model is a centralised, underlay, heterogeneous CRN consisting of K heterogeneous SUs and L PUs, all located within the coverage range of the CRN. The SUBS is responsible for instructing the SUs on the resources (subchannels, data rate, transmit power, modulation scheme etc.) that have been allocated or allotted to them. The SUs operate within the interference range of the PUs' network but transmit at such low power that they cause no significant harm to the PUs and thus, their operations are permissible by the PUs. There are N OFDMA subchannels within the coverage region of the SUBS. The SUBS selects the subchannels for each SU and relays this decision to each SU through a separate control channel. The assumption is that the communication between SUs and the SUBS over the control channel is error-free and subchannels are in slow fading. Each subchannel data rate c is dependent on the modulation scheme assigned to the subchannel. Also, each category of SUs has a rate weight $w(w > 0)$ associated with it. The modulation schemes considered are binary phase shift keying (BPSK), 4-quadrature amplitude modulation (QAM), 16-QAM and 64-QAM, which transmit $c = 1,2,4$ and 6 bits per OFDMA symbol respectively. To achieve a given bit error rate (BER) ρ value at the receiver, the minimum amount of power $P(c,\rho)$ required over any given subchannel for the modulation schemes can be determined easily from their power equations [66]. The minimum power for BPSK modulation is obtained from the equation $P(c,\rho) = N_\phi[c \times erfc^{-1}(2\rho)]^2$ (where $c = 1$), while for the M-ary QAM, the minimum power is given as $P(c,\rho) = \frac{2(2^c-1)N_\phi}{3}[erfc^{-1}(\frac{c\rho\sqrt{2^c}}{2(\sqrt{2^c}-1)})]^2$ ($c = 2,4$ or 6 for 4-QAM, 16-QAM and 64-QAM respectively) where, $erfc(x) = (\frac{1}{\sqrt{2\pi}})\int_x^\infty e^{\frac{-t^2}{2}}dt$ is the complementary error function, $\pi = (22/7)$ and N_ϕ is the single-sided noise power spectral density, which is assumed to be the same for all subchannels.

For a given BER ρ value, the amount of power needed to achieve the QoS requirement generally increases (albeit non-linearly) as the number of bits (or modulation scheme) increases. The subchannel power gain matrix between the SUBS and the SUs is given as $\boldsymbol{H}^s \in R^{K\times N}$. The vector $\boldsymbol{H}^s_{k,n}$ therefore denotes the power gain between the SUBS and the kth SU at the nth subchannel. The minimum power $P_{k,n}(c_{k,n},\rho)$ required at the kth SU over the nth subchannel to transmit $c_{k,n}$ bits is obtained by dividing the power $P(c_{k,n},\rho)$ of that user k on the nth subchannel by the channel gain $H^s_{k,n}$ between the SUBS and the user k over that subchannel n. This is given as:

$$P_{k,n}(c_{k,n},\rho) = \frac{P(c_{k,n},\rho)}{H^s_{k,n}}. \tag{3.1}$$

The power gain matrix between the SUBS and the PUs is given by $\boldsymbol{H}^p \in R^{L\times N}$. The vector $\boldsymbol{H}^p_{l,n}$ therefore denotes the subchannel power gain between the SUBS and the lth PU at the nth subchannel.

From the explanations so far presented, both network and channel heterogeneity have been effectively captured in the developed model. To capture the different classes of user heterogeneity (and their effects), the various classifications are developed and analysed one after another, following the categorisation given in the subsequent subsections. But first, for a clear understanding of the RA problem formulations that incorporate heterogeneity, a general representation of the objective function and the constraints for RA problems in heterogeneous CRN is provided.

3.5.1 General representation of the resource allocation formulation for heterogeneous cognitive radio networks

Let the K heterogeneous SUs in a typical CRN be classified into v different categories, based on any given criterion of classification (as already identified in Table 3.1). The different categories of users are thus numbered $1,2,3,\ldots,v$ such that K_1 is the number of SUs in category 1, K_2 is the number of SUs in category 2 and so on. Let a weight w_i be attached to satisfying users in category $i \in v$. This implies therefore that w_1 is the weight attached to category 1 users, and w_v the weight attached to category v users. Given that the objective is to maximise the total data rate for all users in all categories of the network, the objective function can then be written as follows:

$$\max z = \sum_{n=1}^{N}\left(\sum_{k=1}^{K_1} w_1 c_{k,n} + \sum_{k=K_1+1}^{(K_1+K_2)} w_2 c_{k,n} + \sum_{k=K_1+K_2+1}^{(K_1+K_2+K_3)} w_3 c_{k,n} + \ldots + \sum_{k=K_1+\ldots+K_{v-1}+1}^{(K_1+K_2+\ldots+K_v)} w_v c_{k,n}\right); \tag{3.2}$$
$$c_{k,n} \in \{0,1,2,4,6\}$$

Assume that these K heterogeneous SUs are classified based on their minimum data rate requirement. Let R_1 be the minimum rate that must be satisfied for users in category 1, R_2 the minimum rate that must be satisfied for users in category 2 and so on, so that R_v is the minimum rate requirement for category v SUs. The minimum rate constraints for the different categories of SUs can now be written as follows:

$$\sum_{n=1}^{N} c_{k,n} \geq R_1;\ k = 1, 2, \cdots, K_1 \tag{3.3}$$

$$\sum_{n=1}^{N} c_{k,n} \geq R_2;\ k = K_1+1, K_1+2, \cdots, K_1+K_2 \tag{3.4}$$

$$\sum_{n=1}^{N} c_{k,n} \geq R_3;\ k = K_1+K_2+1, K_1+K_2+2, \cdots, K_1+K_2+K_3 \tag{3.5}$$

$$\vdots$$

$$\sum_{n=1}^{N} c_{k,n} \geq R_v;\ k = (K_1+K_2+\ldots+K_{v-1}+1), (K_1+K_2+\ldots+K_{v-1}+2), \cdots, (K_1+K_2+\ldots+K_v) \tag{3.6}$$

In the following subsections, the actual RA formulations are presented one after another, based on the different heterogeneous user classifications provided in Table 3.1. In the considerations, the heterogeneous classes have been limited to two categories for each case. This is simply to make the model more manageable, and for the results to be easier to understand and compare. The models developed can however be easily extended to three, four or any given number of user categories, following the general formulation presented above, without a significant change in the results of the network.

3.5.2 Classification based on minimum rate requirement

In this subsection, the heterogeneous CRN are classified based on their minimum rate requirements. Hence, the K heterogeneous SUs are sub-divided into two categories and are differentiated as K_1: high-rate demand (HD) users, and $(K - K_1)$: low-rate demand (LD) users. The categories are differentiated in that, they have different minimum data rate demand.

Using the representations already defined in the system model, the RA optimisation problem for heterogeneous CRN with the user heterogeneity based on the minimum rate demands of the different user categories is thus formulated as:

$$\max z = \sum_{n=1}^{N} \left(\sum_{k=1}^{K_1} w_1 c_{k,n} + \sum_{k=K_1+1}^{K} w_2 c_{k,n} \right); c_{k,n} \in \{0, 1, 2, 4, 6\} \tag{3.7}$$

subject to

$$\sum_{n=1}^{N} c_{k,n} \geq R_I;\ k = 1,2,\cdots,K_1 \tag{3.8}$$

$$\sum_{n=1}^{N} c_{k,n} \geq R_{II};\ k = K_1+1, K_1+2,\cdots,K \tag{3.9}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} P_{k,n} \leq P_{\max} \tag{3.10}$$

$$\sum_{n=1}^{N} \Phi_n H_{l,n}^{p} \leq \varepsilon_l;\ l = 1,2,...,L \tag{3.11}$$

$$c_{k,n} = 0\ if\ c_{k',n} \neq 0,\ \forall k' \neq k;\ k = 1,2,...,K \tag{3.12}$$

where R_I is the minimum data rate that must be assigned to the kth SU in category one and R_{II} is the minimum data rate that must be assigned to the kth SU in category two, w_1 is the weight attached to the SUs in category one and w_2 is the weight attached to the SUs in category two, $\Phi_n = \Sigma_{k=1}^{K} P_{k,n}$ is the total power of the nth subchannel, $P_{k,n}$ is the transmit power of the kth SU over the nth subchannel, $H_{l,n}^{p}$ is the magnitude of the interference channel gain between the lth PU and the SUBS over the nth subchannel, ε_l is the threshold interference power to the lth PU from all the SUs in the network and $P_{\max}$ is the maximum transmit power at the SUBS.

The objective function in equation (3.7) gives the total weighted data rate achievable by all the SUs in the network. Constraints of equations (3.8) and (3.9) show that the respective minimum data rate for category one and category two users must be met. The constraint in equation (3.10) explains that the total transmit power of all the SUs cannot be greater than the maximum transmit power of the SUBS. The constraint in equation (3.11) shows that the interference from all the SUs to each PU must not be greater than that which each PU can accommodate. The constraint in equation (3.12) is the mutually exclusive constraint, which implies that no single subchannel can be assigned to two or more SUs. In other words, data rate in subchannel n must be 0 for user k if the subchannel n has been assigned to any other user k' that is not k.

The above formulation of the RA problem is non-linear because the power constraint in equation (3.10) is not a linear function. To make the problem solvable, after studying the problem structure, the non-linear optimisation problem is reformulated as an ILP problem. The reformulation is carried out next.

3.5.2.1 Integer linear programming reformulation of problem

By a careful study of the structure of the non-linear, complex NP-hard problem, two important facts are identified and used in achieving the ILP reformulation of the original problem. Firstly, it is observed that the bit allocation to the various subchannels is actually integer in nature. Secondly, the subchannels may either be allocated bit(s) to transmit (usually when their channel interference to PUs is within some acceptable limit) or they may not be assigned any bit to transmit (if their channel interference to PUs is too high). These facts are exploited in achieving a linear reformulation of the original problem. The ILP reformulation of the developed problem is carried out as follows:
Let $\boldsymbol{x}_1$ be a bit allocation vector for all the subchannels of category one users and $\boldsymbol{x}_2$ be a bit allocation vector for all the subchannels of category two users. $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are defined as:

$$\boldsymbol{x}_1 = [(\boldsymbol{x}_{1,N}^1)^T \ (\boldsymbol{x}_{1,N}^2)^T \ \cdots \ (\boldsymbol{x}_{1,N}^N)^T]^T \ \in \{0,1\}^{NK_1C\times 1} \tag{3.13}$$

$$\boldsymbol{x}_2 = [(\boldsymbol{x}_{2,N}^1)^T \ (\boldsymbol{x}_{2,N}^2)^T \ \cdots \ (\boldsymbol{x}_{2,N}^N)^T]^T \ \in \{0,1\}^{N(K-K_1)C\times 1} \tag{3.14}$$

where $\boldsymbol{x}_{1,N}^n = [x_{1,1,n}^T \ x_{1,2,n}^T \ \cdots \ x_{1,K_1,n}^T]^T \ \in \{0,1\}^{K_1C\times 1}$ shows that the nth subchannel is allocated with $\boldsymbol{x}_{1,k,n} = [x_{k,n,1} \ x_{k,n,2} \ \cdots \ x_{k,n,C}\ ]^T \ \in \{0,1\}^{C\times 1}$; $n = 1,\cdots,N$; $k = 1,\cdots,K_1$; C is the number of modulation schemes considered and in this chapter, $C = 4$. This implies that, $\boldsymbol{x}_{1,k,n} = [x_{k,n,1} \ x_{k,n,2} \ x_{k,n,3} \ x_{k,n,4}\ ]^T$. Similar explanations apply for $\boldsymbol{x}_2$. The combined bit allocation vector $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$. Because of the mutually exclusive constraint, $\boldsymbol{x}_{1,N}^n$ and $\boldsymbol{x}_{2,N}^n$ can only take values from $\{[0\,0\ldots0]^T, [1\,0\ldots0]^T, [0\,1\ldots0]^T, \ldots, [0\,0\ldots1]^T\}$. This indicates that, at most one component in $\boldsymbol{x}_{1,N}^n$ is 1 and the other components are 0s (same applies for $\boldsymbol{x}_{2,N}^n$). $x_{k,n,c}$ being 1 means that the nth subchannel is assigned to the kth user, which transmits c number of bits per OFDMA symbol. If all the components of $\boldsymbol{x}_{1,N}^n$ (or $\boldsymbol{x}_{2,N}^n$) are 0s, the nth subchannel is not assigned to any user.

The modulation order vectors for the two categories of users $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are defined as:

$$\boldsymbol{b}_1 = [(\boldsymbol{b}_{1,N}^1)^T \ (\boldsymbol{b}_{1,N}^2)^T \ \cdots \ (\boldsymbol{b}_{1,N}^N)^T]^T \ \in \mathbb{Z}^{NK_1C\times 1} \tag{3.15}$$

$$\boldsymbol{b}_2 = [(\boldsymbol{b}_{2,N}^1)^T \ (\boldsymbol{b}_{2,N}^2)^T \ \cdots \ (\boldsymbol{b}_{2,N}^N)^T]^T \ \in \mathbb{Z}^{N(K-K_1)C\times 1} \tag{3.16}$$

where $\boldsymbol{b}_{1,N}^{n} = [b_{1,1,n}^{T}\, b_{1,2,n}^{T} \cdots b_{1,K_1,n}^{T}]^{T} \in \mathbb{Z}^{K_1C\times 1}$ and $\boldsymbol{b}_{1,k,n} = [b_{k,n,1}\, b_{k,n,2} \cdots b_{k,n,C}]^{T} \in \mathbb{Z}^{C\times 1}$. Similar explanations also apply for $\boldsymbol{b}_2$. Since only four modulation schemes (BPSK, 4-QAM, 16-QAM and 64-QAM) are considered, $b_{k,n} = [1\ 2\ 3\ 4]^{T}$. The data rate matrices for the two categories of SUs, $\boldsymbol{B}_i \in \mathbb{Z}^{K_1\times NK_1C}$ and $\boldsymbol{B}_j \in \mathbb{Z}^{(K-K_1)\times N(K-K_1)C}$ are defined respectively as:

$$\boldsymbol{B}_i = \begin{bmatrix} b_1 & b_1 & \cdots & b_1 \\ b_2 & b_2 & \cdots & b_2 \\ \vdots & \vdots & \ddots & \vdots \\ b_{K_1} & b_{K_1} & \cdots & b_{K_1} \end{bmatrix}, \boldsymbol{B}_i \in \mathbb{Z}^{K_1\times NK_1C} \tag{3.17}$$

$$\left\{ \begin{array}{llll} b_1 = [b^T & 0_C^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times K_1C} \\ b_2 = [0_C^T & b^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times K_1C} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K_1} = [0_C^T & 0_C^T & \cdots & b^T] \in \mathbb{Z}^{1\times K_1C} \end{array} \right\}$$

$$\boldsymbol{B}_j = \begin{bmatrix} b_{K_1+1} & b_{K_1+1} & \cdots & b_{K_1+1} \\ b_{K_1+2} & b_{K_1+2} & \cdots & b_{K_1+2} \\ \vdots & \vdots & \ddots & \vdots \\ b_K & b_K & \cdots & b_K \end{bmatrix}, \boldsymbol{B}_j \in \mathbb{Z}^{(K-K_1)\times N(K-K_1)C} \tag{3.18}$$

$$\left\{ \begin{array}{llll} b_{K_1+1} = [b^T & 0_C^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times (K-K_1)C} \\ b_{K_1+2} = [0_C^T & b^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times (K-K_1)C} \\ \vdots & \vdots & \ddots & \vdots \\ b_K = [0_C^T & 0_C^T & \cdots & b^T] \in \mathbb{Z}^{1\times (K-K_1)C} \end{array} \right\}$$

Given that the rate weight for category one SUs is w_1 and the rate weight for category two SUs is w_2, the total data rate in the objective function (3.7) can thus be written as $\max_x[(w_1 \odot \boldsymbol{b}_1)^T \boldsymbol{x}_1 + (w_2 \odot \boldsymbol{b}_2)^T \boldsymbol{x}_2]$, where $\odot$ is the Schur-Hadamard (or entry-wise) product. By defining $\boldsymbol{R}_I \triangleq [R_1\ R_2 \cdots R_{K_1}]^T \in \mathbb{R}^{K_1\times 1}$ and $\boldsymbol{R}_{II} \triangleq [R_{K_1+1}\ R_{K_1+2} \cdots R_K]^T \in \mathbb{R}^{(K-K_1)\times 1}$, the data rate per user constraint of equation (3.8) can be written as $\boldsymbol{B}_i\boldsymbol{x}_1 \geq \boldsymbol{R}_I$ while for equation (3.9), the data rate constraint can be written as $\boldsymbol{B}_j\boldsymbol{x}_2 \geq \boldsymbol{R}_{II}$.

For the constraint in equation (3.10), the power transmission vector $\boldsymbol{p}$ is defined as:

$$\boldsymbol{p} = [(\boldsymbol{p}_N^1)^T\ (\boldsymbol{p}_N^2)^T\ \cdots\ (\boldsymbol{p}_N^N)^T]^T\ \in \mathbb{R}^{NKC\times 1} \tag{3.19}$$

where $\boldsymbol{p}_N^n = [\boldsymbol{p}_{1,n}^T\ \boldsymbol{p}_{2,n}^T\ \cdots\ \boldsymbol{p}_{K,n}^T]^T\ \in \mathbb{R}^{KC\times 1}$ and $\boldsymbol{p}_{k,n} = [p_{k,n,1}\ p_{k,n,2}\ \cdots\ p_{k,n,C}\ ]^T\ \in \mathbb{R}^{C\times 1}$; $p_{k,n,c}$ is the required power to transmit c number of bits over the nth subchannel for the kth user. The power constraint in equation (3.10) can then be written as $\boldsymbol{p}^T\boldsymbol{x} \leq P_{\max}$.

In order to write the interference power constraint in equation (3.11) in terms of the vector $\boldsymbol{x}$, a matrix $\boldsymbol{A} \in \{0,1\}^{N\times NKC}$ is defined as follows:

$$\boldsymbol{A} = \begin{bmatrix} 1_{KC}^T & 0_{KC}^T & \cdots & 0_{KC}^T \\ 0_{KC}^T & 1_{KC}^T & \cdots & 0_{KC}^T \\ \vdots & \vdots & \ddots & \vdots \\ 0_{KC}^T & 0_{KC}^T & \cdots & 1_{KC}^T \end{bmatrix}, \boldsymbol{A} \in \{0,1\}^{N\times NKC} \tag{3.20}$$

$$1_{KC} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \{1\}^{KC\times 1}, \qquad 0_{KC} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \{0\}^{KC\times 1}$$

Given that $\boldsymbol{p} \odot \boldsymbol{x}$ is the Schur-Hadamard product of $\boldsymbol{p}$ and $\boldsymbol{x}$, $\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})$ is therefore an $N \times 1$ vector whose nth element characterises the total power used for the nth subchannel. Defining $\boldsymbol{\varepsilon}_l \triangleq [\varepsilon_1\ \varepsilon_2\ ...\ \varepsilon_L]^T \in \mathbb{R}^{L\times 1}$, the interference power constraint in equation (3.11) can be written as:

$$\boldsymbol{H}^p[\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})] \leq \boldsymbol{\varepsilon}_l. \tag{3.21}$$

The RA problem for heterogeneous CRN given in equations (3.7) - (3.12) can now be rewritten in the ILP form as:

$$z^* = \max_x [(w_1 \odot \boldsymbol{b}_1)^T \boldsymbol{x}_1 + (w_2 \odot \boldsymbol{b}_2)^T \boldsymbol{x}_2] \tag{3.22}$$

subject to

$$\boldsymbol{B}_i\boldsymbol{x}_1 \geq \boldsymbol{R}_I;\ k = 1,2,\cdots,K_1 \tag{3.23}$$

$$\boldsymbol{B}_j\boldsymbol{x}_2 \geq \boldsymbol{R}_{II};\ k = K_1+1, K_1+2,\cdots,K \tag{3.24}$$

$$\boldsymbol{p}^T\boldsymbol{x} \leq P_{\max} \tag{3.25}$$

$$\boldsymbol{H}^p[\boldsymbol{A}(\boldsymbol{p}\odot\boldsymbol{x})] \leq \boldsymbol{\varepsilon}_l \tag{3.26}$$

$$\boldsymbol{0}_N \leq \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{1}_N \tag{3.27}$$

$$\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x} \in \{0,1\}\ w_1,\ w_2 \in \mathbb{R}^+. \tag{3.28}$$

The optimisation problem in equations (3.22) - (3.28) is a combinatorial ILP problem which, in this thesis, is solved by the branch-and-bound (BnB) method, a very adequate technique for solving such linear programming problems. To reduce the complexity in computation, the implicit enumeration method, which is a special case of BnB that solves binary integer LP problems, is employed [51]. Implicit enumeration makes use of the fact that each variable (in this case, the bit allocation vector $\boldsymbol{x}$) must be equal to 0 or 1 and uses this information to simplify both the branching and bounding components of the BnB process, and to determine efficiently when a node is infeasible, thus reducing the overall computational complexity of the network.

3.5.3 Classification based on user priority or sensitivity

In this subsection, the heterogeneous classification of users is based on either the priority of the SUs or their sensitivity to changes within the network. In terms of priority, the SUs are categorised into two - high priority (HP) users and best effort service (BE) users. With this priority classification, category one HP SUs do have the higher priority and their demands are first met. The remaining resources are thereafter proportionally shared among the category two BE SUs based on a proportional rate constraint. In terms of sensitivity, the users are categorised as either sensitive users (XU) or general users (GU). The sensitivity in this classification is dependent on the data transfer rate requirement. Users in the XU category are indeed more sensitive in that they require guaranteed QoS, hence, a minimum transfer rate must be assigned to them to meet their demands at all times. The users in this category may have applications like audio and video communications that require constant data transfer at an acceptable rate for satisfactory QoS delivery. Users in the GU category are less sensitive and have less QoS requirement as compared to the XU users. GU users may be users that provide services like emails, short (text) messaging, web surfing or downloading etc.

While these two classifications (i.e., in terms of priority or sensitivity) are slightly different from one another, the problem formulations and analyses for both classifications are however similar, hence, it is appropriate to group and study them together in this subsection. The K heterogeneous SUs in the two categories are differentiated as K_1: HP or XU users, and K_2: BE or GU users. In both considerations, the corresponding sets of the two categories of SUs are denoted as κ_A and κ_B respectively. The explanations of the system model given in the previous section is applicable in this consideration as well.

Let R_k be the minimum data rate that must be assigned to the kth SU in κ_A, γ_k be the predetermined value of the normalised proportional fairness factor for each SU in κ_B, data rate R_i indicate the rate for the element i in κ_B, let w_1 be the weight of the kth SU in κ_A and w_2 be the weight of the kth SU in κ_B. All other representations previously defined in the system model are equally applicable. The RA optimisation problem for heterogeneous CRN with priority or sensitivity considerations is thus formulated as:

$$\max z = \sum_{n=1}^{N} \left(\sum_{k=1}^{K_1} w_1 c_{k,n} + \sum_{k=1}^{K_2} w_2 c_{k,n} \right); c_{k,n} \in \{0,1,2,4,6\} \tag{3.29}$$

subject to

$$\sum_{n=1}^{N} c_{k,n} \geq R_k;\ \forall k \in \kappa_A \tag{3.30}$$

$$\frac{R_k}{\sum\limits_{i \in \kappa_B} R_i} = \gamma_k;\ \forall k \in \kappa_B \tag{3.31}$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} P_{k,n} \leq P_{\max} \tag{3.32}$$

$$\sum_{n=1}^{N} \Phi_n H_{l,n}^{p} \leq \varepsilon_l;\ l = 1,2,\ldots,L \tag{3.33}$$

$$c_{k,n} = 0\ if\ c_{k',n} \neq 0,\ \forall k' \neq k;\ k = 1,2,\ldots,K \tag{3.34}$$

The objective function (3.29) gives the total weighted data rate (throughput) achievable by all the SUs (in both categories) of the network. The constraint in equation (3.30) shows that the minimum data transfer rate for each HP or XU user of category one must be met. In equation (3.31), a proportional fairness factor is used to determine how much of the capacity left is assigned to each user in category two, the BE or GU user category. As earlier explained, equation (3.32) is the total transmit power constraint for all the SUs, equation (3.33) is the maximum interference constraint and equation (3.34) is the mutually exclusive constraint. It is easy to show that equation (3.31) can be equivalently rewritten as:

$$R_k = \gamma_k \times \sum_{i \in \kappa_B} R_i,$$

where $\sum_{i \in \kappa_B} R_i$ is the constant value of the sum of all the data rates of all category two users. Let the product $\gamma_k \times \sum_{i \in \kappa_B} R_i$ be represented as $\tilde{\gamma}_k$, then,

$$R_1 : R_2 : \ldots : R_{K_2} = \tilde{\gamma}_1 : \tilde{\gamma}_2 : \ldots : \tilde{\gamma}_{K_2} \; \forall k \in \kappa_B. \tag{3.35}$$

Similar to the formulation in the previous heterogeneous consideration, the new formulation of the RA problem presented above is a non-linear programming problem because the power constraint in equation (3.32) is not a linear function. Again, just as in the previous subsection, to make the problem solvable, it is reformulated as an ILP problem and then solved using BnB. The ILP reformulation follows the same procedure as described in the previous subsection and it is therefore not necessary to repeat the process. The newly reformulated ILP problem of RA for heterogeneous CRN, given priority or sensitivity considerations, is therefore presented as:

$$z^* = \max_x [(w_1 \odot \boldsymbol{b}_1)^T \boldsymbol{x}_1 + (w_2 \odot \boldsymbol{b}_2)^T \boldsymbol{x}_2] \tag{3.36}$$

subject to

$$\boldsymbol{B}_i \boldsymbol{x}_1 \geq R_k; \; \forall k \in \kappa_A \tag{3.37}$$

$$\boldsymbol{B}_j \boldsymbol{x}_2 = \tilde{\gamma}_k; \; \forall k \in \kappa_B \tag{3.38}$$

$$\boldsymbol{p}^T \boldsymbol{x} \leq P_{\max} \tag{3.39}$$

$$\boldsymbol{H}^p [\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})] \leq \boldsymbol{\varepsilon}_l \tag{3.40}$$

$$\boldsymbol{0}_N \leq \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{1}_N \tag{3.41}$$

$$\boldsymbol{x}_{1,2} \in \{0,1\},\ w_1, w_2 \in \mathbb{R}^+. \tag{3.42}$$

As established in the previous subsection, the ILP problem in equations (3.36) - (3.42) is a combinatorial linear programming problem which, in this thesis, is solved using the BnB method for solving ILP problems.

3.5.4 Classification based on delay tolerance

In this section, the SUs are classified based on their delay characteristics. The K heterogeneous SUs are differentiated as: K_1, representing the delay-sensitive (DS) users, and K_2, representing the delay-tolerant (DT) users. The corresponding sets of these two categories of SUs are also denoted as κ_A and κ_B respectively. The DS SUs in category one, because of their delay sensitivity, constantly have a minimum rate guarantee for their service to be acceptable. The DT SUs in category two could have a flexible data rate demand. Furthermore, the SUs in both categories might all have buffered data (i.e. data in a queue waiting to be transmitted), but the category two SUs, being DT, can accommodate a longer waiting period than the category one SUs. Users that fit into category one could be SUs that require services that need to be attended to urgently (for instance, in emergency service deliveries like hospital or fire-service ambulances, or service providers during disasters or crises). Such users would therefore prefer that their communications not be initiated than be interrupted or delayed for a long duration before they can be completed. The traffic model of the SUs is described next.

For the DS SUs, their data buffer has a finite capacity. The arrival process of packets is modelled as a Poisson process [37]. The process has an independent arrival rate λ_k (packets/slot) $\forall k \in \kappa_A$, the set of DS SUs. For a user k that falls within this category of SUs, the sum of the average time that its data packets wait in the queue and the time required for service completion by the user gives the average delay duration of data packets, and is represented by the expectation value $\mathbb{E}[D_k]$. The data buffer for DT SUs is defined to be infinitely large such that at every given time, there will always be data packets for them to transmit. The available resources for these SUs are therefore shared proportionately, using a predetermined proportional fairness factor γ_k. Hence, for the set of DT SUs, data rate R_i indicates the rate for the element i in κ_B.

Let the maximum permissible delay duration for an acceptable QoS for each DS SU k (i.e., the delay constraint) be T_k. To meet this required QoS, the average delay during data packet transmission for the DS SU must therefore not exceed the delay constraint. Hence,

$$\mathbb{E}[D_k] \leq T_k,\ \forall k \in \kappa_B. \tag{3.43}$$

From the explanations given above, the optimisation problem of RA for heterogeneous CRN, having SUs with different delay characteristics, is presented as follows:

$$\max z = \sum_{n=1}^{N} \left(\sum_{k=1}^{K_1} w_1 c_{k,n} + \sum_{k=1}^{K_2} w_2 c_{k,n} \right); c_{k,n} \in \{0,1,2,4,6\} \tag{3.44}$$

subject to

$$\sum_{n=1}^{N} c_{k,n} \geq R_k;\ \forall k \in \kappa_A \tag{3.45}$$

$$\mathbb{E}[D_k] \leq T_k,\ \forall k \in \kappa_A \tag{3.46}$$

$$\frac{R_k}{\sum_{i \in \kappa_B} R_i} = \gamma_k;\ \forall k \in \kappa_B \tag{3.47}$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} P_{k,n} \leq P_{\max} \tag{3.48}$$

$$\sum_{n=1}^{N} \Phi_n H_{l,n}^{p} \leq \varepsilon_l;\ l = 1,2,\ldots,L \tag{3.49}$$

$$c_{k,n} = 0\ if\ c_{k',n} \neq 0,\ \forall k' \neq k;\ k = 1,2,\ldots,K \tag{3.50}$$

where R_k is the minimum data rate that must be assigned to the kth SU of DS users, w_1 is the weight of the kth SU in κ_A and w_2 is the weight of the kth SU in κ_B. The other representations are as previously defined.

The objective function (3.44) gives the total data rate that the CRN can deliver. Equations (3.45) and (3.46) are specifically for the DS SUs. The constraint in equation (3.45) gives the minimum rate, while equation (3.46) is the permissible time delay constraint for the DS SUs. In equation (3.47), the proportional fairness factor is used to assign data rates to each user in the DT category of SUs. Similar to the previous cases considered, equation (3.48) shows that there is a total transmit power constraint for all SUs, equation (3.49) gives the constraint on the permissible interference to PUs and equation (3.50) is the mutual exclusivity constraint.

Again, the formulated problem given in equations (3.44) - (3.50) is a non-linear optimisation problem since the power constraint in equation (3.48) is not linear. Similar to the other problems already discussed, to solve this problem, an ILP reformulation of the initial problem is realised. The reformulation follows the process already explained in the previous subsections. The ILP reformulated problem is thus given as:

$$z^* = \max_{x}[(w_1 \odot \boldsymbol{b}_1)^T \boldsymbol{x}_1 + (w_2 \odot \boldsymbol{b}_2)^T \boldsymbol{x}_2] \tag{3.51}$$

subject to

$$\boldsymbol{B}_i \boldsymbol{x}_1 \geq R_k;\ \forall k \in \kappa_A \tag{3.52}$$

$$\mathbb{E}[D_k] \leq T_k,\ \forall k \in \kappa_A \tag{3.53}$$

$$\boldsymbol{B}_j \boldsymbol{x}_2 = \tilde{\gamma}_k;\ \forall k \in \kappa_B \tag{3.54}$$

$$\boldsymbol{p}^T \boldsymbol{x} \leq P_{\max} \tag{3.55}$$

$$\boldsymbol{H}^p[\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})] \leq \boldsymbol{\varepsilon}_l \tag{3.56}$$

$$\boldsymbol{0}_N \leq \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{1}_N \tag{3.57}$$

$$\boldsymbol{x}_{1,2} \in \{0,1\},\ w_1,\ w_2 \in \mathbb{R}^+ \tag{3.58}$$

Similar to the earlier ones presented, the ILP problem in equations (3.51) - (3.58) is a combinatorial ILP problem. Therefore, the BnB method is also employed in obtaining solutions, as used in solving previous problems.

3.6 RESULTS AND DISCUSSION

This section presents results for the RA solutions of all the heterogeneous CRN considerations analysed in this chapter. The underlay, heterogeneous, OFDMA-based CRN, as described in the system model, is simulated using the MATLAB software and the optimisation is carried out using the YALMIP toolbox developed for solving optimisation problems [15]. The general simulation parameters for all the results presented are: number of OFDMA subchannels $N = 64$, PUs $L = 4$ and SUs $K = 4$. The SUs, from the earlier classifications, are categorised as: category one $K_1 = 2$ (representing the HD, HP, XU or DS SUs) and category two $K - K_1(\text{or } K_2) = 2$ (representing the LD, BE, GU or DT SUs). The choice of the number of PUs, SUs and other parameters used in the simulation is informed by the need to compare results obtained in this chapter with similar works in the literature so as to validate the results. For all simulation results presented in this chapter, random multipath fading channels of length six were generated for the PUs and SUs using statistically independent Gaussian random variables. The average channel gain between SUBS and PUs was set at 0.1 while the gain between the SUBS and SUs was set at 1. The maximum interference limit to PUs was set as 0.001mW while the interference caused by the PUs, considered as noise by the SUs, had a power spectral density of $(0.01/64)$mW/subchannel. All the simulation results were obtained using 100 randomly generated channel pairs H^s and H^p. The required BER ρ has a value of 0.01 for all SUs. A weight of unity for for all SU categories in considered, except in the final results where the effects of weight are explored. The results are discussed in subsequent subsections based on the various classifications carried out in the previous section, and in the order of their presentation.

3.6.1 Results based on minimum data rate classification

For the results discussed in this subsection, the minimum data rate for the HD category one SUs is 64bits/user and for the LD category two SUs, it is 32 bits/user. Generally, because they require a higher data rate, the category one SUs might be the users who are billed higher, or there might be some other criteria by which they are charged to pay for the better QoS being provided for them.

First, the work in [66] is re-simulated and its results reproduced as Fig. 3.3, to validate the simulation results obtained and discussed in this chapter. For instance, the results presented in Figs. 3.4 - 3.6 are very similar to, and consistent with results of [66] and [65], thus validating the simulations. Since other results presented in this chapter (and in subsequent chapters) are derived based on the principles on which the results presented in Figs. 3.4 - 3.6 are obtained, these reproduced results of [66] serves as a strong validation for the analyses and simulations carried out and presented in the thesis.

Fig. 3.4 shows the interference channel gain between the SUBS and the PUs while Fig. 3.5 shows the channel gain between the SUBS and the SUs. The channel gain is important in that it influences the

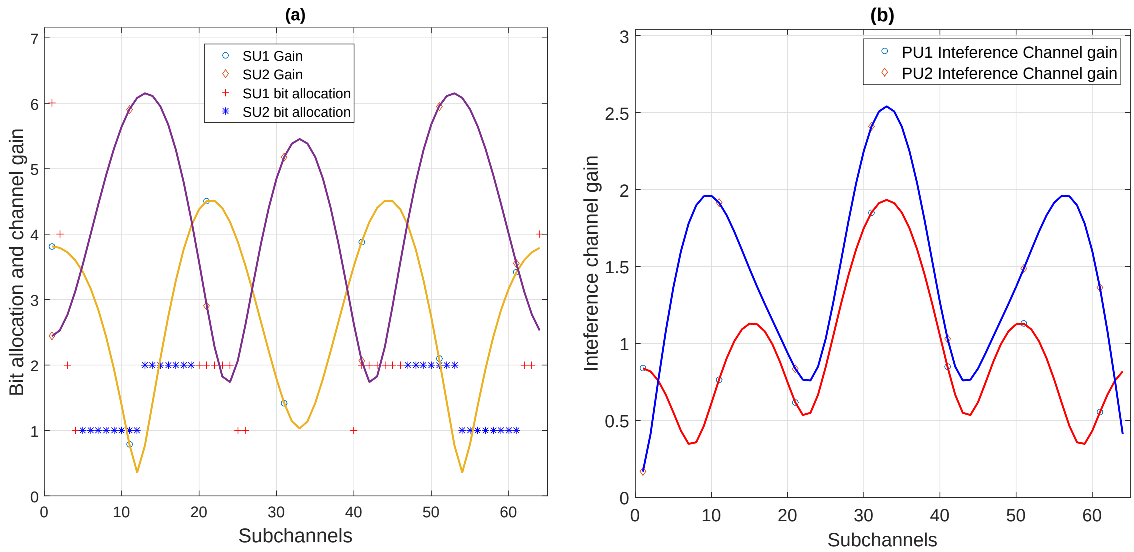

Figure 3.3. Simulation results from [66] reproduced to validate the simulations carried out in this thesis. (a) Subchannel gains between the SUs and SUBS, as well as bit allocation on each subchannel for each SU (b) Interference channel gain between the PUs and the SUBS.

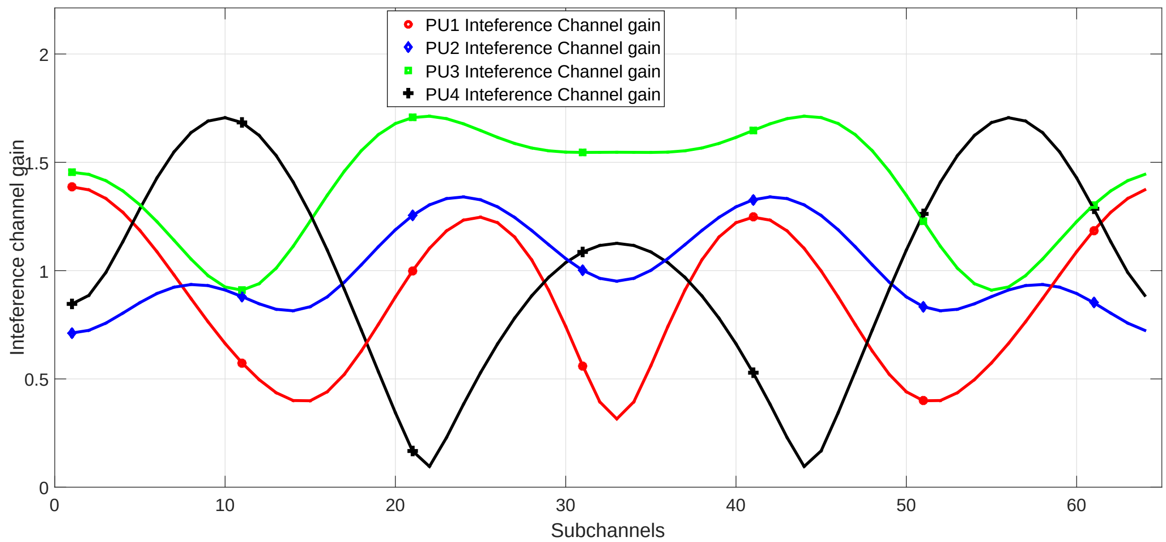

Figure 3.4. Interference channel gain between the PUs and the SUBS.

allocation of data rates to each of the users. The actual data rate (bits per symbol) allocated to each of the SUs is shown in Fig. 3.6. To explain the bit allocation in Fig. 3.6, an 'x' at a bit allocation of 6 for subchannel 9 means that subchannel 9 has been allocated to SU 3 to transmit 6 bits. It is significant to note that the bit allocation is done with careful consideration of the interference gains to the PUs. At high interference gain (which signifies low or less fading), the subchannels are allocated low data rates.

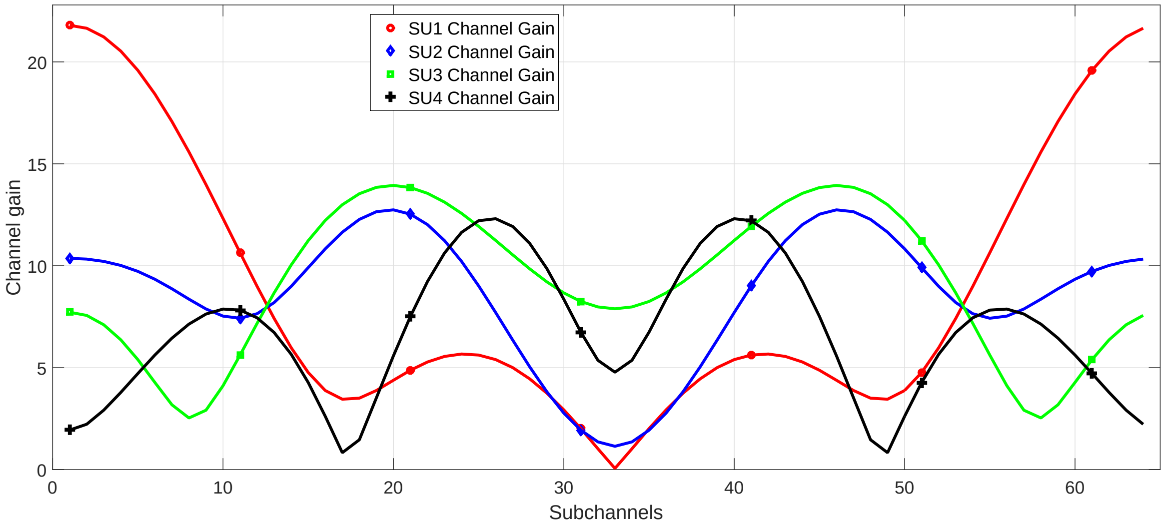

Figure 3.5. Interference channel gain between the SUs and the SUBS.

This is because high data rates (from a high modulation scheme) on a subchannel will require high power and if the interference gain on such subchannel is high, the adverse effect on the PUs will be considerably high. The converse of this is also true - low interference gains (signifying high or deep fading) can accommodate a high data rate on a subchannel (to transmit with high modulation), and this with minimal interference effect on the PUs. The SUBS thus avoids higher order modulation (e.g. 64-QAM) to the subchannels with high interference channel gains, in order to reduce the amount of interference to the PUs.

One important contribution from this optimal allocation algorithm developed in this chapter, that needs to be stressed, is the 'smartness' with which the RA procedure is carried out in order to achieve optimality for the heterogeneous CRN. The simple but profound 'sense' the algorithm catches in on is the fact that higher order modulations require more power and therefore, employing such on those subchannels with high interference channel gains to PUs will cause more harm to the PUs occupying them; hence the need to either avoid them completely or assign low data rates to those subchannels. Examples of this smart exploitation can be seen in subchannels 2, 3, 9, 57, 63 and 64 of Fig. 3.6 where a high data rate has been allocated. The combined interference to the PUs on those subchannels, as seen from Fig. 3.4, is lower than the combined interference on the other subchannels. On subchannels 14 - 27 and 39 - 52, the combined interference to PUs is quite high and the subchannels have been allocated low data rates to transmit. This is the basic principle by which the bit allocation is carried out to obtain optimal results on the overall utility (average data rates, total data rates etc.) of the network.

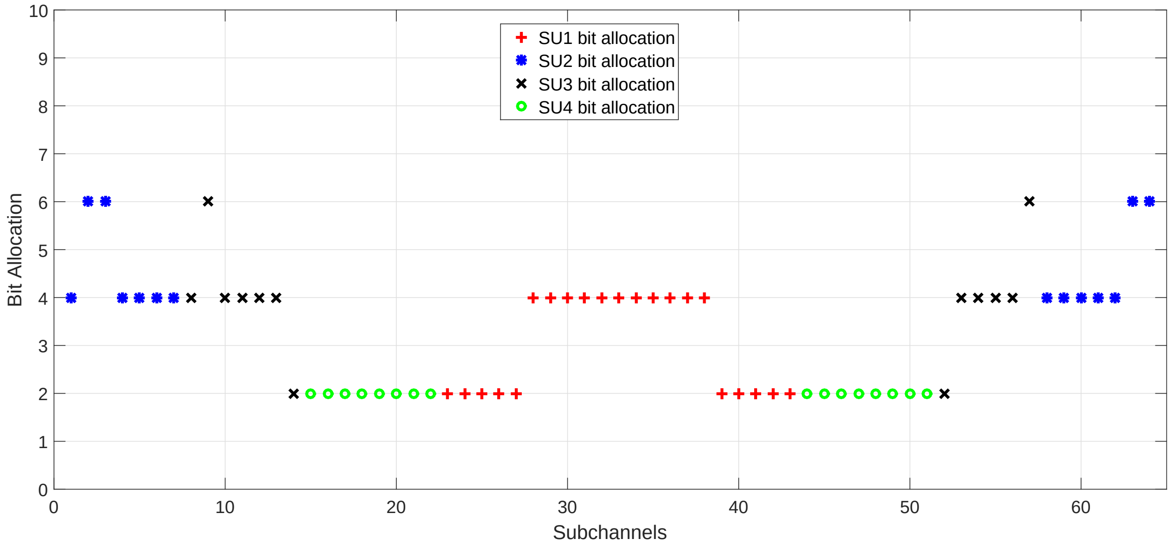

Figure 3.6. SUBS bit allocation for each of the SUs. The bit allocation is carried out with consideration for the PU's current interference channel gain.

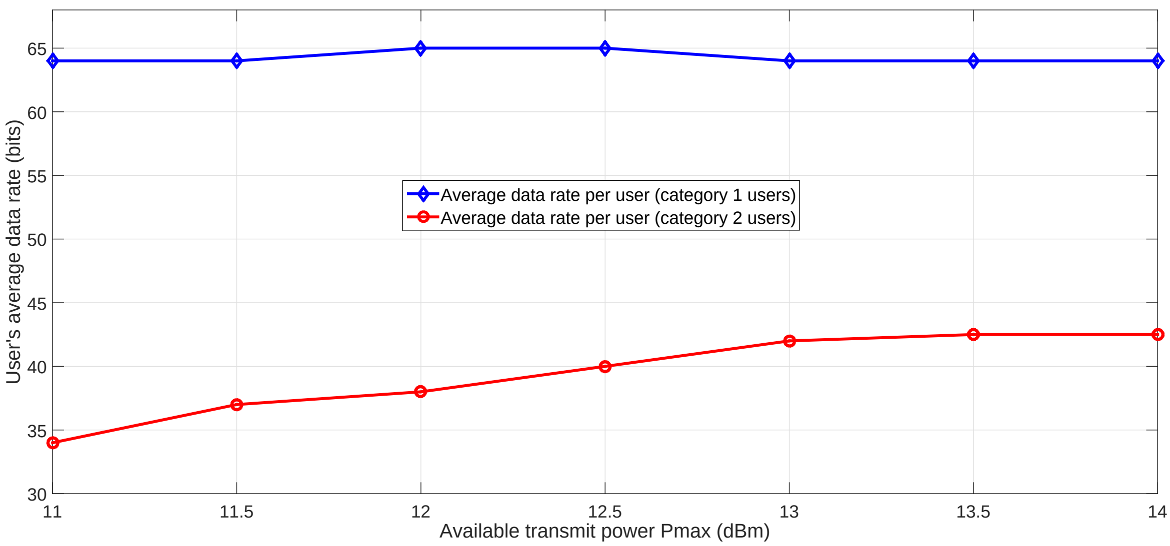

Figure 3.7. Average data rate as a function of available transmit power at the SUBS for the two categories of SUs.

The average data rate of each SU against the maximum transmit power at the SUBS is shown in Fig. 3.7 for the two categories of SUs considered. These results are comparable to similar ones in [65]. To obtain an accurate result, the interference channel gain between the PUs and the SUBS was kept constant as the transmit power of the SUBS was varied. In the plot, the minimum data rate requirement for each category of SUs must at least be satisfied for the optimisation problem to be feasible. The plot

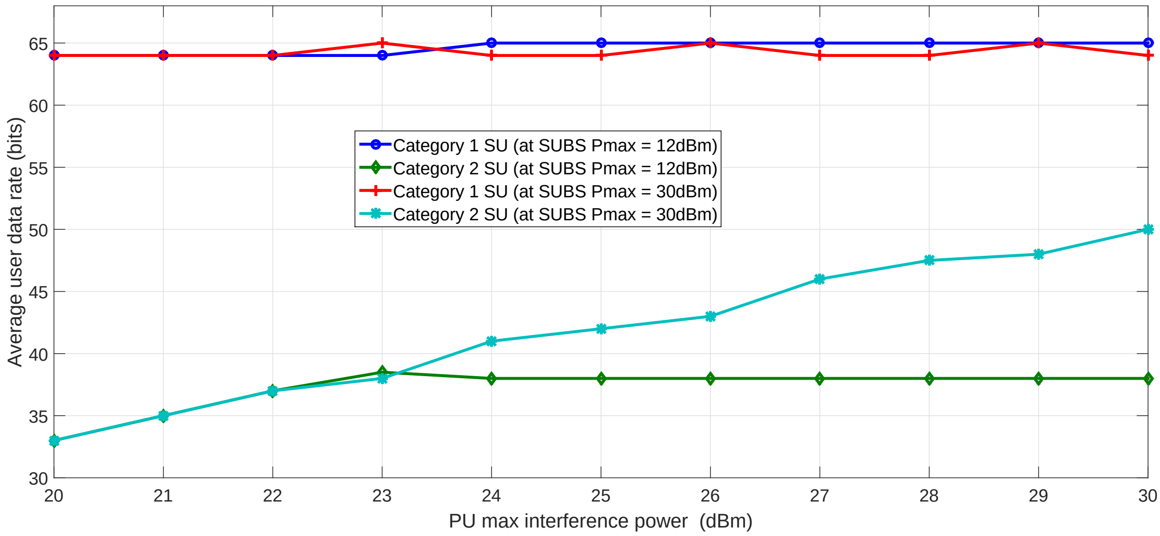

Figure 3.8. Average data rate of users versus maximum interference power to PUs at different SUBS power for the categories of SUs.

also shows that the average data rate increases gradually as the transmit power of the SUBS increases until it gets to a saturation point. After that point, an increase in the transmit power at the SUBS does not cause any further increase in the average data rate of the users. This is because, the other constraints (e.g. the maximum amount of interference power leaked to the PUs) also come into play in the optimisation problem, thereby making it impossible for the SUs' data rates to keep increasing indefinitely with an increase in SUBS transmit power.

3.6.2 Results based on priority and sensitivity classifications

For the results presented in this section, the HP or XU category one SUs K_1 have a minimum data transfer rate requirement of 64 bits/user while BE or GU category two SUs K_2 have the remaining resources proportionately distributed between them with a normalised proportional fairness factor $\gamma_k = 1$.

The average user data rate achieved for each category of SUs over a varying interference power to the PUs is shown in Fig. 3.8. The maximum acceptable interference power to each PU, i.e. ε_l, was varied between $20dBm$ and $30dBm$ with the available SUBS power set at $12dBm$, and then later increased to $30dBm$. It is important to first note that below $20dBm$ interference the problem becomes infeasible. Also, it can be observed that, when the problem is feasible, the minimum data rate requirement for

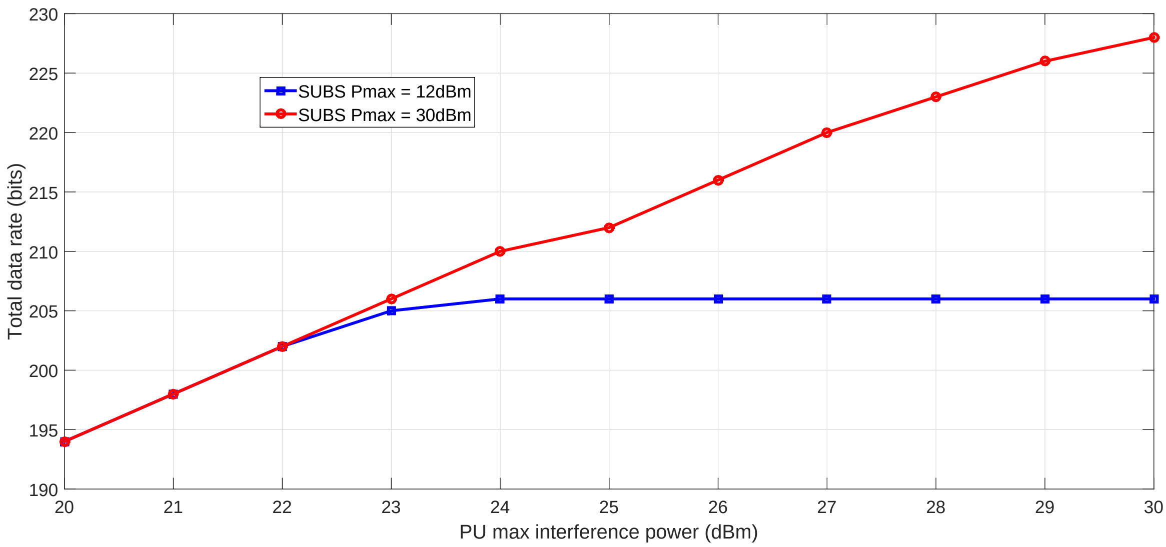

Figure 3.9. Total data rate versus maximum interference power to PUs at different SUBS power.

category one SUs is achieved at all points. Furthermore, the plot shows that the algorithm achieves a similar trend (continuous improvement) until about 24dBm of maximum interference power. Beyond this limit, the average rate for users in both categories begins to stabilise when the SUBS maximum power is at 12dBm. However, the average rate for users in category two SUs increases further and further when the SUBS maximum power is at 30dBm (it would also reach a saturation point if the interference it increased beyond the range used in this result). The reason for this is that, with a higher power at the SUBS, the average data rate of the users is greatly improved if all the other constraints are unchanged. It is also very significant to observe that the algorithm would rather increase the average rate of the category of SUs with BE or GU demand when it has a slightly higher resource than it would have with the category of SUs with a HP or XU demand. This signifies that it is easier to slightly (or even significantly) improve resource allocations to the category of SUs that have the most flexibility (such as the BE or GU SUs) because their demands are a lot easier to satisfy than the demands of the more rigid HP or XU SUs.

In Fig. 3.9, the total data rate or throughput of the system against varying values of interference power to the PUs, is presented. The PUs' maximum interference power is varied between 20dBm and 30dBm for values of SUBS power at 12dBm and 30dBm. The result clearly shows that the CRN will generally achieve a better QoS in terms of throughput as the amount of permissible interference power to the PUs is relaxed (i.e. when the permissible interference power to PUs assume higher values). Also, it can be seen that, for a higher SUBS power (30dBm), the throughput keeps improving, unlike its lower

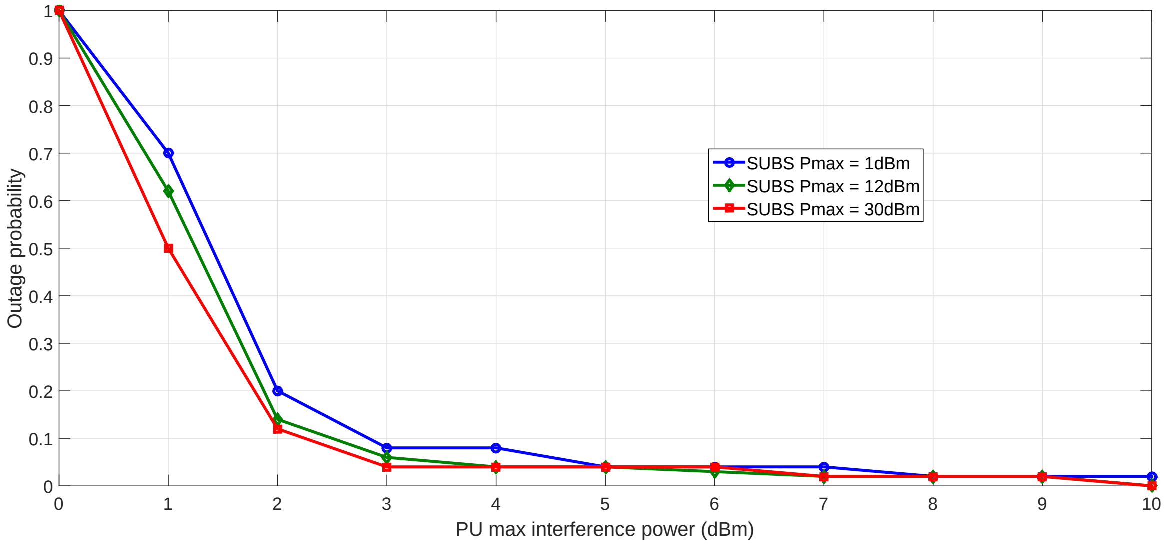

Figure 3.10. Outage probability versus maximum interference power to PUs at different SUBS power.

SUBS power (12dBm) counterpart where the throughput quickly stabilises, even with an increasing interference limit.

The outage probability is the probability that the formulated problem will be infeasible, given the prevalent and/or immediate constraints and conditions under consideration. In Fig. 3.10, the outage probability over a varying amount of interference power to the PUs is shown for different values of SUBS power. From the plot, it can be depicted that the outage probability decreases with an increasing interference power limit to the PUs. It can also easily be observed that the outage probability generally improves (by achieving lower values) with an increasing SUBS power ($P_{\max}$). This implies that, for a given value of interference power to PUs, the outage probability would be better at a higher SUBS power than it would be at a lower SUBS transmit power.

Fig. 3.11 describes the total data rate of the CRN against the maximum transmit power at the SUBS when the number of available SUs in the various categories are differently combined. The maximum permissible interference to PUs has been pegged at 50dBm. From the plot, it can be observed that at an increasing SUBS power, the total data rate of the CRN increases steadily until it saturates. The reason for this is that at a larger value of SUBS power a higher modulation rate (and hence, a larger data rate) is achieved for the SUs. However, the total data rate does not increase indefinitely because at some point other constraints such as the maximum interference to PUs, which are also not to be violated, come into play. The results show further that the more the number of category two users in the network

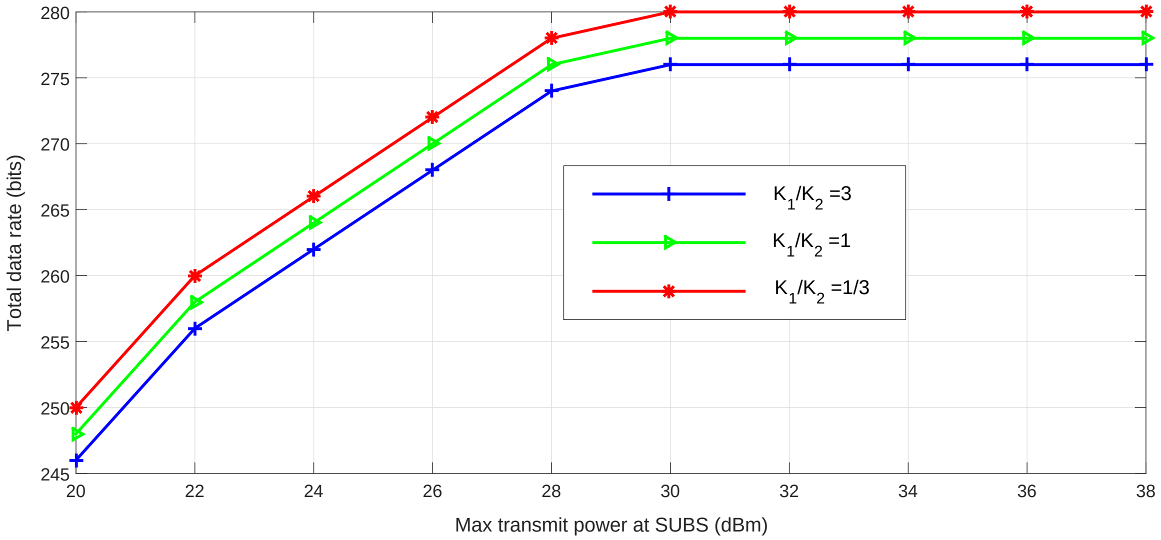

Figure 3.11. Total data rate against the maximum transmit power at the SUBS for different possible combinations of categories of SUs. Maximum permissible interference to PUs is set at 50dBm.

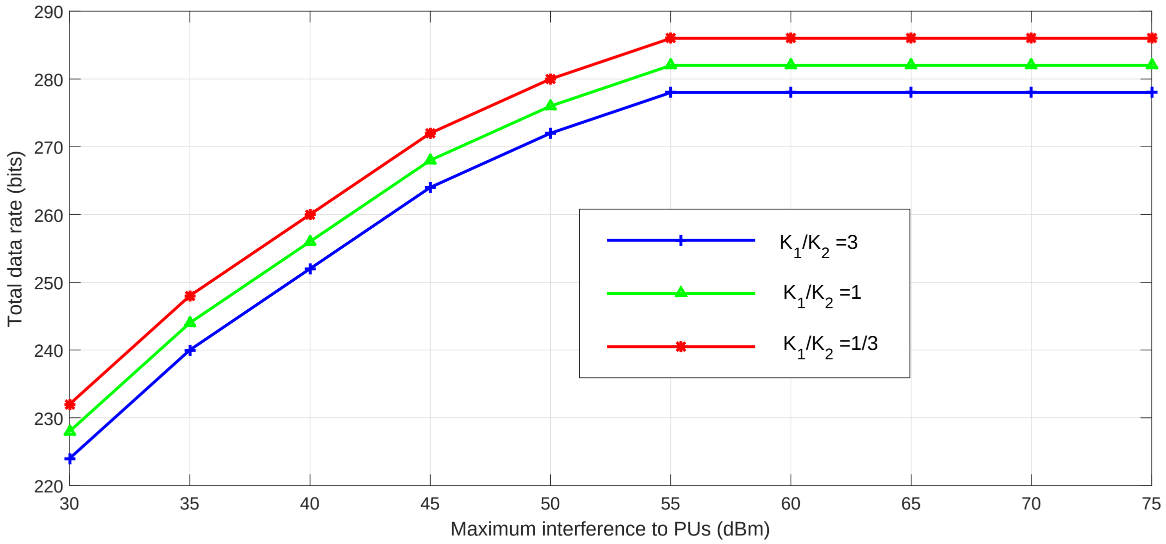

Figure 3.12. Total data rate against the maximum interference to PUs for different possible combinations of categories of SUs. Maximum transmit power at SUBS is set at 30dBm

(in comparison with the category one users), the better the overall throughput of the system. This can be seen in that at $\frac{K_1}{K_2} = 3$ the overall best throughput is achieved. The reason for this is that it is easier to satisfy category two users because of the flexibility in their demand, as compared to the category one users whose rate expectations are higher and quite static.

Fig. 3.12 shows the total data rate of the CRN against the maximum permissible interference to the PUs when the number of the various categories of SUs available in the network are also combined differently. The plot is similar to the previous one, the difference being that the maximum transmit power at the SUBS is fixed in this case and is set to $30dBm$. Again, the plot shows that as the permissible interference to PUs increases, the total data rate of the CRN also increases until it achieves a maximum possible value. Similar to the previous plot, the total data rate does not increase indefinitely because other constraints are also being considered. The plot shows further that the more the number of category two users in the network (in comparison with category one users) the better the overall throughput of the system. The same reason deduced for the previous plot applies - it is easier to satisfy category two users because of the flexibility in their demands.

3.6.3 Results based on delay tolerance classification

The simulation is carried out with the number of category one DS SUs $K_1 = 2$ and given that their minimum data rate requirement is 64 bits/user, while the maximum permissible delay time $T_k = 20ms$. The number of category two DT SUs $K_2 = 2$ and the remaining resources are proportionally distributed among them.

Fig. 3.13 gives the average data rate of each SU against the maximum transmit power at the SUBS for the two categories of SUs considered. The results for the delay tolerance classification are compared with those obtained using the minimum data rate classification already presented in Fig. 3.7. The plot shows that it takes a higher transmit power for the delay tolerant classification to become feasible, as the problem only begins to be solvable at about $12dBm$ SUBS transmit power. Furthermore, the performance of the system with delay tolerance classification was constantly below comparative results obtained from the minimum rate classification. The reason that can be given for these observations is that, for the delay tolerant consideration, a further constraint in terms of the maximum permissible delay duration for the DS SUs is also incorporated into the problem formulation and its effect is what makes the overall performance of the network to be slightly degraded, as compared to only when the minimum rate requirement is considered.

Fig. 3.14 gives the total data rate of each SU against the maximum transmit power at the SUBS for the two categories of SUs considered. The results for the delay tolerance classification are also compared

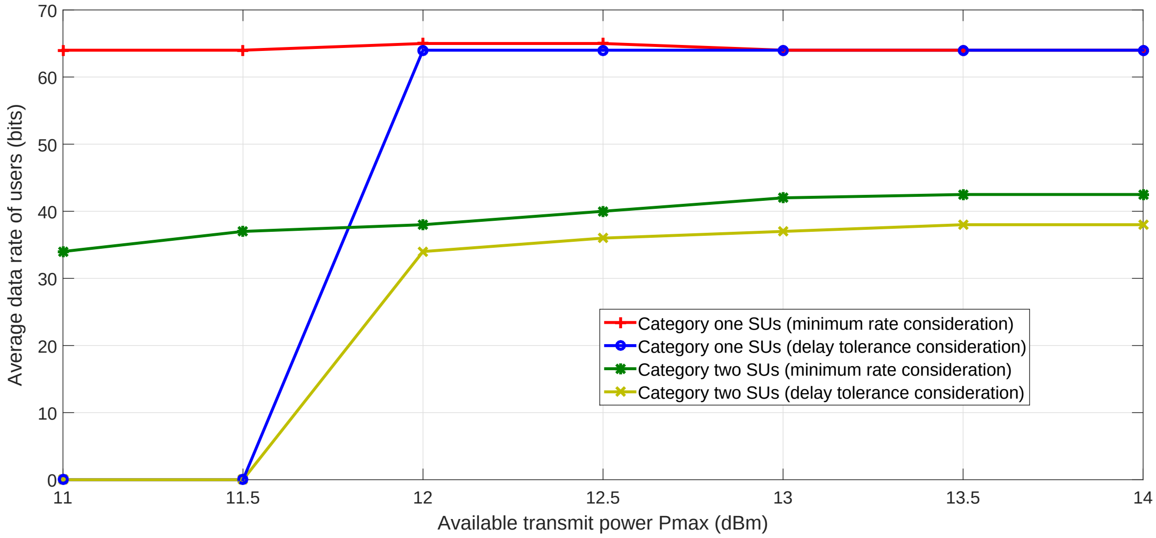

Figure 3.13. Average data rate against maximum transmit power at the SUBS for the delay tolerant consideration.

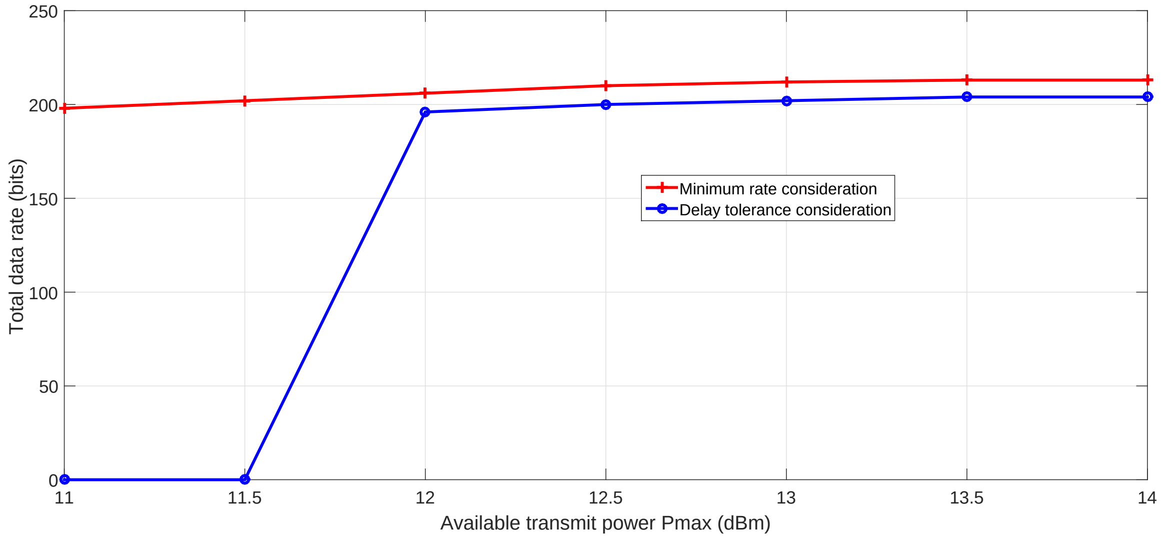

Figure 3.14. Total data rate against maximum transmit power at the SUBS for the delay tolerant consideration.

with those obtained using the minimum data rate classification. The results and reasoning for the observations are similar to those given in Fig. 3.13.

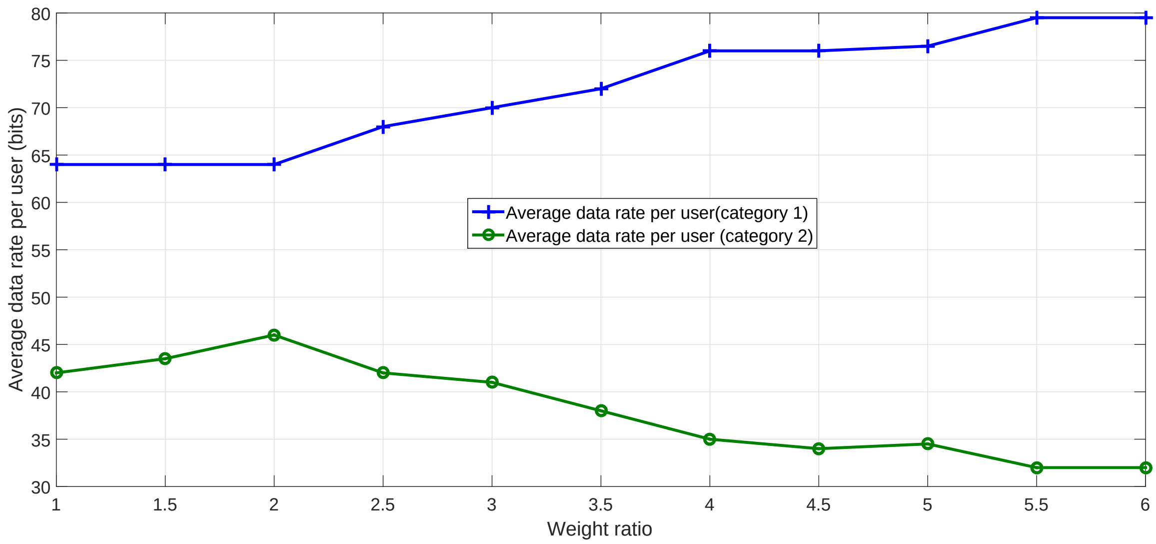

Figure 3.15. Average data rate at different weight ratios for the two categories of users.

3.6.4 Effects of weight on resource allocation in heterogeneous cognitive radio networks

Weight is an important factor in the allocation of resources to various user categories in heterogeneous CRN. This is because weight can be used effectively in a number of ways to influence the decision of the allocation algorithm to favour some user categories over other categories. Weight can therefore be used as a powerful bias mechanism in the RA decision making for CRN to provide options for further improvement that would not have been available should the user categories not have been given such weight considerations.

In Fig. 3.15, the average data rate is plotted against the weight ratio to demonstrate the importance of weight on the data rate achieved by the different categories of users. The minimum data rate classification has been employed (results can thus be compared with the ones in Fig. 3.7 and Fig. 3.8), while the weight ratio between the two user categories has been steadily increased from unity to some higher values. It can be observed that, for larger values of weight ratio, the average data rate for category one users increases while the average data rate for category two users decreases. This implies therefore that, contrary to the results presented in Fig. 3.7, a higher weight in this case compels the algorithm to give a higher data rate (or resources) to users with the higher demand (the category one SUs). Indisputably, users in category one are the most valuable, since they, in some way, pay a higher price in order to get a better service. It therefore becomes meaningful to give them preference when the available resources are slightly improved and this is achieved by the impact of the weight.

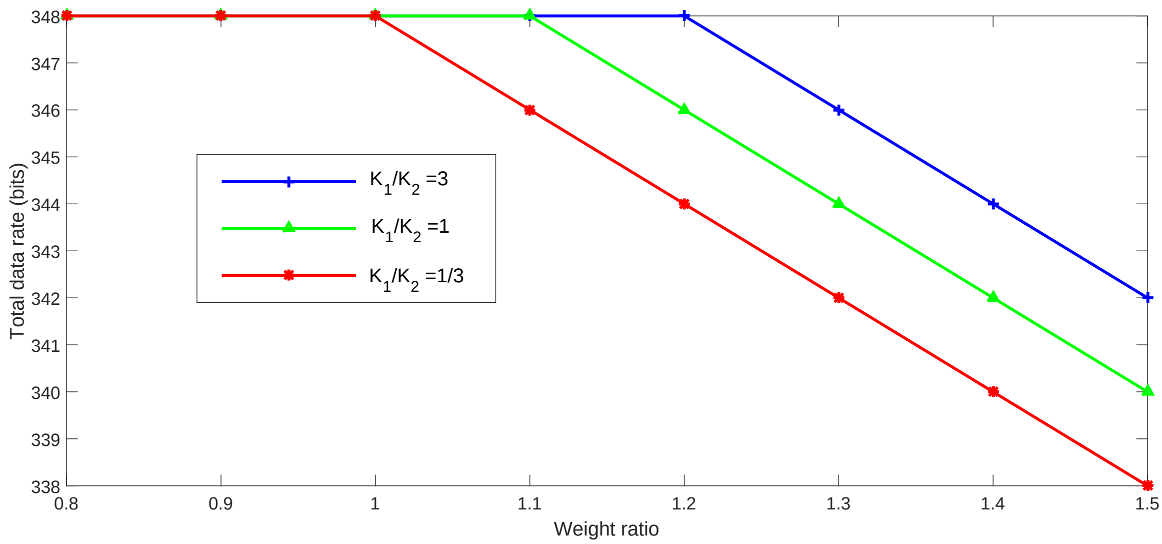

Figure 3.16. Total data rate of the CRN against the weight ratios for different possible combinations of categories of SUs. Maximum transmit power at SUBS is set at $100dBm$, interference to PUs is at $120dBm$.

The minimum data rate requirement for each category of users is, however, still satisfied in all cases, otherwise the problem becomes infeasible.

In Fig. 3.16, the effect of weight on different possible combinations of the number of SUs in the various categories of users is demonstrated. In the plot it can be observed that, at a higher weight, the network achieves better results in terms of the total data rate when the number of category one users is more than the number of category two users available in the network. This is because, although the category one users have a higher demand, the weight still influences the network to satisfy them more. With the weight, the network achieves or gains a lot more by satisfying a higher number of category one users than category two users. The reason is the same as discussed for the previous plot - the category one users are usually the ones that pay more! This is, in fact, a kind of cost-benefit realisation. In essence therefore, the weight is a potent tool for influencing how much of the excess resources are allocated to one category of users over another category in order to achieve the best utility for the network.

As a final consideration, Fig. 3.17 gives a comparison of the performance of different weight distributions. The authors in [115] used weights randomly chosen between 0 and 1 and normalised so that the sum of all user weights equalled 1. In this plot, as a significant improvement, three different weight

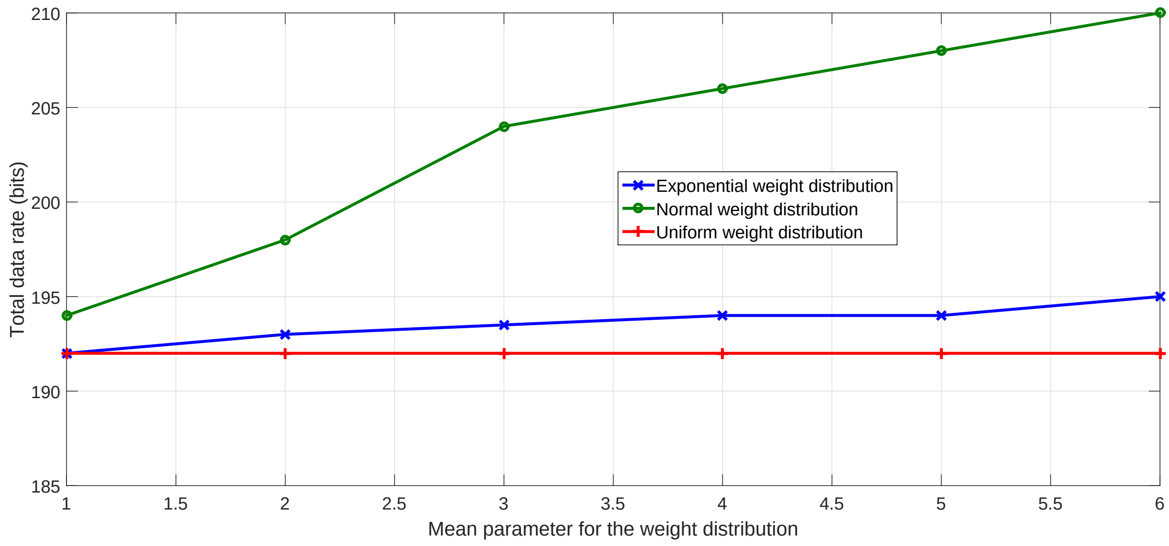

Figure 3.17. Total data rate performance for different weight distributions.

distributions - uniform, normal and exponential distributions - are compared for the SUs. From the result, it can be observed that the normal weight distribution outperforms the exponential and uniform distributions, with the uniform distribution performing the least. This would imply therefore that the performance of CRN with heterogeneous users could be slightly influenced by the choice of the weight distributions employed for the network.

3.7 CONCLUSION

CRN, being an emerging next-generation wireless communication paradigm, must be capable of delivering optimal productivity with the limited resources at its disposal to a wide variety of user categories. Heterogeneous CRN, which incorporates various concepts of heterogeneity as applicable to CRN, is therefore the more realistic CRN consideration. In this chapter, appropriate RA models that capture the various heterogeneous considerations for CRN are developed and analysed. The models are such that heterogeneous SUs in each classification are adequately serviced within the limits of the network's available resources. The optimisation problems developed from the RA formulations are all NP-hard and obtaining optimal solutions to such problems are, in all reality, very difficult to achieve. In the chapter, however, an extensive investigation into how to solve such RA problems is conducted. In the developed solution models, by carefully studying the problems' structure, easier-to-solve ILP reformulations of the original problems are realised. The BnB approach for solving ILP problems is then used to determine optimal solutions for all the classifications of heterogeneity considered. The optimal results of the average data rate, throughput, outage probability, the impact of the number of available users in each category, and the effect of weight on the overall performance of the network that were obtained were extensively discussed.

CHAPTER 4 RESOURCE ALLOCATION IN HETEROGENEOUS COOPERATIVE COGNITIVE RADIO NETWORKS

4.1 CHAPTER OVERVIEW

In CRN, resources such as spectrum, time-slot and bandwidth available for use are usually very limited. This is generally because of the very tight constraints by which CRN operate. Of all the constraints, arguably the most critical of them all is the level of permissible interference to the original owners (or PUs) of the spectrum due to the activities of the SUs. Attempts to mitigate the limiting effects of this constraint, thus achieving higher productivity for CRN, is a current research focus and in this chapter, cooperative diversity is investigated as a promising solution. Cooperative diversity itself has recently attracted attention because of its capability to achieve diversity gain and a much better channel quality for wireless networks. In the chapter, therefore, the possibility of and mechanism for achieving greater utility in CRN when cooperative diversity is incorporated are studied. To accomplish this, a RA model is developed and analysed for an underlay, heterogeneous, cooperative CRN. In the model, during cooperation, a best relay is selected to assist only the SUs that have poor channel conditions (because of their high interference gain to PUs). Overall, the cooperation makes it feasible for virtually all the SUs to transmit at a high data rate while still causing minimal harm to the PUs. This would have been unachievable were they to transmit only directly. The results presented show a marked improvement in the RA performance of CRN when user cooperation is employed in contrast to when CRN operate only by direct communication.

4.2 BACKGROUND

Fundamentals on CRN have already been established in previous chapters. From the explanations provided in those chapters, one may summarise that in CRN, unlicensed cognitive users or SUs are generally made to access and utilise the same spectrum space that has been preallocated to some licensed, original owners or PUs of the spectrum, provided certain preconditions (such as the amount of permissible interference) already agreed upon are not violated by the SUs [61, 116, 117]. While CRN is certainly a very promising wireless communication paradigm, especially because of its promise of addressing the spectrum challenge, several issues with its design, implementation, application and eventual productivity have equally arisen. One such issue, very germane, is the possibility of very low productivity (in terms of throughput, for instance) that the SUs network can actualise, especially when the conditions of the PUs network are rather stringent. An example of this possible limitation can be observed in the underlay CRN arrangement where the SUs are made to transmit over the entire PUs spectrum, but at such low power so that the PUs are not in any way adversely affected by the SUs' transmission [118]. In such a case, it becomes extremely difficult for CRN to achieve great results, if the PUs network has high sensitivity and/or low interference tolerance temperatures. In such situations where the permissible interference to PUs is very low or where there are other very strict conditions under which the SUs must operate, the throughput or system capacity of CRN can become very limited, diminishing the overall network productivity [119]. It might therefore be arguable whether CRN is, in fact, a worthwhile investment, unless such issues are adequately addressed. Several research activities on how to make CRN achieve its ends, even within such tight constraints, are currently being undertaken.

In addressing some of these limitations, it was shown in the previous chapter that it is usually best to allocate low data rates (or none at all) to subchannels where the interference gains to PUs are quite high, so as to achieve optimal or near-optimal productivity in the allocation of the rather scarce and/or limited resources of CRN. This same point can also be observed in some of the earlier published works of the author, for example, references [48, 120]. The point raised above is quite reasonable, as allocating high data rates to such subchannels would imply high transmit power by the SUs and consequently, high interference to the PUs because of the high interference gain. This smart move by the allocating algorithms of the SUs greatly increases the productivity of CRN. However, the productivity achieved is usually still very limited, as there are some subchannels which, because of their high interference channel gains to PUs, are either completely unallocated or are only allocated to transmit very low data rates. If the productivity of CRN is to be improved further, it is imperative to investigate how the use of better channelling techniques can be employed so as to make higher data rates possible for virtually all the available subchannels, but still without causing too much interference to the PUs network.

To address the problem raised above, in this chapter, cooperative diversity is investigated as a promising solution. Essentially, by bringing cooperative diversity into CRN, the chapter reveals how the limiting effects of the interference constraint in its RA problems can be adequately mitigated, thus achieving much better productivity for the network.

4.3 RELATED LITERATURE ON COOPERATIVE DIVERSITY IN COGNITIVE RADIO NETWORKS

Cooperative diversity is a recent but comprehensive proposition on how to achieve better wireless channel networking. Cooperative diversity defines and describes how diversity gains can be realised among spatially dispersed users in a wireless communication system [121, 122]. This is actualised by the cooperating users (called nodes or relays) forming a kind of ‘virtual multiple input, multiple output (virtual-MIMO)’ arrangement. These cooperating users use their antennas, as carried out in conventional MIMO systems, to assist each other in transmitting (or retransmitting) or relaying their data to a given destination user. Overall, a significant increase in reliability and capability of the system is realised. Several cooperative diversity strategies have been developed and studied, some of which are store-and-forward, amplify-and-forward, decode-and-forward and coded cooperation [123]. Similarly, cooperative diversity has been classified in terms of the number of cooperators selected, or on whether or not the cooperation happens opportunistically or incrementally [124]. The important thing is that, at the destination, a much better signal quality is achieved and network capacity is significantly improved. While cooperative diversity has been mainly employed in CRN for spectrum sensing, this chapter investigates and develops how it can be used in improving the effective capacity (that is, the RA optimisation) of CRN by addressing the limitations due to the stringent conditions on the level of permissible interference to PUs.

RA in CRN has been described as devising mechanisms for assigning resources (frequency spectrum, transmit power, bandwidth, time slot, modulation scheme, etc.) fairly and optimally to all users so that the highest possible productivity level is achieved. A number of RA problems for underlay CRN have been developed, and attempts at solving them (both optimally and sub-optimally) have been investigated. In [78] for example, an approach for obtaining optimal solutions for RA problems in an underlay CRN is developed. A centralised algorithm, which makes use of the Lagrangian duality, is first employed to solve the problem. Thereafter, a distributed algorithm that uses dual decomposition is developed to solve the same problem. Both algorithms produce near optimal solutions. Other similar works that have developed RA models for underlay CRN can be found in [66, 84, 104]. The major challenge with underlay, as observed in the above-mentioned works and other similar ones in literature, is the low level of utility that is achievable in its network owing to the stringent conditions of the PUs and the power limitations of the secondary network.

RA problems in overlay CRN have been studied in [27, 60, 62, 87]. In overlay CRN generally, the SUs use free and/or available spectrum (spectrum holes) of the PUs for transmission. Both subchannel and power control were jointly considered in [27] with the intent of maximising the throughput of their CRN. Developing on this, authors in [60] and [62] extended the work to make room for possible imperfections in the CRN's sensing capabilities of PUs and developed models with some robustness to accommodate such imperfections. In [87], the problem is studied even further to include QoS provisioning. A major problem with overlay networks, as observable in the above-mentioned works and similar ones in the literature, is the possibility of PUs' interference and possible disruption in service delivery of the SUs in CRN.

As a means of addressing some of the limitations of the underlay and overlay architectures, recent attempts at introducing user cooperation into RA in CRN have been made. References [57, 59, 61, 79, 125] have all developed models that describe possible cooperation between SUs in CRN to help achieve a higher utility level. In [59] and [61], relays using decode-and-forward protocol were made to assist the SUs in CRN. For the resulting optimisation problem to be solvable, the subchannels were first assigned to the SUs based on their channel gains and possible interference to PUs. Thereafter, power was allocated to each subchannel. A similar model was developed in [57], where a decode-and-forward cooperative relay network was used to assist the SUs, thereby improving throughput. The non-convex optimisation problem that was developed was solved by first dualising, then decomposing into relay assignment and power allocation. A primary decomposition method was also used in [79], after the power allocation problem in the developed model had been formulated. The sum rate of both PUs and SUs was jointly maximised in [125], while the SUs cooperated to transmit their signals. To achieve a result close to optimal, subchannels were first allocated to the SUs; thereafter, power was assigned to each SU and PU iteratively. While the above-reviewed works have incorporated some kind of cooperation, the work developed in this chapter differs from them in that the cooperative diversity approach developed is targeted directly and specifically at addressing the problem of PUs' interference limitations. Thus, the interference problem is first taken care of by the cooperative diversity model even before the RA to SUs is carried out.

More specifically, in this chapter, through SUs cooperation, the impact of the interference to PUs is mitigated, thereby achieving greater productivity for the heterogeneous CRN. The heterogeneity in the CRN has been approached from two perspectives. Firstly, the channels are assumed to be heterogeneous, meaning that the available channels for the CRN do not all have the same characteristics. To capture the differing effects of channel heterogeneity, the network has been developed using an OFDMA platform. With the OFDMA, the system can dynamically and optimally utilise different portions of the spectrum (heterogeneous channels) for different users at the same time. Secondly, the SUs in the network are assumed to be heterogeneous. This implies that the users do have different priorities, requirements or demands, thus necessitating that they be categorised. Users in each category are then serviced based on their priority and/or their varying demands. During cooperation, the selection scheme employed is the single-best-relay selection scheme used alongside the store-and-forward cooperative diversity technique. With this scheme, a best relay among the SUs is selected as the cooperator, which, at cooperation receives data from the source user and transmits to the destination. Overall, the heterogeneous cooperative CRN model, as developed and studied, reveals that much greater productivity is achievable in CRN when its users cooperate. The most-important contributions of this chapter are highlighted:

- Investigating the use of cooperative diversity as a means of mitigating the limiting effects of interference to PUs in the RA problem of CRN, thereby making much better productivity possible for the network.

- Developing and analysing methods for obtaining solutions (optimal and suboptimal) to the RA problem in heterogeneous, cooperative CRN. The solutions are obtained through a thorough study of the structure of the problem.

The remainder of this chapter is organised as follows: Section 4.4 describes the system model, Section

4.5 deals with the problem formulation and reformulation to obtain optimal solutions, Section 4.6 presents the heuristic developed to reduce the computational complexity, Section 4.7 presents and discusses the results and finally, Section 4.8 gives the concluding remarks.

4.4 SYSTEM MODEL

The underlay, heterogeneous, cooperative CRN model developed consists of K heterogeneous SUs and L PUs, all located within the coverage range of the SUBS. N OFDMA heterogeneous subchannels are available for the SUs, to which any of them can be assigned. The K heterogeneous SUs have different demands and priorities. These SUs are thus categorised as K_1: SUs with minimum rate guarantee, and $(K-K_1)$: SUs with best effort service. Users in category one have a minimum rate requirement and their demands are first met. They therefore have the higher priority. Users in category two are best effort users, hence the remaining resources are proportionally shared among them (using a proportional rate constraint). All subchannels are also modelled to be in slow fading. During transmission, the network decides whether to employ direct or cooperative communication based on the prevalent condition of the network. At cooperation, the direct link is ignored in the model because of the high interference to PUs that will be introduced if the direct link is employed, which would potentially limit the entire CRN resourcefulness.

Fig. 4.1. shows the network when cooperation is to be employed. The cooperative scheme employed in this chapter is the incremental, single-best-relay selection cooperative diversity scheme. The scheme being ‘incremental’ means that cooperation is strictly restricted to only when it is needed. And the single-best-relay selective cooperation means that only one best relay is selected in such instances. The reason for this cooperative diversity choice is to ensure that the model is feasible, as well as to minimise overhead. The SU that requires cooperation, as it intends to communicate with a destination terminal (D), is referred to as the source secondary user (SSU). This SSU has a potentially high interference channel gain to the PU on the direct link and would therefore either have not been allocated subchannels at all or would have been given only a few subchannels to transmit at low data rates if direct communication alone had been considered. To help mitigate that limitation, the SSU, at cooperation, selects a cooperating secondary user (CSU) with good channel quality (SSU to CSU as well as CSU to D) and poor interference channel gain to the PU. Through this cooperation, the effects of the poor channel condition are mitigated. Next, a description of how the best relay (i.e. the CSU) is

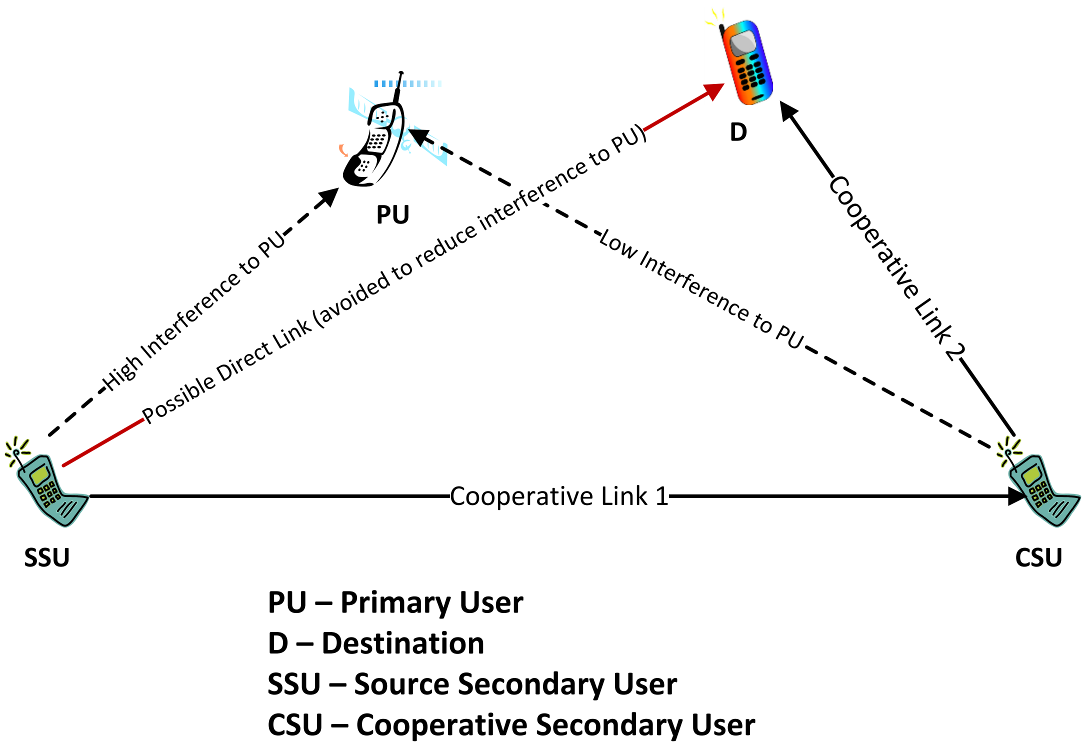

Figure 4.1. System model of the cooperative CRN

selected from among the other users is presented.

In the model developed, the CRN operates a centralised control system with the SUBS as its communicating hub. Communication between the SU and the SUBS is assumed to be error free. All the SUs estimate and communicate their channel conditions as well as their interference gains to PUs to the SUBS. Any of the SUs can be the potential CSU for any other one. The SUBS determines which of the SUs need a cooperator and chooses and contacts the best of the other SUs, which is then assigned as its CSU. It is assumed that at the moment of cooperation, the SU employed as the cooperating SU is idle and has no data of its own to transmit. This information, along with the estimated channel condition and PU interference, is relayed on the control channel to the SUBS by each SU. The choice of a CSU is usually based on the SU with the best channel condition and the least interference gain to the PU. The chosen CSU is thereafter relayed by the SUBS to the SSU, while the other SUs, which have not been contacted, carry on with their normal transmission (or simply maintain their idle state, as the case may be). The SSU transmits to the CSU, which then forwards the transmitted data to the destination terminal D over the assigned subchannels. The transmission is in two time slots. In the first time slot the SSU transmits to the CSU and in the second time slot the CSU transmits to D. The combined channel condition of the SSU and the CSU is obtained as follows:

Denote $H^s_{k,n}$ as the channel gain between the SSU and the kth SU, employed as the CSU, at the nth subchannel and $H^r_{k,n}$ as the channel gain between the CSU and the destination terminal D over the nth subchannel. The SSU transmits signals to the kth relay on the nth subchannel with power $P^s_{k,n}$ in the first slot, while the kth relay (CSU) transmits signals to D on the nth subchannel with power $P^r_{k,n}$ in the second slot. Thus, the data rate for each transmission slot is given in [61] as:

$$\begin{aligned} c^s_{k,n} &= \log_2\left(1 + \frac{P^s_{k,n}|H^s_{k,n}|^2}{\sigma_r^2 + \sum_{l=1}^{L} J^l_{k,n}}\right), \\ c^r_{k,n} &= \log_2\left(1 + \frac{P^r_{k,n}|H^r_{k,n}|^2}{\sigma^2 + \sum_{l=1}^{L} J^l_{n}}\right) \end{aligned} \tag{4.1}$$

where σ_r^2 and σ^2 are the variance values of the noise at the kth relay (CSU) and D respectively. Similarly, the interference to the kth relay and that to D on the nth subchannel by the lth PU's signal are denoted by $J^l_{k,n}$ and J^l_n. This is regarded as noise and measured by the receivers of the CSU and D. It is important to state that the effective data rate during cooperative transmission is actually limited by the minimum of the two hops:

$$c_{k,n,C} = \min\left(c^s_{k,n}, c^r_{k,n}\right). \tag{4.2}$$

If no cooperation is needed, transmission is directly from the SU to D over the assigned subchannels and the data rate is simply $c_{k,n,D}$. This data rate c for each subchannel, using either direct or cooperative transmission, is dependent on the modulation scheme to which the subchannel has been assigned. In this chapter, four modulation schemes, which are BPSK, 4-QAM, 16-QAM and 64-QAM, are considered. The modulation schemes transmit $c = 1,2,4$ and 6 bits per OFDMA symbol respectively. For a given BER (ρ) value to be met at the receiver end of communication, the minimum amount of power required over any given subchannel is dependent on the modulation scheme employed. The minimum power for BPSK modulation is given as $P(c,\rho) = N_\phi[c \times erfc^{-1}(2\rho)]^2$ (where $c = 1$), while for the M-ary QAM, the minimum power is given as $P(c,\rho) = \frac{2(2^c-1)N_\phi}{3}[erfc^{-1}(\frac{c\rho\sqrt{2^c}}{2(\sqrt{2^c}-1)})]^2$ ($c = 2,4$ or 6 for 4-QAM, 16-QAM and 64-QAM respectively) where $erfc(x) = (\frac{1}{\sqrt{2\pi}})\int_x^\infty e^{\frac{-t^2}{2}}dt$ is the complementary error function, $\pi = (22/7)$, and N_ϕ is the single-sided noise power spectral density. In this chapter, N_ϕ assumes the same value for all subchannels. For a given value of ρ, as the number of bits assigned to a subchannel increases, the transmit power also increases, albeit non-linearly. The minimum power $P_{k,n}(c_{k,n},\rho)$ required at the kth SU over the nth subchannel to transmit $c_{k,n}$ bits is obtained by dividing the power $P(c_{k,n},\rho)$ of that user k on the nth subchannel by the channel gain $H^c_{k,n}$ between the SUBS and the user k over that subchannel n. This is thus given as:

$$P_{k,n}(c_{k,n},\rho) = \frac{P(c_{k,n},\rho)}{H^c_{k,n}}. \tag{4.3}$$

4.5 RESOURCE ALLOCATION PROBLEM FORMULATION

Let the minimum data rate assigned to user k in category one be R_k and the normalised proportional fairness factor for each SU in category two be γ_k with data rate R_i indicating the rate for the element i. The total power on the nth subchannel is represented as $\Phi_n = \Sigma_{k=1}^K P_{k,n}$ with $P_{k,n}$ being the transmit power of user k over the nth subchannel ($P_{k,n,C}$ is the power utilised when cooperation is employed, $P_{k,n,D}$ is the transmit power for direct transmission). Also let the interference power gain matrix between the SUBS and the available PU be represented as $\boldsymbol{H}^p \in R^{L\times N}$. The vector $\boldsymbol{H}^p_{l,n}$ therefore denotes the subchannel interference power gain between the SUBS and PU l over subchannel n ($\boldsymbol{H}^p_{l,n,D}$ is the gain matrix when direct transmission is used, $\boldsymbol{H}^p_{l,n,C}$ is the gain matrix when cooperation is employed). The maximum permissible level of interference to the lth PU from all the transmitting SUs is represented as ε_l while $P_{\max}$ denotes the maximum power available for transmission at the SUBS. Also, let $X_{k,n,D}$ be a binary $(0,1)$ variable employed to limit each subchannel to either direct communication or cooperative communication (since each subchannel can only transmit using either of the two, but not both). The RA problem for the heterogeneous cooperative CRN is therefore formulated as:

$$z = \max \sum_{n=1}^{N} \left(\sum_{k=1}^{K_1} [X_{k,n,D} c_{k,n,D} + (1 - X_{k,n,D}) c_{k,n,C}] + \sum_{k=K_1+1}^{K} [X_{k,n,D} c_{k,n,D} + (1 - X_{k,n,D}) c_{k,n,C}] \right); \quad c_{k,n,D}, c_{k,n,C} \in \{0,1,2,4,6\} \tag{4.4}$$

subject to

$$\sum_{n=1}^{N} (c_{k,n,D} + c_{k,n,C}) \geq R_k;\ k = 1,2,\cdots,K_1 \tag{4.5}$$

$$\frac{R_k}{\sum\limits_{i=K_1+1}^{K} R_i} = \gamma_k;\ k = K_1+1, K_1+2, \cdots, K \tag{4.6}$$

$$\sum_{n=1}^{N} \left(\sum_{k=1}^{K} [X_{k,n,D} P_{k,n,D} + (1 - X_{k,n,D}) P_{k,n,C}] \right) \leq P_{\max} \tag{4.7}$$

$$\sum_{n=1}^{N} \Phi_n \boldsymbol{H}^{p}_{l,n,D} \leq \varepsilon_l;\ l = 1,2,\cdots,L \tag{4.8}$$

$$\sum_{n=1}^{N} \Phi_n \boldsymbol{H}^{p}_{l,n,C} \leq \varepsilon_l;\ l = 1,2,\cdots,L \tag{4.9}$$

$$c_{k,n,D} = 0\ if\ c_{k',n,D} \neq 0,\ c_{k,n,C} = 0\ if\ c_{k',n,C} \neq 0, \quad \forall k' \neq k;\ k = 1,2,\cdots,K \tag{4.10}$$

$$X_{k,n,D} \in \{0,1\},\ X_{k,n,D} = 1\ if\ c_{k,n,D} \neq 0, \quad X_{k,n,D} = 0 \text{ otherwise.} \tag{4.11}$$

The objective function (4.4) captures the throughput or total data rate that all the SUs in the network for both direct and cooperative transmission can realise. Equation (4.5) is the minimum data rate constraint, indicating that for each SU in category one, their minimum data rate requirement must be met. Equation (4.6) is the best effort service constraint where, to determine how the remaining resources are to be assigned to each user in category two, a proportional fairness factor is being employed. Equation (4.7) is used to limit the total transmit power of all the SUs during direct and cooperative transmission to the SUBS's maximum transmit power available. Equation (4.8) shows that when the SUs are transmitting, the amount of interference permissible to each PU during direct transmission must not be greater than the predetermined threshold limit. Equation (4.9) is similar to (4.8), but captures the interference constraint during cooperative transmission. To restrict the allocation of subchannels to only one user per subchannel for each user, the mutually exclusive constraint in equation (4.10) is given. The constraint shows that once subchannel n has been assigned to a user

$k' \neq k$, the data rate for subchannel n must therefore be 0 for any other user k. The equation in constraint (4.6) can be equally expressed as:

$$R_k = \gamma_k \times \sum_{i=K_1+1}^{K} R_i$$

where $\sum_{i=K_1+1}^{K} R_i$ is the sum value of all the data rates for all SUs in category two. Representing $\gamma_k \times \sum_{i=K_1+1}^{K} R_i$ by $\tilde{\gamma}_k$, Equation (4.6) becomes:

$$R_{K_1+1} : R_{K_1+2} : \ldots : R_K = \tilde{\gamma}_{K_1+1} : \tilde{\gamma}_{K_1+2} : \ldots : \tilde{\gamma}_K. \tag{4.12}$$

The above formulation of the RA problem in equations (4.4 - 4.11) is not a linear programming problem because the power constraint in equation (4.7) is not linear. However, by a careful consideration of the problem structure, the problem has been reformulated as an ILP problem. The reformulated problem can easily be solved by using any of the classical optimisation techniques. The BnB approach is used in this chapter. The reformulation, which develops on the approach employed in the previous chapter, is achieved in the next section.

4.6 REFORMULATION AS AN INTEGER LINEAR PROGRAMMING PROBLEM

The reformulation of the original problem to an ILP problem is carried out as follows:

Define $\boldsymbol{x}_I$ as the bit allocation vector for all subchannels assigned to all users in category one (both for direct and cooperative transmission, hence, $\boldsymbol{x}_I = (\boldsymbol{x}_{I,D} + \boldsymbol{x}_{I,C})$) and also define $\boldsymbol{x}_{II}$ as the bit allocation vector for all subchannels assigned to all users in category two (for direct and cooperative transmission so that $\boldsymbol{x}_{II} = (\boldsymbol{x}_{II,D} + \boldsymbol{x}_{II,C})$). $\boldsymbol{x}_I$ and $\boldsymbol{x}_{II}$ are given as:

$$\boldsymbol{x}_I = [(\boldsymbol{x}_{I,N}^1)^T \ (\boldsymbol{x}_{I,N}^2)^T \ \cdots \ (\boldsymbol{x}_{I,N}^N)^T]^T \ \in \{0,1\}^{NK_1C\times 1} \tag{4.13}$$

$$\boldsymbol{x}_{II} = [(\boldsymbol{x}_{II,N}^1)^T \ (\boldsymbol{x}_{II,N}^2)^T \ \cdots \ (\boldsymbol{x}_{II,N}^N)^T]^T \ \in \{0,1\}^{N(K-K_1)C\times 1} \tag{4.14}$$

where $\boldsymbol{x}^n_{I,N} = [x^T_{I,1,n}\, x^T_{I,2,n}\, \cdots\, x^T_{I,K_1,n}]^T \in \{0,1\}^{K_1 C\times 1}$ indicates that subchannel n has been assigned to a category one SU with $\boldsymbol{x}_{I,k,n} = [x_{k,n,1}\, x_{k,n,2}\, \cdots\, x_{k,n,M}\,]^T \in \{0,1\}^{C\times 1}$; $n = 1,\cdots,N$; $k = 1,\cdots,K_1$; M indicates the overall number of modulation schemes being employed (for this work, $M = 4$). The implication is that $\boldsymbol{x}_{I,k,n} = [x_{k,n,1}\, x_{k,n,2}\, x_{k,n,3}\, x_{k,n,4}\,]^T$. Similar explanations apply to $\boldsymbol{x}_{II}$. The combined bit allocation vector $\boldsymbol{x} = \boldsymbol{x}_I + \boldsymbol{x}_{II}$. As a result of the mutually exclusive constraint, $\boldsymbol{x}^n_{I,N}$ and $\boldsymbol{x}^n_{II,N}$ can be any of the vectors $\{[0\,0\cdots 0]^T, [1\,0\cdots 0]^T, [0\,1\cdots 0]^T, \cdots, [0\,0\cdots 1]^T\}$. Hence, only one component in $\boldsymbol{x}^n_{I,N}$ is 1, while the other components are all 0s (same applies for $\boldsymbol{x}^n_{II,N}$). If $x_{k,n,c}$ is 1, it means that subchannel n has been assigned to user k to transmit c bits per symbol. If $\boldsymbol{x}^n_{I,N}$ (or $\boldsymbol{x}^n_{II,N}$) has all its components as 0s, subchannel n is not being assigned to any user.

For the two user categories, define the modulation order vectors $\boldsymbol{b}_I$ and $\boldsymbol{b}_{II}$ as:

$$\boldsymbol{b}_I = [(\boldsymbol{b}^1_{I,N})^T\, (\boldsymbol{b}^2_{I,N})^T\, \cdots\, (\boldsymbol{b}^N_{I,N})^T]^T \in \mathbb{Z}^{NK_1C\times 1} \tag{4.15}$$

$$\boldsymbol{b}_{II} = [(\boldsymbol{b}^1_{II,N})^T\, (\boldsymbol{b}^2_{II,N})^T\, \cdots\, (\boldsymbol{b}^N_{II,N})^T]^T \in \mathbb{Z}^{N(K-K_1)C\times 1} \tag{4.16}$$

where $\boldsymbol{b}^n_{I,N} = [b^T_{I,1,n}\, b^T_{I,2,n}\, \cdots\, b^T_{I,K_1,n}]^T \in \mathbb{Z}^{K_1C\times 1}$ and $\boldsymbol{b}_{I,k,n} = [b_{k,n,1}\, b_{k,n,2}\, \cdots\, b_{k,n,C}\,]^T \in \mathbb{Z}^{C\times 1}$. Similar explanations also apply to $\boldsymbol{b}_{II}$. Having considered only four modulation schemes (i.e. BPSK, 4-QAM, 16-QAM and 64-QAM), $\boldsymbol{b}_{1,k,n} = [1\;2\;3\;4]^T$ (the same applies to $\boldsymbol{b}^n_{II,N}$). For the two categories of SUs, data rate matrices $\boldsymbol{B}_i \in \mathbb{Z}^{K_1\times NK_1C}$ and $\boldsymbol{B}_j \in \mathbb{Z}^{(K-K_1)\times N(K-K_1)C}$ are defined respectively as;

$$\boldsymbol{B}_i = \begin{bmatrix} b_1 & b_1 & \cdots & b_1 \\ b_2 & b_2 & \cdots & b_2 \\ \vdots & \vdots & \ddots & \vdots \\ b_{K_1} & b_{K_1} & \cdots & b_{K_1} \end{bmatrix}, \boldsymbol{B}_i \in \mathbb{Z}^{K_1\times NK_1C} \tag{4.17}$$

$$\left\{ \begin{array}{llll} b_1 = [b^T & 0^T_C & \cdots & 0^T_C] \in \mathbb{Z}^{1\times K_1C} \\ b_2 = [0^T_C & b^T & \cdots & 0^T_C] \in \mathbb{Z}^{1\times K_1C} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K_1} = [0^T_C & 0^T_C & \cdots & b^T] \in \mathbb{Z}^{1\times K_1C} \end{array} \right\}$$

$$\boldsymbol{B}_j = \begin{bmatrix} b_{K_1+1} & b_{K_1+1} & \cdots & b_{K_1+1} \\ b_{K_1+2} & b_{K_1+2} & \cdots & b_{K_1+2} \\ \vdots & \vdots & \ddots & \vdots \\ b_K & b_K & \cdots & b_K \end{bmatrix}, \boldsymbol{B}_j \in \mathbb{Z}^{(K-K_1)\times N(K-K_1)C} \tag{4.18}$$

$$\begin{Bmatrix} b_{K_1+1} = [b^T & 0_C^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times(K-K_1)C} \\ b_{K_1+2} = [0_C^T & b^T & \cdots & 0_C^T] \in \mathbb{Z}^{1\times(K-K_1)C} \\ \vdots & \vdots & \ddots & \vdots \\ b_K = [0_C^T & 0_C^T & \cdots & b^T] \in \mathbb{Z}^{1\times(K-K_1)C} \end{Bmatrix}$$

Equation (4.4), which gives the total data rate achievable by the network, can thus be written as $\max_x[(\boldsymbol{b}_I)^T\boldsymbol{x}_I + (\boldsymbol{b}_{II})^T\boldsymbol{x}_{II}]$. Define $\boldsymbol{R}_k \triangleq [R_1\ R_2\ \cdots\ R_{K_1}]^T \in \mathbb{R}^{K_1\times 1}$ and $\tilde{\boldsymbol{\gamma}}_k \triangleq [\tilde{\gamma}_{K_1+1}\ \tilde{\gamma}_{K_1+2}\ \cdots\ \tilde{\gamma}_K]^T \in \mathbb{R}^{(K-K_1)\times 1}$, the constraint of equation (4.5), which explains the data rate per user for category one SUs, can be written as $\boldsymbol{B}_i\boldsymbol{x}_I \geq \boldsymbol{R}_k$, while the data rate constraint for category two SU given in equation (4.6) can be written as $\boldsymbol{B}_j\boldsymbol{x}_{II} = \tilde{\boldsymbol{\gamma}}_k$.

Next, a power transmission vector $\boldsymbol{p}$ is defined as:

$$\boldsymbol{p} = [(\boldsymbol{p}_N^1)^T\ (\boldsymbol{p}_N^2)^T\ \cdots\ (\boldsymbol{p}_N^N)^T]^T\ \in \mathbb{R}^{NKC\times 1} \tag{4.19}$$

where $\boldsymbol{p}_N^n = [\boldsymbol{p}_{1,n}^T\ \boldsymbol{p}_{2,n}^T\ \cdots\ \boldsymbol{p}_{K,n}^T]^T\ \in \mathbb{R}^{KC\times 1}$ and $\boldsymbol{p}_{k,n} = [p_{k,n,1}\ p_{k,n,2}\ \cdots\ p_{k,n,C}]^T\ \in \mathbb{R}^{C\times 1}$; $p_{k,n,c}$ is the power required to transmit c bits over subchannel n for user k. Equation (4.7), which describes the power constraint can now be written as $\boldsymbol{p}^T\boldsymbol{x} \leq P_{\max}$. Given that the transmit power is the summation of the power used for both direct and cooperation transmission, $\boldsymbol{p} = \boldsymbol{p}_D + \boldsymbol{p}_C$, where $\boldsymbol{p}_D$ and $\boldsymbol{p}_C$ are the transmit power vectors during direct and cooperation transmission respectively. The power constraint therefore becomes $(\boldsymbol{p}_D + \boldsymbol{p}_C^T\boldsymbol{x}) \leq P_{\max}$.

To write equation (4.8), the interference power constraint (which is also applicable to equation (4.9)), in terms of the bit allocation vector $\boldsymbol{x}$, define a matrix $\boldsymbol{A} \in \{0,1\}^{N\times NKC}$ as below:

$$\boldsymbol{A} = \begin{bmatrix} 1_{KC}^T & 0_{KC}^T & \cdots & 0_{KC}^T \\ 0_{KC}^T & 1_{KC}^T & \cdots & 0_{KC}^T \\ \vdots & \vdots & \ddots & \vdots \\ 0_{KC}^T & 0_{KC}^T & \cdots & 1_{KC}^T \end{bmatrix}, \boldsymbol{A} \in \{0,1\}^{N \times NKC} \tag{4.20}$$

$$1_{KC} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \{1\}^{KC \times 1}, \qquad 0_{KC} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \{0\}^{KC \times 1}$$

Let $\boldsymbol{p} \odot \boldsymbol{x}$ be the Schur-Hadamard (or entry-wise) product of $\boldsymbol{p}$ and $\boldsymbol{x}$, $\boldsymbol{A}(\boldsymbol{p} \odot \boldsymbol{x})$ will therefore be that $N \times 1$ vector whose nth element gives the total power the nth subchannel uses while transmitting. By defining $\boldsymbol{\varepsilon}_l \triangleq [\varepsilon_1 \ \varepsilon_2 \ ... \ \varepsilon_L]^T \in \mathbb{R}^{L \times 1}$, equation (4.8), which describes the interference power constraint for the direct transmission, can then be written as:

$$[\boldsymbol{H}_{l,n,D}^p(\boldsymbol{A}(\boldsymbol{P}_D \odot \boldsymbol{x}))] \leq \boldsymbol{\varepsilon}_l. \tag{4.21}$$

Likewise, the constraint in equation (4.9), which describes the interference power for the cooperative transmission, can be written as:

$$[\boldsymbol{H}_{l,n,C}^p(\boldsymbol{A}(\boldsymbol{P}_C \odot \boldsymbol{x}))] \leq \boldsymbol{\varepsilon}_l. \tag{4.22}$$

Thus, the RA problem for the modelled heterogeneous cooperative cognitive CRN described in equations (4.4) - (4.11) can be described in the ILP form as given by the following formulation:

$$z^* = \max_x [(\boldsymbol{b}_I)^T \boldsymbol{x}_I + (\boldsymbol{b}_{II})^T \boldsymbol{x}_{II}] \tag{4.23}$$

subject to

$$\boldsymbol{B}_i \boldsymbol{x}_I \geq \boldsymbol{R}_k; \ k = 1, 2, \cdots, K_1 \tag{4.24}$$

$$\boldsymbol{B}_j\boldsymbol{x}_{II} = \tilde{\boldsymbol{\gamma}}_k;\ k = K_1+1, K_1+2, \cdots, K \tag{4.25}$$

$$(\boldsymbol{p}_D + \boldsymbol{p}_C)^T\boldsymbol{x} \leq P_{\max} \tag{4.26}$$

$$[\boldsymbol{H}^p_{l,n,D}(\boldsymbol{A}(\boldsymbol{p}_D \odot \boldsymbol{x}))] \leq \boldsymbol{\varepsilon}_l \tag{4.27}$$

$$[\boldsymbol{H}^p_{l,n,C}(\boldsymbol{A}(\boldsymbol{p}_C \odot \boldsymbol{x}))] \leq \boldsymbol{\varepsilon}_l \tag{4.28}$$

$$\boldsymbol{0}_N \leq \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{1}_N \tag{4.29}$$

$$\boldsymbol{x}_I, \boldsymbol{x}_{II}, \boldsymbol{x} \in \{0,1\}. \tag{4.30}$$

The formulation above is an ILP problem of which, in this chapter, the BnB approach has been employed to obtain solutions. The BnB optimisation approach is a very useful and well-developed technique for solving such problems. However, although the BnB approach can yield optimal solutions, it can be very poor in finding such solutions timeously, especially in large networks. It is imperative to investigate a much faster approach for achieving near-optimal solutions. This is done by developing a heuristic for solving the problem, as explained in the next section.

4.7 ITERATIVE-BASED HEURISTIC

In this section, a fast, iterative-based heuristic is developed to solve the formulated ILP problem. Even though the results obtained are only near-optimal, the heuristic gives a good trade-off between optimality and complexity, especially for large systems. The approach employed in the heuristic is an extension of the work presented in [65]. The algorithm involves two steps:

- subchannel allocation
- iterative bit and power allocation.

4.7.1 Subchannel allocation

In carrying out the subchannel allocation for the different categories of SUs, the constraint $\boldsymbol{x} \in [0,1]$ is integer-relaxed such that the constraint becomes:

$$0 \leq \boldsymbol{x} \leq 1. \tag{4.31}$$

In other words, $\boldsymbol{x}$ is allowed to take any value from 0 to 1 and not necessarily restricted to either 0 or 1. All the other parts of the formulation are unchanged. By solving this integer-relaxed formulation at the first iteration, the values of $\boldsymbol{x}$ are obtained, which implies that the subchannels have been allocated to the various users. The data rate for the kth SU at the nth subchannel becomes $(\boldsymbol{b}_{k,n}^T\boldsymbol{x}_{k,n})$. It is not impossible that there is a user $m \neq k$ whose data rate $(\boldsymbol{b}_{m,n}^T\boldsymbol{x}_{m,n})$ on subchannel n could be larger than user k's data rate on n. It would therefore be more appropriate to give subchannel n to user m rather than k. Hence, subchannel n is only allocated to user k after ascertaining that $(\boldsymbol{b}_{k,n}^T\boldsymbol{x}_{k,n}) \geq (\boldsymbol{b}_{m,n}^T\boldsymbol{x}_{m,n})\forall m \neq k$. Clearly then, each subchannel is allocated to the SU that has the highest achievable data rate over that subchannel. It is important to realise too that once the suchannels have been allocated to the different SUs using the above criterion at the first iteration, the dimension of $\boldsymbol{x}$ reduces from its initial value of $\boldsymbol{x} \in [0,1]^{KNC\times 1}$ to the smaller value of $\boldsymbol{x} \in [0,1]^{NC\times 1}$.

4.7.2 Iterative bit and power allocation

Once the subchannels have been assigned to the various SUs, it then remains to determine how many bits (or by inference, what modulation scheme) and what power can be associated with each subchannel. This is carried out in an iterative manner. The algorithm starts by assigning a possible number of bits (rather unambitiously) to each user, then it determines the power used, checks if other constraints are not violated, determines if there is some excess power left, and increases the bits gradually where possible. Then it checks the power again and the whole iterative process is repeated until no further improvement on bit allocation is possible.

The whole optimisation process occurs in a number of iterations, say y. In general, therefore, the following optimisation problem has to be solved at the yth iteration step:

$$\max_{\boldsymbol{x}^y} [(\boldsymbol{b}_I^y)^T\boldsymbol{x}_I^y + (\boldsymbol{b}_{II}^y)^T\boldsymbol{x}_{II}^y] \tag{4.32}$$

subject to

$$\boldsymbol{B}_i\boldsymbol{x}_I^y \geq [\boldsymbol{R}_k - \boldsymbol{f}^{(y-1)}]^+;\ k = 1,2,\cdots,K_1 \tag{4.33}$$

$$\boldsymbol{B}_j\boldsymbol{x}_{II}^y = [\tilde{\boldsymbol{\gamma}}_k - \boldsymbol{g}^{(y-1)}]^+;\ k = K_1+1, K_1+2,\cdots,K \tag{4.34}$$

$$(\boldsymbol{p}^{(y-1)})^T \boldsymbol{x}^y \leq P_{\max} - \|\boldsymbol{u}^{(y-1)}\|_1 \tag{4.35}$$

$$\boldsymbol{H}^p[\boldsymbol{A}(\boldsymbol{p}^{(y-1)} \odot \boldsymbol{x}^y)] \leq \boldsymbol{\varepsilon}_l - \boldsymbol{H}^p \boldsymbol{u}^{(y-1)} \tag{4.36}$$

$$\boldsymbol{0}_N \leq \boldsymbol{A}\boldsymbol{x}^y \leq \boldsymbol{1}_N \tag{4.37}$$

$$\boldsymbol{0}_{KNC} \leq \boldsymbol{x}^y \leq \boldsymbol{1}_{KNC} \tag{4.38}$$

where $\boldsymbol{f}^{(y-1)}$ and $\boldsymbol{g}^{(y-1)}$ are the allocated bits for category one and category two users at the yth iteration respectively, and $\boldsymbol{u}^{(y-1)}$ is the allocated power at the yth iteration.

Here, a detailed explanation on the iteration process is given. Recall that the bit allocation to the nth subchannel assigned to a category one SU, $\boldsymbol{b}_{I,n} = [b_{1,n}^T \cdots b_{K_1,n}^T]^T$ is a vector of size $K_1C \times 1$ with possible entries $1, 2, 4$ and 6. Assume that during the subchannel allocation carried out in the last subsection, the first subchannel has been allocated to the second user, which happens to be a category one SU. Then, $\boldsymbol{b}_{I,1} = [0\,0\,0\,0, 1\,2\,4\,6, 0\,0\,0\,0, 0\,0\,0\,0]$ for users in category one (assuming there are four users). If it had been the third subchannel that was allocated to the first user, which happens to be a category two SU, then $\boldsymbol{b}_{II,3} = [1\,2\,4\,6, 0\,0\,0\,0, 0\,0\,0\,0, 0\,0\,0\,0]$ (assuming there are also four users in this category) and so on. Once this has been done and certain elements of $\boldsymbol{b}_I$ and $\boldsymbol{b}_{II}$ are zeros according to the subchannel allocation, the vectors $\boldsymbol{b}_I$ and $\boldsymbol{b}_{II}$ are renamed $\boldsymbol{b}_I^1$ and $\boldsymbol{b}_{II}^1$ respectively. Consequently, at the first iteration (i.e. when $y = 1$), the following optimisation problem is solved:

$$\max_{\boldsymbol{x}^1} \, [(\boldsymbol{b}_I^1)^T \boldsymbol{x}_I^1 + (\boldsymbol{b}_{II}^1)^T \boldsymbol{x}_{II}^1] \tag{4.39}$$

subject to

$$\boldsymbol{B}_i \boldsymbol{x}_I^1 \geq \boldsymbol{R}_k; \; k = 1, 2, \cdots, K_1 \tag{4.40}$$

$$\boldsymbol{B}_j \boldsymbol{x}_{II}^1 = \tilde{\boldsymbol{\gamma}}_k; \; k = K_1 + 1, K_1 + 2, \cdots, K \tag{4.41}$$

$$\boldsymbol{p}^T \boldsymbol{x}^1 \leq P_{\max} \tag{4.42}$$

$$[\boldsymbol{H}_{l,n,D}^p \boldsymbol{A}(\boldsymbol{p}_D \odot \boldsymbol{x}^1)] \leq \boldsymbol{\varepsilon}_l \tag{4.43}$$

$$[\boldsymbol{H}_{l,n,C}^p \boldsymbol{A}(\boldsymbol{p}_C \odot \boldsymbol{x}^1)] \leq \boldsymbol{\varepsilon}_l \tag{4.44}$$

$$\boldsymbol{0}_N \leq \boldsymbol{A}\boldsymbol{x}^1 \leq \boldsymbol{1}_N \tag{4.45}$$

$$\boldsymbol{0}_{KNC,1} \leq \boldsymbol{x}^1 \leq \boldsymbol{1}_{KNC,1}. \tag{4.46}$$

$\boldsymbol{f}^{(0)}$, $\boldsymbol{g}^{(0)}$ and $\boldsymbol{u}^{(0)}$ are all going to be 0 at the first iteration, hence they are not reflected in the formulation above.

The rates $\boldsymbol{B}_i x_I^1$ and $\boldsymbol{B}_j x_{II}^1$ and power $\boldsymbol{p}^T \boldsymbol{x}^1$ obtained at the first iteration are passed on as $\boldsymbol{f}^1$, $\boldsymbol{g}^1$ and $\boldsymbol{u}^{(1)}$ respectively for the second iteration. Vector $\boldsymbol{x}^1$ is used along with the power vector $\boldsymbol{p}$ to determine the initial modulation scheme (invariably, the number of bits) for each SU at various subchannels. From the explanation given earlier, the first subchannel, say, has been allocated to the second SU. Hence, all entries of $\boldsymbol{x}_I^1$ are zeros except the elements in $x_{2,1}^1$. The total power allocated to the first subchannel can then be calculated as $(p_{2,1}^T x_{2,1}^1)$. To generalise, if the nth subchannel is allocated to the kth SU, the total power allocated to it is calculated as $(p_{k,n}^T x_{k,n}^1)$. The modulation scheme η (with bits c_η) that can be employed without exceeding the power $p_{k,n}^T x_{k,n}^1$ can be obtained as:

$$\eta = arg \max_{\eta} \left\{ \eta \in [0,1,2,3,4] : \; p_{k,n,\eta} \leq p_{k,n}^T x_{k,n}^1 \right\}. \tag{4.47}$$

The value η answers the question, ‘what is the highest modulation scheme that can be assigned to subchannel n that will require a transmit power not exceeding the power already allocated to this subchannel?’ Since the modulation sizes and their corresponding powers are finite and predetermined, the set of power levels that $p_{k,n,\eta}$ can take will be finite. Once the bits corresponding to this $p_{k,n,\eta}$ are determined, the total power used up to that point will still be less than $P_{\max}$. The interference leakage to PUs will also still be less than ε. As a result, it is most likely that there will be some residual power available for use (thus implying that further iterations can still be carried out to increase the number of bits already allocated to each subchannel). Hence $y = 2$ (i.e., the second iteration) becomes feasible. Since (from the subchannel allocation) the first subchannel has been allocated to the second user, which happens to be in category one, to transmit 2 bits (4-QAM modulation), then, $\boldsymbol{b}_{I,2,1}$ can be modified as $\boldsymbol{b}_{I,2,1}^2 = [0\ 0\ (4-2)\ (6-2)]^T = [0\ 0\ 2\ 4]^T$. To have allocated 2 bits to this subchannel, the power $\boldsymbol{p}_{2,1,2}$ must have been used. With the realisation of excess power available for use, the allocation might then be upgraded to, say, a 16-QAM (to transmit 2 more bits) or 64-QAM (to transmit 4 more bits). For this to take place, it would require an additional power of $(p_{2,1,3} - p_{2,1,2})$ (for 16-QAM) or $(p_{2,1,4} - p_{2,1,2})$ (for 64-QAM) respectively. Hence, the new power vector at the second iteration $\boldsymbol{p}_{2,1}^1 = [p_{2,1,1}\ p_{2,1,2}\ (p_{2,1,3} - p_{2,1,2})\ (p_{2,1,4} - p_{2,1,2})]^T$. The values of the vector $\boldsymbol{p}^1$ are thus determined. If u_n^1 denotes the power that was allocated to the nth subchannel in the first iteration, then $\boldsymbol{u}^1 \triangleq [u_1^1 \cdots u_N^1]^T$. It therefore implies that $P_{\max} - \sum_{n=1}^{N} u_n^1$, which can rather be written as $P_{\max} - \|\boldsymbol{u}^1\|_1$, is the residual power available for the second iteration step. After this second iteration, the amount of power allocated to the nth subchannel is the sum of the power allocated at the first iteration and that allocated at the second iteration. This total power is given as:

$$v_n^2 = u_n^1 + (\boldsymbol{p}_{k,n}^1)^T \boldsymbol{x}_{k,n}^2.$$

This new power is used to decide what the modulation scheme η of the nth subchannel should be upgraded to.

$$\eta = arg \max_{\eta} \left\{ \eta \in [0,1,2,3,4] : \, p_{k,n,\eta} \leq v_n^2 \right\}. \tag{4.48}$$

Similarly, the interference to PUs as a result of the power allocated in the first iteration step is given as $\boldsymbol{H}^p\boldsymbol{u}^1$. The remaining interference permissible must be less than $(\boldsymbol{\varepsilon}_l - \boldsymbol{H}^p\boldsymbol{u}^1)$ for the second iteration. Since, at this second iteration, f_k^1 already becomes the data rate allocated to the kth SU in category one during the first iteration and g_k^1 becomes the data rate already allocated to the kth SU in category two during the first iteration, $\boldsymbol{f}^1$ and $\boldsymbol{g}^1$ are defined as $\boldsymbol{f}^1 \triangleq [f_1^1 \, \cdots \, f_1^k]^T$ and $\boldsymbol{g}^1 \triangleq [g_1^1 \, \cdots \, g_1^k]^T$ respectively. Hence, the data rate requirement at the second iteration for category one users would be $(\boldsymbol{R}_k - \boldsymbol{f}^1)$, while the available data rate for category two users at the second iteration would be $(\tilde{\boldsymbol{\gamma}}_k - \boldsymbol{g}^1)$. The constraints on data rate then become $\boldsymbol{B}_i\boldsymbol{x}_I^2 \geq [\boldsymbol{R}_k - \boldsymbol{f}^1]^+$ for category one users and $\boldsymbol{B}_j\boldsymbol{x}_{II}^2 = [\tilde{\boldsymbol{\gamma}}_k - \boldsymbol{g}^1]^+$ for category two users.

This whole iteration process is repeated continuously and only stopped when no further improvement can be achieved on the total achievable data rate for each user (in other words, the throughput of the system cannot be improved any further). The stopping criterion is thus given as:

$$[(\boldsymbol{b}_I^y)^T \boldsymbol{x}_I^y + (\boldsymbol{b}_{II}^y)^T \boldsymbol{x}_{II}^y] - [(\boldsymbol{b}_I^{y-1})^T \boldsymbol{x}_I^{y-1} + (\boldsymbol{b}_{II}^{y-1})^T \boldsymbol{x}_{II}^{y-1}] = \varsigma, \tag{4.49}$$

where ς is a predetermined (very small) value.

Table 4.1. Pseudo-code for the proposed iterative-based heuristic (part I)

Pseudo-code for the subchannel allocation
solve for $\boldsymbol{x}$ using equations (4.23) - (4.29) and (4.31)
set subchannel index $n = 0$
repeat
$n \leftarrow n+1$
if $(\boldsymbol{b}_{k,n}^T \boldsymbol{x}_{k,n}) \geq (\boldsymbol{b}_{m,n}^T \boldsymbol{x}_{m,n}) \forall m \neq k$
nth subchannel is allocated to user k
end if
until $n < N+1$

After the yth iteration, the vectors $\boldsymbol{f}^{(y+1)}$ and $\boldsymbol{g}^{(y+1)}$ will contain the allocated bits for each subchannel assigned to category one and category two users respectively. The vector $\boldsymbol{u}^{(y+1)}$ will contain the power allocated to each subchannel. The respective pseudo-codes given in Tables 4.1 and 4.2 summarise the subchannel allocation and the iterative bit and power allocation that form the heuristic.

4.8 RESULTS AND DISCUSSION

The RA model for the underlay, heterogeneous, cooperative CRN is simulated in MATLAB while the YALMIP simulator is used in carrying out the optimisation. For the simulation, the parameters used are given as: OFDMA subchannels $N = 64$, PUs $L = 4$, SUs $= 8$ in all, with category one SUs $K_1 = 2$, category two SUs $(K - K_1) = 2$ and SUs which act as possible cooperators from which the best relay (CSU) is selected $= 4$. The minimum data rate requirement for each SU in category one is 64 bits while the remaining resources are proportionately distributed among the category two SUs with a normalised proportional rate constant γ_k summed to unity. The BER requirement for all SUs $\rho = 0.01$. The choice of the number of PUs, SUs and other parameters used in the simulation is informed by the need to compare results obtained with the works already presented in the previous chapter, as well as in comparative literatures such as [66] and [120].

Figs. 4.2 and 4.3 give the average data rate (bits) for each category of SUs against the maximum interference power to the PUs for both direct communication and cooperative communication. Cases

Table 4.2. Pseudo-code for the proposed iterative-based heuristic (part II)

Pseudo-code for the bit and power allocation (i.e. at $y = 1, 2, 3, \ldots$)
set $n = 0, y = 0, \boldsymbol{u}^{(0)} = \boldsymbol{0}_N, \boldsymbol{p}^{(0)} = \boldsymbol{p}$
repeat
$y \leftarrow y + 1$
set $\boldsymbol{f}^y = \boldsymbol{0}_K, \boldsymbol{g}^y = \boldsymbol{0}_K, \boldsymbol{v}^y = \boldsymbol{0}_N$
solve the problem (4.32) - (4.38)
repeat
$N \leftarrow n + 1$
$v_n^y = u_n^{y-1} + (\boldsymbol{p}_{k,n}^{y-1})^T \boldsymbol{x}_{k,n}^y$
if $\eta = arg\max_\eta \{\eta \in [0,1,2,3,4] : p_{k,n,\eta} \leq v_n^y\}$ then
use modulation scheme η (i.e. with c_η bits) on nth subchannel
set $u_{k,n}^y = p_{k,n,l}$; $f_k^y = f_k^y + c_\eta$; $g_k^y = g_k^y + c_\eta$
set $p_{k,n,m}^y = p_{k,n,m} - p_{k,n,l}$, $\forall m > l$
set $b_{k,n,m}^{y+1} = b_{k,n,l} - c_\eta$, $\forall m > l$
set $b_{k,n,m}^{y+1} = 0$, $\forall m \leq l$
end if
until $n < N + 1$
until no further improvement on total data rate (equation (4.49))
the vectors $\boldsymbol{f}^{y+1}$ and $\boldsymbol{g}^{y+1}$ contain the bits allocated for each subchannel in category one and two respectively
the vector $\boldsymbol{u}^{y+1}$ contains the power allocated for each subchannel

when the SUBS maximum transmit power is at $20\,dBm$ and $40\,dBm$ are considered in Fig. 4.2 and Fig. 4.3 respectively. The results of the direct communication are similar to the ones obtained in [48, 65, 66, 120], therefore validating the simulations. From the results obtained, it is obvious that for the developed RA problem to have feasible solutions, the constraint of category one users, i.e. their minimum rate guarantee, has to be met at all times (since they have the higher priority). It can also be observed that at a higher permissible interference limit to the PUs, the average data rates for the categories of SUs improve. The improvement however tends to favour the category two users, because it is more convenient to improve their performance gradually at a slightly higher resource than it is to improve the performance of the category one users. Again, it can be seen that at a higher SUBS

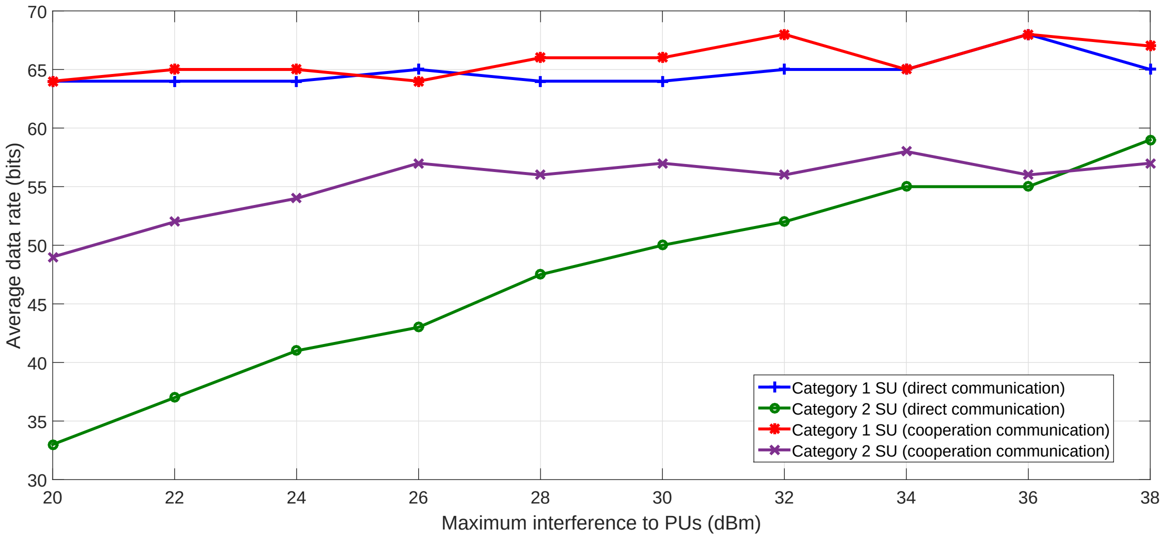

Figure 4.2. Average data rate for different categories of SUs, and with both direct and cooperative communication considered at 20 dBm maximum SUBS power.

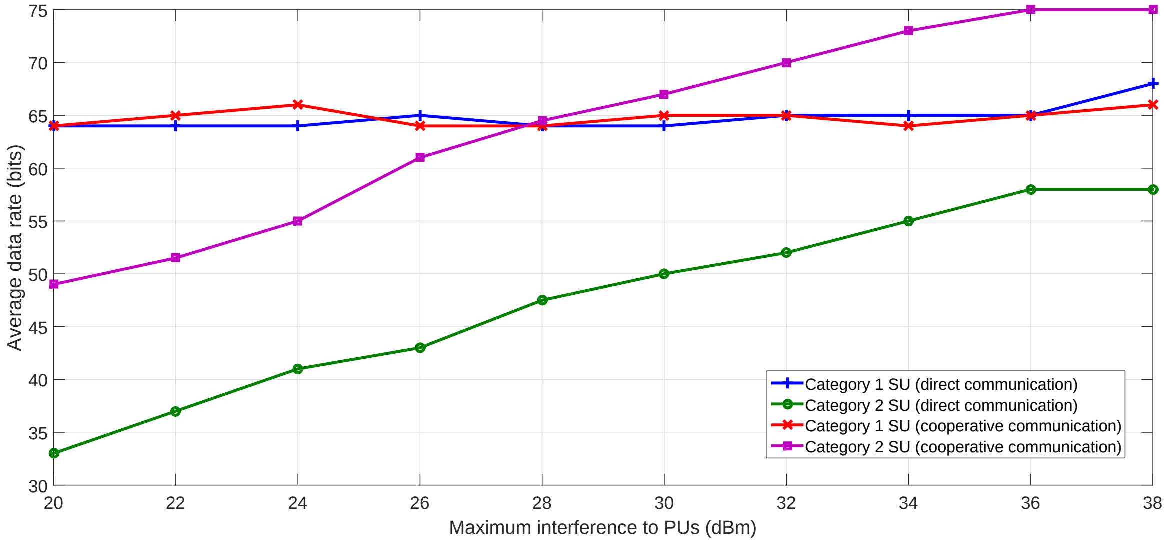

Figure 4.3. Average data rate for different categories of SUs, and with both direct and cooperative communication considered at 40 dBm maximum SUBS power.

power (40 dBm), the average data rate is better than at a lower SUBS power (20 dBm). Importantly, the result shows that a marked improvement in performance of the network is achieved during cooperation, compared to when direct communication alone is employed. Both categories of SUs realised a higher average data rate during cooperation. The reason for this is the improvement in the interference gain to PUs that is achieved during cooperation. As a result, with cooperative communication, the

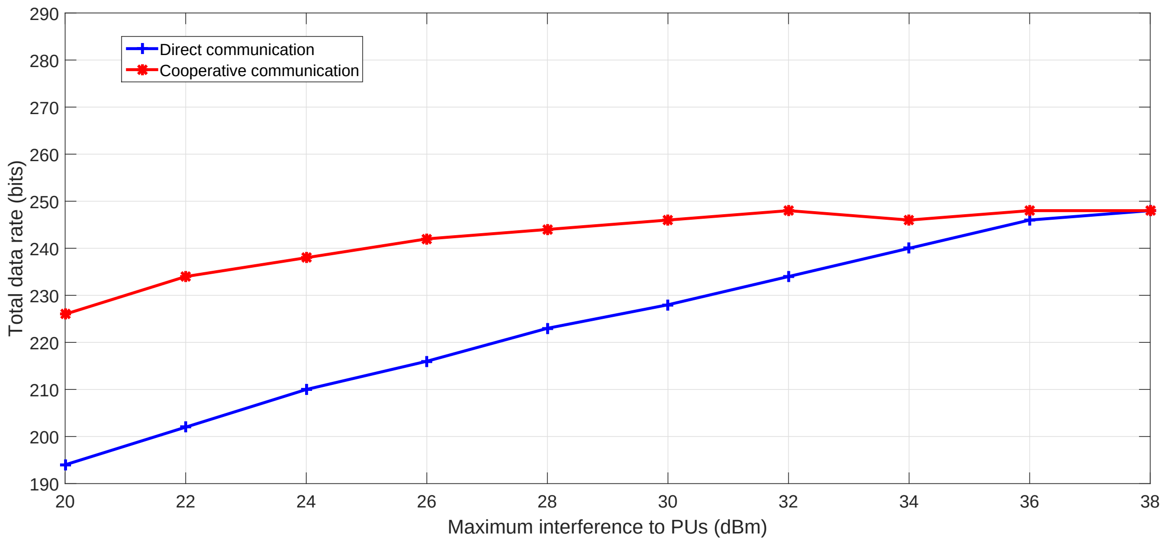

Figure 4.4. Total data rate for different categories of SUs, and with both direct and cooperative communication considered at 20dBm maximum SUBS power.

subchannels are able to transmit at a higher rate than they would ordinarily have been able to were they to communicate only directly. It is also worth noting that in Fig. 4.2, the average data rate during cooperation eventually converges to nearly that of direct communication (same would have happened in Fig. 4.3 if the permissible PU interference is increased much further. A similar trend is observable in Fig. 4.7 and Fig 4.9). This shows that as the permissible interference level to PUs increases, the need for and/or effect of cooperation diminishes. It would be better to transmit directly if the PUs are robust to the SUs' interference than to transmit using cooperation, as cooperation generally requires much more signalling overhead than direct communication.

In Figs. 4.4 and 4.5, the total data rate (bits) for each category of SUs and the maximum interference power to the PUs for both direct communication and cooperative communication are compared. Similar to Figs. 4.2 and 4.3, cases when the SUBS maximum transmit power is at 20dBm and 40dBm are considered in Fig. 4.4 and Fig. 4.5 respectively. The explanations given for Figs. 4.2 and 4.3 are also applicable in this instance, as the total data rate during cooperation generally outperforms that of its direct communication counterpart. Hence, similar reasoning and deductions about the throughput and the better performance of cooperation compared with direct communication can also be made.

Figs. 4.6 and 4.7 describe the average data rate performance for an increasing SUBS power. The two categories of SUs are covered and both direct and cooperative communication are considered. In Fig.

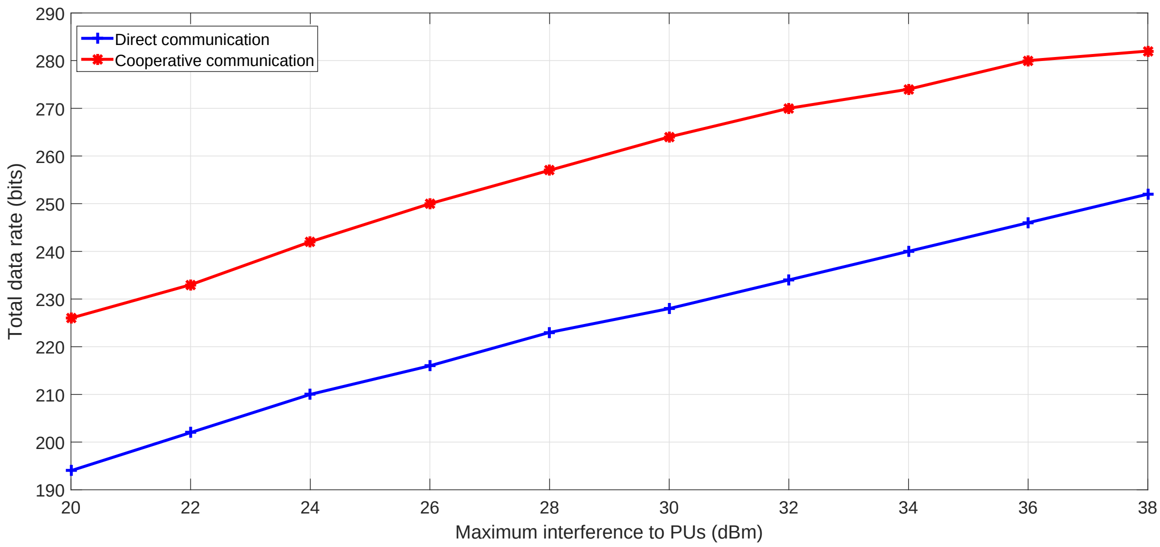

Figure 4.5. Total data rate for different categories of SUs, and with both direct and cooperative communication considered at 40 dBm maximum SUBS power.

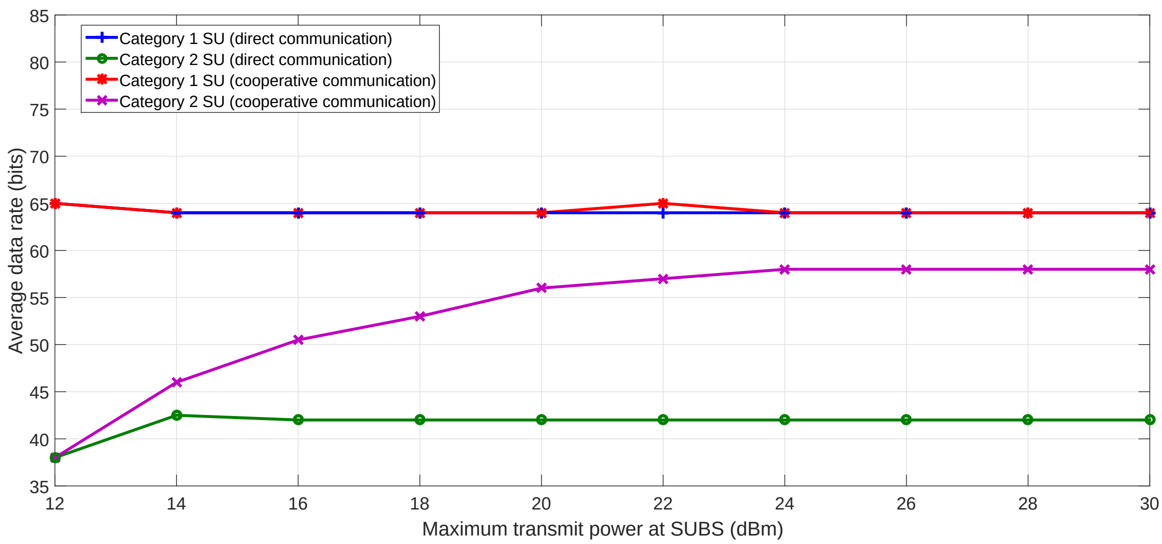

Figure 4.6. Average data rate for different categories of SUs, and with both direct and cooperative communication considered at 25 dBm maximum interference to PUs.

4.6, a maximum interference power to PUs of 25 dBm is employed, while the maximum interference power to PUs in Fig. 4.7 has been increased to 45 dBm. The plots present some of the feasible regions of the established problem. At all times, the minimum rate guarantee of the category one SUs must be met for the problem to have feasible solutions. It is noted that for the given parameters, at an SUBS of less than 12 dBm, it becomes infeasible to obtain solutions to the RA problem. Again, as the SUBS

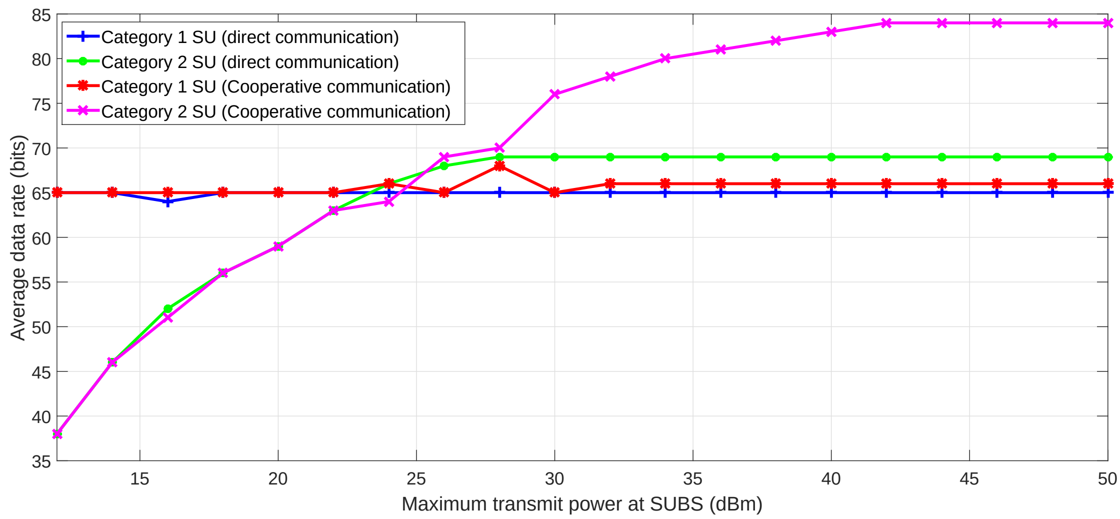

Figure 4.7. Average data rate for different categories of SUs, and with both direct and cooperative communication considered at 45 dBm maximum interference to PUs.

power is increased, the average data rate improves, particularly for category two SUs. The SUs in category two are much easier to satisfy and improve, which is the obvious reason for their continuous upgrading with a slight increase in available resources. After a while though, the data rate reaches a peak value and stabilises. No further improvement can be observed irrespective of whether or not the SUBS power is increased. The reason for this is that the other constraints are considered as well, thus making it impossible for the data rate to keep increasing indefinitely with increasing SUBS power. It is significant to note the sizeable improvement that cooperative communication achieves over direct communication. This improvement can be seen both when the interference limit is at 25 dBm (Fig. 4.6) and when it is at 45 dBm (Fig. 4.7) though for the latter, the improvement due to cooperation only begins to be observable at an SUBS power of about 26 dBm. It shows therefore that the network would rather transmit using direct communication when the SUBS power is limited so as to maximise the power usage and reduce signalling overhead. At higher power, however, cooperative communication is preferred, as the overall capacity is remarkably better. It is also worth noting that in Fig. 4.7 (and Fig 4.9), the average data rate (and total data rate) during cooperation eventually saturates (remains the same value) as the maximum transmit power is increased further. The increase is not indefinite because other constraints coming into play eventually limits the improvement in performance that can be realised.

Figs. 4.8 and 4.9 present the total data rate as realised by the network for all the categories of SUs,

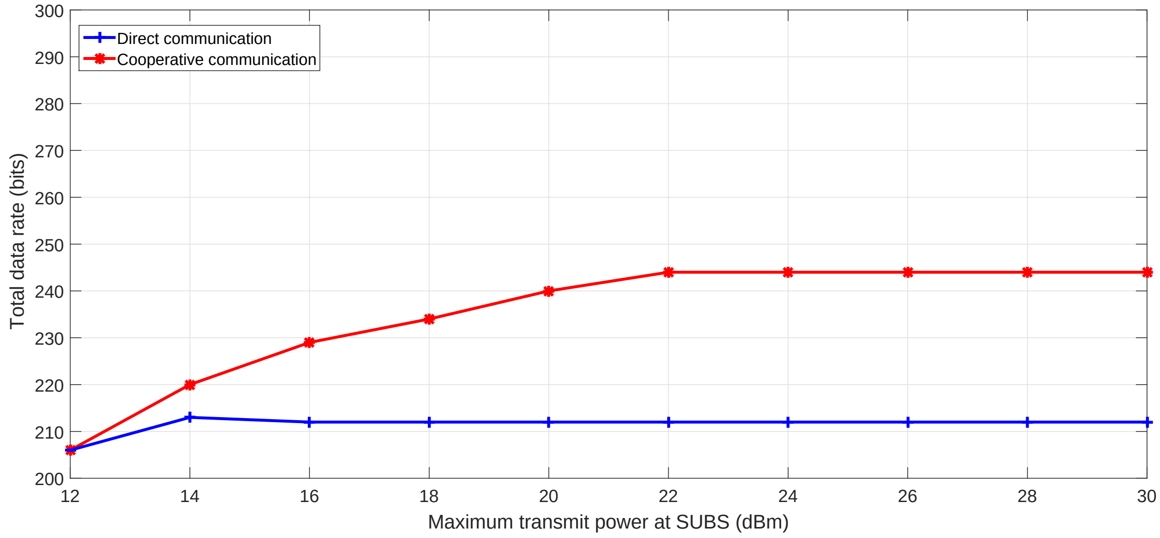

Figure 4.8. Total data rate for different categories of SUs, and with both direct and cooperative communication considered at 25 dBm maximum interference to PUs.

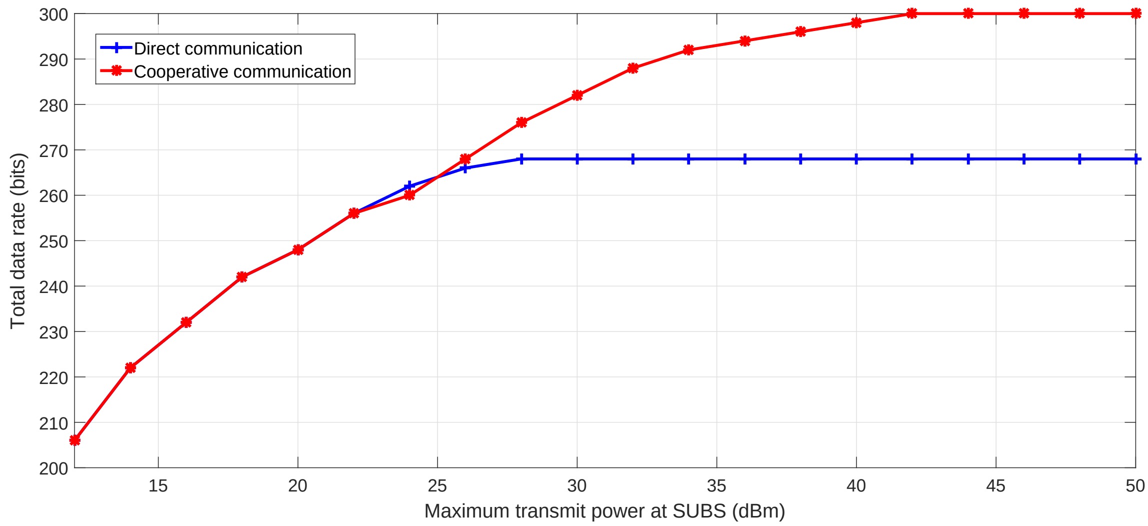

Figure 4.9. Total data rate for different categories of SUs, and with both direct and cooperative communication considered at 45 dBm maximum interference to PUs.

with both direct and cooperative communication. Also similar to Figs. 4.6 and 4.7, the cases presented are when the maximum interference to PUs is at 25 dBm and 45 dBm, as seen in Fig. 4.8 and Fig. 4.9 respectively. The explanations given for Figs. 4.6 and 4.7 are also very appropriate for describing the performance of the network. The total data rates during cooperation generally outperform rates achieved during direct communication. The reasoning and inferences given for Figs. 4.6 and 4.7 are also applicable in understanding the results of Fig. 4.8 and Fig. 4.9.

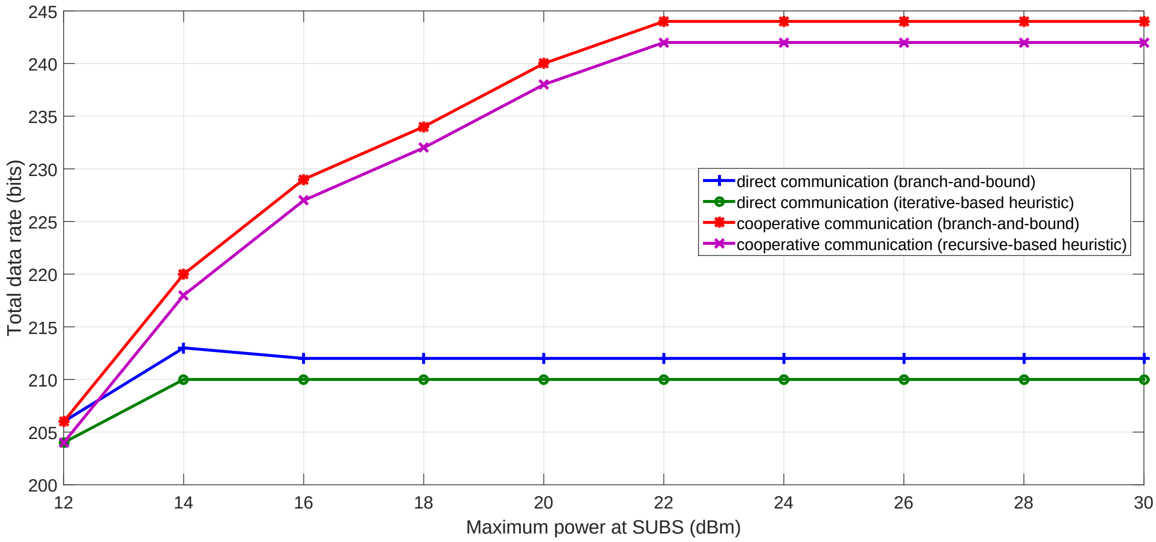

Figure 4.10. Performance of the BnB method is compared with the iterative-based heuristic approach using the total data rate for an increasing SUBS power (max interference to PUs is at $25\,dBm$).

In Figs. 4.10 and 4.11, the optimality of the network and the complexity are compared for the ILP (using BnB) and the iterative-based heuristic. The results are comparative to the ones obtained in [65]. The computational complexity is obtained from the number of arithmetic operations that the network undergoes before arriving at the solution [65]. For the heuristic, the total complexity is the sum of the complexities of the two parts (subchannel allocation and the iterative bit and power allocation). The results presented show that while the heuristic performs very close to optimality in its total data rate for the network, the complexity is significantly less, especially as the network gets larger. For such large CRN systems therefore, developing appropriate heuristic(s) to solve them, thereby providing both feasible and timeous solutions with lower complexities, is recommended.

4.9 CONCLUSION

RA models that can yield outstanding productivity, even with very stringent constraints, are critical for a meaningful CRN realisation and eventual deployment. This chapter develops such a model whereby, in a heterogeneous CRN environment, cooperative diversity is employed in mitigating the limiting effects of interference to the PUs of the network, thereby achieving optimality in the RA solutions. To make the model feasible and close to practical, only one single best relay is selected from the available ones as the cooperating relay. Also, cooperation is only employed by users that have subchannels with a high interference gain to the PUs. The RA problem developed is first solved by a careful re-formulation

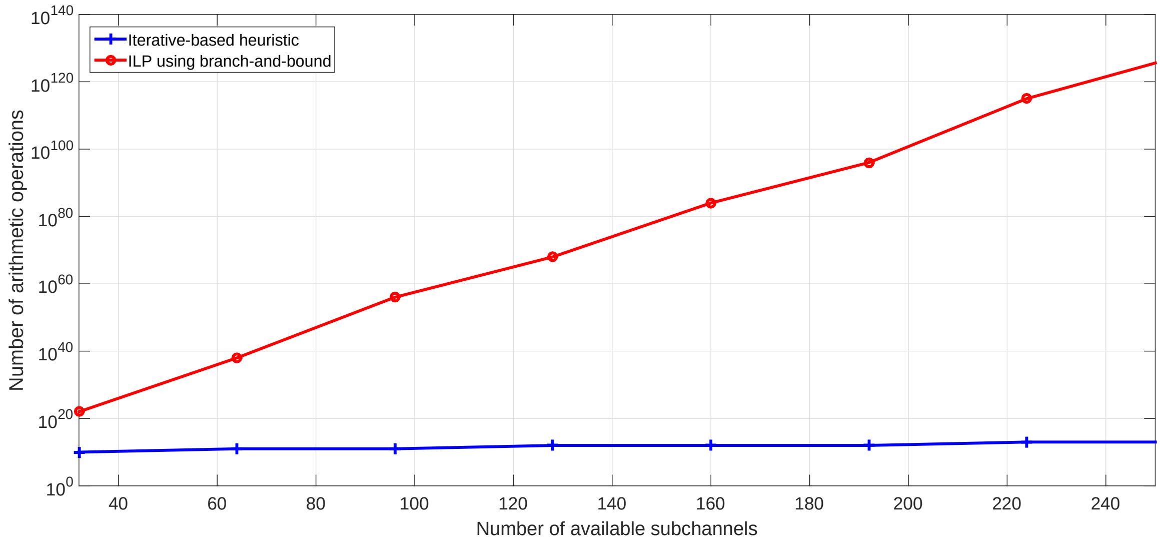

Figure 4.11. Performance of the BnB method is compared with the iterative-based heuristic approach using the computational complexity for different number of subchannels.

of the NP-hard problem into an ILP problem and optimal solutions are obtained using the BnB method for solving such ILP problems. To reduce computational complexity, an iterative-based heuristic is then developed to solve the problem in a much reduced time duration. The results presented compare the average data rates and the total data rates for the different categories of SUs when direct and cooperative communications are employed. Also, the optimality and computational complexity of the developed heuristic are compared with those obtained using ILP. The improvement in the performance of the network when cooperation is used is quite remarkable, as the results have shown.

CHAPTER 5 RESOURCE ALLOCATION SOLUTION FOR HETEROGENEOUS BUFFERED COGNITIVE RADIO NETWORKS

5.1 CHAPTER OVERVIEW

Resources available for operation in CRN are generally limited, making it imperative for efficient RA models that address its peculiar limitations to be designed for them. In the previous chapters, a considerable amount of work has been dedicated to developing such RA models. However, in those RA designs, a significant limiting factor has still been mostly ignored - the fact that different users or user categories do have different delay tolerance profiles. To address this limitation, in this chapter, a RA model for heterogeneous CRN with delay considerations is developed and analysed. In the model, the demands of users are first categorised and then, based on the distances of users from the controlling SUBS and with the assumption that the users are mobile, the user demands are placed in different queues having different service capacities and the resulting network is analysed using queueing theory. Furthermore, to achieve optimality in the RA process, an important concept is introduced whereby some demands from one queue are moved to another queue where they have a better chance of enhanced service, thus giving rise to an improvement in the overall performance of the network. The performance results obtained from the analysis, particularly the blocking probability and network throughput, show that the queueing model incorporated into the RA process can help in mitigating the effects of time delays and in achieving better productivity for the heterogeneous CRN with buffered data.

5.2 BACKGROUND

In recent times, there has been a deservedly growing interest in CRN as a possible driver for xG wireless communications. This special interest in CRN hinges on its promise of much higher resourcefulness, particularly in the spectrum usage. In its design, CRN enables an allotted spectrum space to be used by two different networks - a primary and a secondary one - under certain preconditions agreed upon by both networks [126]. With this kind of arrangement, a much better utilisation of the rather scarce spectrum becomes inevitable. This promise has triggered considerable research effort to develop and describe the CRN, as well as to address possible challenges to its introduction and implementation.

In CRN, PUs generally take priority in the usage of the resources, especially spectrum, because they are the original or licensed owners of it. The network of SUs must of necessity devise how to achieve and maintain an acceptable QoS, despite the stringent conditions under which it has to operate [78]. In heterogeneous CRN especially, the demands of the SUs are usually different from one SU to another, or from one category of users to another, and the CRN should be capable of meeting the different demands efficiently and timeously [112]. To make this possible, RA models that capture the essential peculiarities and dynamics of the heterogeneous users in CRN, and that can optimally assign the available resources fairly and favourably, are required [84].

In developing RA models for CRN, an important and realistic criterion for categorising demands of the heterogeneous users is their level of delay tolerance that is permissible for an acceptable QoS. Depending on the kind of service being provided, different users may have differing delay tolerance characteristics. Usually, because resources for SUs' transmission are limited or sometimes even temporarily unavailable (because of PUs' transmission, for instance), SUs, depending on the kind of service intended to be provided, may keep their data for transmission in a buffer (queue) and wait (usually for an acceptable time duration) for the requisite resources to be available for their transmission to be completed. Those delay instances and/or durations for SUs, and the properties of the queue developed as a result, may stand as a critical limitation in the CRN's QoS provisioning, unless it is adequately addressed. Applying queueing models in analysing the queue characteristics can help in addressing the limitation and improving the performance of CRN significantly. That is the main interest of this chapter. The chapter therefore develops and analyses an appropriate queueing model for heterogeneous CRN with users having different delay priorities or delay profiles. The analyses of the model show that, by investigating and exploiting the delay characteristics of users in CRN using queueing theory, a significant improvement in the optimal allocation of resources for the heterogeneous buffered CRN can be realised.

5.3 RELATED LITERATURE ON BUFFERING IN RESOURCE ALLOCATION FOR COGNITIVE RADIO NETWORKS

Developing appropriate RA models for CRN has been a recent research focus and some related works are briefly reviewed in this section. The authors in [55–57] analysed RA problems in heterogeneous MIMO CRN. In their works, the dual optimisation problems developed focused on optimising the transmit power and transmission time allocation of the SUs. To achieve their objectives, firstly, power allocation to each SU was optimised (on assumption of a constant transmission time), and secondly, optimal scheduling of the SUs was realised. In [59, 60], the authors studied RA in CRN, first with cooperative relays and then with imperfect spectrum sensing. The developed problems were non-convex and NP-hard. By separating the problems into two - subchannel allocation and power allocation - a convex programming conversion was achieved and suboptimal solutions realised. However, in the above-mentioned works and in most other related works on RA in CRN, no consideration has been given to the delay requirements of the SUs.

Works on RA in CRN that have incorporated some kind of time delay or data queueing in their RA problem formulations are indeed very few. The available ones, as obtained by the author during the course of this research work, are briefly reviewed. In [94], the authors, in developing their RA model, made provision for users that require real-time communication and gave them priority over non-real-time users. The real-time users were given priority in that they were admitted first into the network, as well as given sufficient resources to transmit their data. Thereafter, an optimal number of non-real-time users were made to share the remaining resources of the network. Thus, the delay profiles of the users were used as a means of controlling the number of users admitted to the network. Authors in [37] classified the SUs into delay-sensitive and delay-tolerant SUs, with the delay-sensitive SUs having the higher priority in the allocation of resources, as their delay requirements must always be met first. In the analysis, the delay requirements of the delay-sensitive SUs were simply transformed into a constant rate requirement using very elementary ideas of queueing theory. A study on band allocation in CRN when both PUs and SUs have data queues was carried out in [127]. In order to satisfy the required QoS for the SUs, the authors proposed that each SU be probabilistically assigned to a PU band, and showed that such arrangement gives a better performance than either random or fixed allocations for the SUs. In [89], the authors developed a model that separates the users by the amount of data backlog they have in their queues waiting to be transmitted. The allocation of resources was carried out based on the size of the backlog - users with small backlogs were given just sufficient resources to transmit their data, while users with large backlogs shared the remaining resources among themselves in a manner that was fair and efficient. In the above-mentioned works, the significance of time delays and the effects on the overall performance of the RA problems had not been studied. Moreover, improving overall network capacity by developing an appropriate queueing model that addresses the limitations in the RA formulation due to time delay was never carried out, making the work presented in this chapter uniquely different.

In the earlier works of the author (some of which have also been discussed in previous chapters), RA models for heterogeneous CRN were developed and studied but with no consideration for data buffering or delay possibilities [48, 120]. In this chapter, the more realistic possibility of users having buffered data is incorporated into the RA problems of CRN and its effects are investigated. Particularly, the RA model developed in this chapter addresses the very likely circumstance where heterogeneous SUs have data in their buffer waiting to be transmitted. With the SUs having different delay tolerance profiles, the model seeks to find a unique method of transmitting the SUs' data in a manner in which QoS is guaranteed for each SU, while the overall productivity of the network is enhanced. By categorising the SUs and developing efficient queueing and RA mechanisms to best satisfy their requirements, a much greater capacity can be realised for the network. The contributions in this chapter are summarised as:

- Developing and analysing a queueing model that captures and addresses the limiting effects of time delay in the various categories of users in buffered heterogeneous CRN.

- Investigating optimality in the RA process for the heterogeneous CRN by studying the impact of varying the values of user demands into different queues and determining their effect on the overall performance of the network. The variation in arrivals is achieved by moving a factor θ of demands from users in a farther queue to a nearer one. θ itself is such that it can be changed continuously within a certain range, and an optimal value for θ that maximizes the productivity of the network is realised.

5.4 SYSTEM MODEL

The system model is shown in Fig. 1. The model is a development on the work presented in [128]. The referenced work had considered an overlay network in which the SUs query a database to ascertain the PUs' transmission and had designed the SUs' connections according to the periodicity of the PUs' traffic patterns. However, in this model, a centralised, underlay heterogeneous CRN is developed. The SUs are made to transmit on the entire PUs' bandwidth within an acceptable interference power limit. This makes it possible to concentrate on analysing the SU network of the CRN separately and intently, so as to achieve the utmost for the SUs, and this without causing significant harm to the PUs. The SUs are classified based on their delay profiles as either delay-sensitive (DS) or delay-tolerant (DT) users, the classes differentiated by the duration of delay time for acceptable service. The SUs are assumed to be both mobile and capable of performing adaptive modulation and coding (AMC) to dynamically change their modulation and coding schemes. Moreover, the SUs, depending on their distance to the SUBS, are placed in virtual rings. The nearest ring to the SUBS operates with the highest AMC scheme, while the farthest ring operates with the lowest AMC scheme. Data transmission requests of users within a ring are placed in a queue of that ring and service (transmission of data) is carried out using the available subchannels. The queues therefore act as a buffer, should there be a possible delay in immediate transmission due to insufficiency in resource availability. There are N subchannels, which automatically correspond to the number of parallel servers in each ring (or queue). Queues are finite with a maximum length Y. As a result of the mobility of SUs, arrival rates into queues can be adjusted so that the maximum productivity of the network can be realised. To accomplish this, a fraction of the demands of users, particularly the DT demands, which have a high delay profile, is moved from a farther ring (queue) to a closer ring (queue), so that the demands can possibly be transmitted at a higher rate. The intent is that, by such an arrangement, a likely reduction in both the time and energy consumed in transmitting the data can be achieved, making it possible for a significant improvement in the capacity of the SUs and the entire CRN.

5.4.1 Queueing model

The queueing model is set up to determine the overall capacity of the system. The queueing analysis shows the significance of the fraction of demands that is moved between queues. To make the model manageable and easier to analyse, only two concentric rings are considered, meaning that there are two

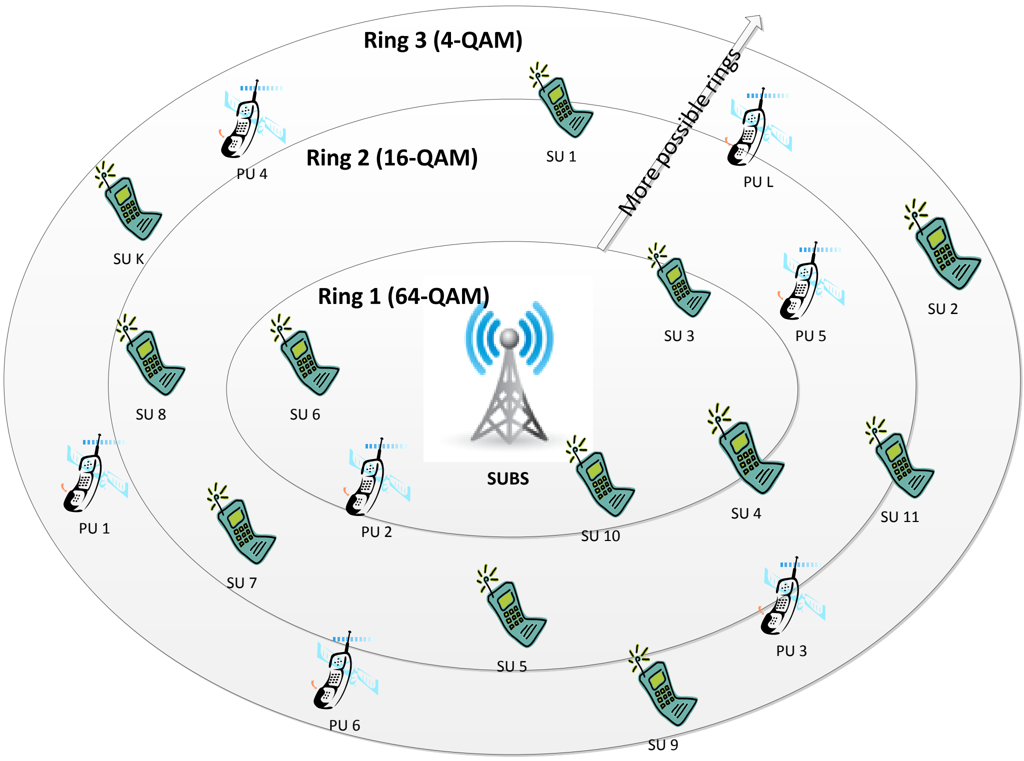

Figure 5.1. System model of the heterogeneous buffered CRN. Users are placed in virtual rings of different distance ranges from the SUBS

parallel queues, each served by multiple servers (subchannels). The analysis can however be extended to three or more rings. The closest ring to the SUBS is initially assigned to transmit at a modulation scheme of 64-QAM (6 bits per symbol) and the farthest ring is assigned to transmit with a modulation scheme of 4-QAM (2 bits per symbol). Arrivals into each of the queues follow a Poisson distribution with arrival rates λ_1 for queue 1 and λ_2 for queue 2. Service is exponential with rates μ_1 for queue 1 and μ_2 for queue 2. μ_1 and μ_2 correspond to the data rate of the AMC scheme operated in each ring, meaning 6 bits per symbol and 2 bits per symbol for queue 1 and queue 2 respectively. Service per unit time in queue 1 is therefore significantly faster than in queue 2, since it operates at a higher service rate. Some of the arrivals into each queue are DT demands and users can move from one ring to another. The essence is to determine whether the CRN's productivity can be improved by moving the DT demands of the farther ring to the ring nearer to the SUBS. In the model therefore, a fraction θ, which represents the DT demands of queue 2, is moved to queue 1 where the demand is capable of being transmitted at a higher data rate. The challenge is to be able to find the value θ, the fraction of the demands from the farther queue to be moved to the nearer queue that will not be counter-productive

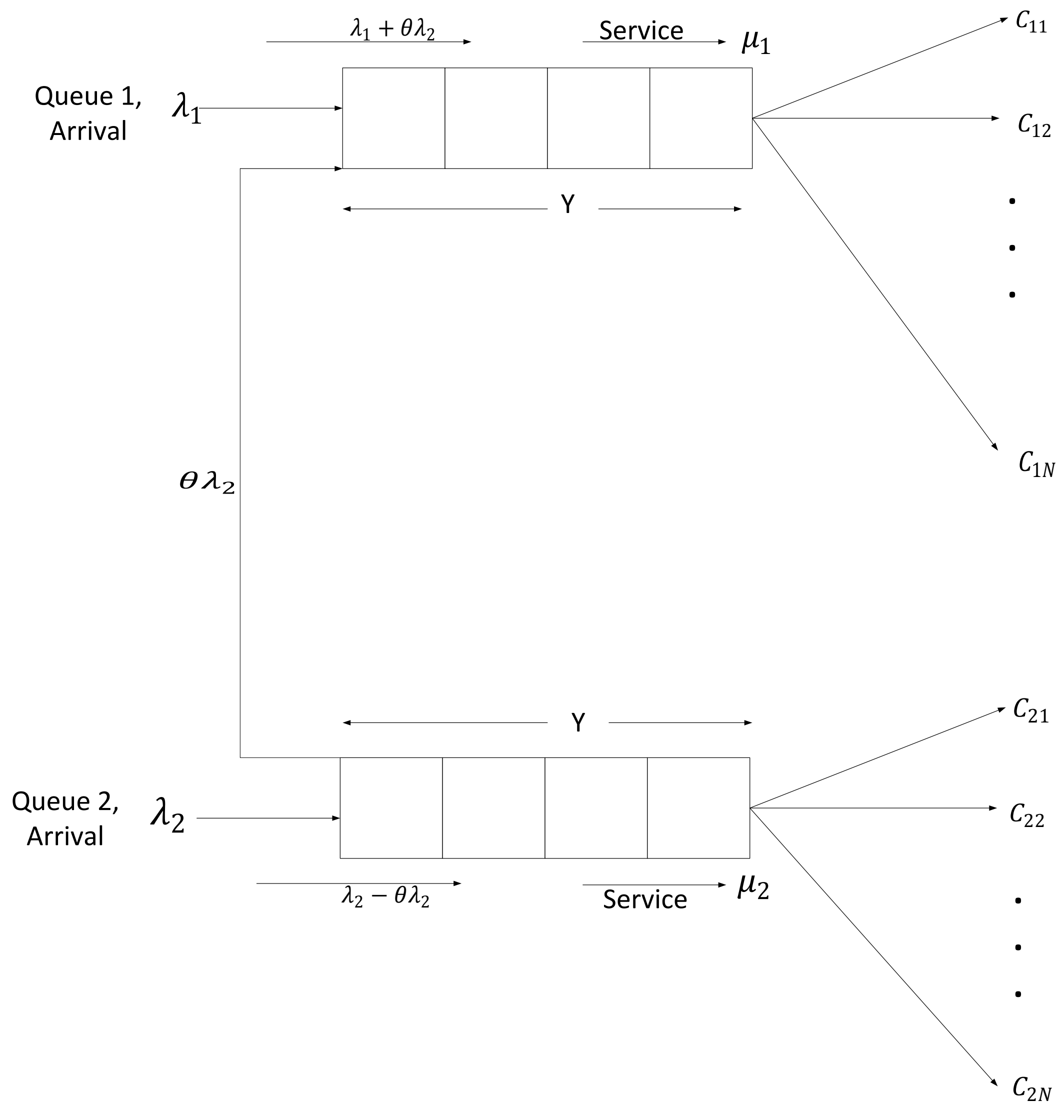

Figure 5.2. Queueing model for the heterogeneous buffered CRN

but will rather optimise the total productivity of the network.

5.4.2 Analysis of model

To analyse the queueing model developed, the following parameters are defined:

Total number of subchannels (which is also equivalent to the number of multiple servers in each queue) $= N$

Arrival rate into queue 1 $= \lambda_1$

Arrival rate into queue $2 = \lambda_2$

Fraction of queue 2 (DT demands) moved to queue $1 = \theta$

Total arrival into queue $1 = \lambda_1 + \theta\lambda_2$

Total arrival into queue $2 = \lambda_2 - \theta\lambda_2$

Total arrival into the network $\lambda = \lambda_1 + \lambda_2$, with the traffic intensities given as $\rho_1 = \frac{(\lambda_1+\theta\lambda_2)}{N\mu_1}$; $\rho_2 = \frac{(\lambda_2-\theta\lambda_2)}{N\mu_2}$; $\rho = \rho_1 + \rho_2$

$$\text{Service rate of queue 1} = \begin{cases} n\mu_1, & 0 \leq n \leq N; \\ N\mu_1, & N \leq n \leq Y \end{cases}$$

$$\text{Service rate of queue 2} = \begin{cases} n\mu_2, & 0 \leq n \leq N; \\ N\mu_2, & N \leq n \leq Y \end{cases}$$

The queueing model is a continuous-time Markov chain (CTMC) queue with a finite buffer. The model is shown in Fig. 5.2. From Fig. 5.2, it can be observed that the total arrival into queue 1 is the addition of the original arrival λ_1 and the fraction of arrival to queue 2 that is redirected to queue 1. Similarly, the total arrival to queue 2 becomes what is left of the original arrival after the fraction $\theta\lambda_2$ has been taken away. If there are no arrivals into queue 2 or no part of queue 2 is moved to queue 1, then $\theta\lambda_2 = 0$ and arrival to queue 1 is simply limited to λ_1.

Let $X = \{(k,l),\ 0 \leq k \leq Y,\ 0 \leq l \leq Y\}$ be the state space of the combined queues, where $k(l)$ represents the number of data packets in the system from queue $1(2)$. Hence, $(k,l) \in X$. The state space diagram is shown in Fig. 5.3. $x_{k,l}(t)$ (for simplicity, this is subsequently written as $x_{k,l}$) is the probability that at time t there are (k,l) data packets in the system, implying that there are k packets in queue 1 and l packets in queue 2 available for transmission (including the packets in service). The system can now be studied at steady state using the stationary distribution conditions. If $\boldsymbol{x} = [x_{0,0}\ x_{0,1}\ \ldots\ x_{0,Y}\ x_{1,0}\ x_{1,1}\ \ldots\ x_{1,Y}\ \ldots\ x_{Y,0}\ x_{Y,1}\ \ldots\ x_{Y,Y}]$ is the steady state vector, the conditions $\mathbf{0} = \boldsymbol{xQ}$ and $1 = \boldsymbol{xe}$ for CTMCs hold, where $\boldsymbol{Q}$ is the generator matrix and $\boldsymbol{e}$ is the identity matrix.

From the state space diagram as shown in Fig. 5.3 and using the steady state conditions, the equilibrium balance equations are obtained as:

$$0 = -\lambda x_{0,0} + \mu_2 x_{0,1} + \mu_1 x_{1,0} \tag{5.1}$$

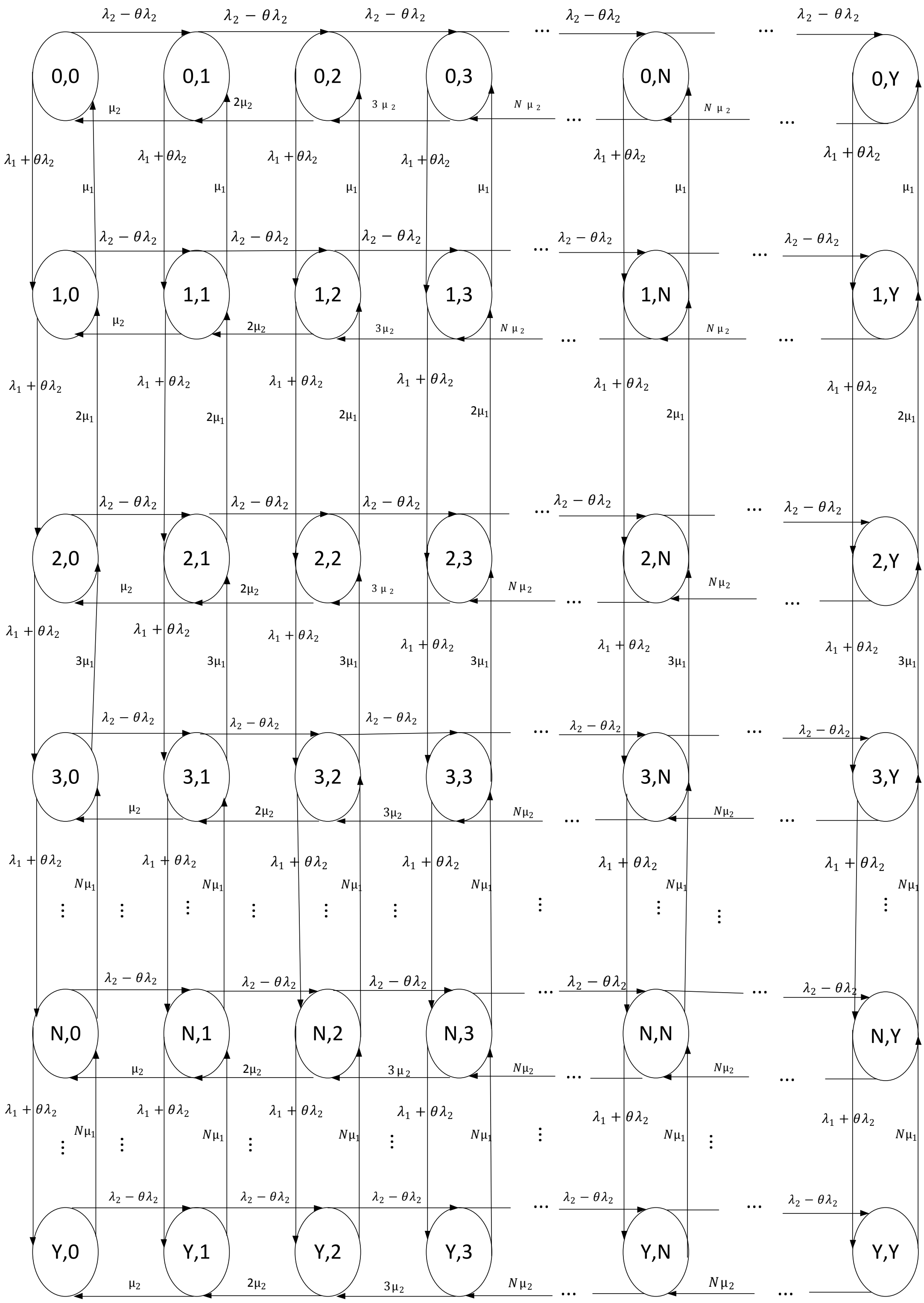

Figure 5.3. State space for the queueing model of RA in heterogeneous buffered CRN

$$0 = (\lambda_2 - \theta\lambda_2)x_{0,0} - (\lambda + \mu_2)x_{0,1} + 2\mu_2 x_{0,2} + \mu_1 x_{1,1} \tag{5.2}$$
$$0 = (\lambda_2 - \theta\lambda_2)x_{0,1} - (\lambda + 2\mu_2)x_{0,2} + 3\mu_2 x_{0,3} + \mu_1 x_{1,2} \tag{5.3}$$
$$0 = (\lambda_2 - \theta\lambda_2)x_{0,N-1} - (\lambda + N\mu_2)x_{0,N} + N\mu_2 x_{0,N+1} + \mu_1 x_{1,N} \tag{5.4}$$
$$0 = (\lambda_2 - \theta\lambda_2)x_{0,Y-1} - (\lambda_1 + \theta\lambda_2 + N\mu_2)x_{0,Y} + \mu_1 x_{1,Y} \tag{5.5}$$

$$0 = (\lambda_1 + \theta\lambda_2)x_{0,0} - (\lambda + \mu_1)x_{1,0} + \mu_2 x_{1,1} + 2\mu_1 x_{2,0} \tag{5.6}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{0,1} + (\lambda_2 - \theta\lambda_2)x_{1,0} - (\lambda + \mu_1 + \mu_2)x_{1,1} + 2\mu_2 x_{1,2} + 2\mu_1 x_{2,1} \tag{5.7}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{0,2} + (\lambda_2 - \theta\lambda_2)x_{1,1} - (\lambda + \mu_1 + 2\mu_2)x_{1,2} + 3\mu_2 x_{1,3} + 2\mu_1 x_{2,2} \tag{5.8}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{0,N} + (\lambda_2 - \theta\lambda_2)x_{1,N-1} - (\lambda + \mu_1 + N\mu_2)x_{1,N} + N\mu_2 x_{1,N+1} + 2\mu_1 x_{2,N} \tag{5.9}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{0,Y} + (\lambda_2 - \theta\lambda_2)x_{1,Y-1} - (\lambda_1 + \theta\lambda_2 + \mu_1 + N\mu_2)x_{1,Y} + 2\mu_1 x_{2,Y} \tag{5.10}$$

$$0 = (\lambda_1 + \theta\lambda_2)x_{N-1,0} - (\lambda + N\mu_1)x_{N,0} + \mu_2 x_{N,1} + N\mu_1 x_{N+1,0} \tag{5.11}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{N-1,1} + (\lambda_2 - \theta\lambda_2)x_{N,0} - (\lambda + N\mu_1 + \mu_2)x_{N,1} + 2\mu_2 x_{N,2} + N\mu_1 x_{N+1,1} \tag{5.12}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{N-1,2} + (\lambda_2 - \theta\lambda_2)x_{N,1} - (\lambda + N\mu_1 + 2\mu_2)x_{N,2} + 3\mu_2 x_{N,3} + N\mu_1 x_{N+1,2} \tag{5.13}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{N-1,N} + (\lambda_2 - \theta\lambda_2)x_{N,N-1} - (\lambda + N\mu_1 + N\mu_2)x_{N,N} + N\mu_2 x_{N,N+1} + N\mu_1 x_{N+1,N} \tag{5.14}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{N-1,Y} + (\lambda_2 - \theta\lambda_2)x_{N,Y-1} - (\lambda_1 + \theta\lambda_2 + N\mu_1 + N\mu_2)x_{N,Y} + N\mu_1 x_{N+1,Y} \tag{5.15}$$

$$0 = (\lambda_1 + \theta\lambda_2)x_{Y-1,0} - (\lambda_2 - \theta\lambda_2 + N\mu_1)x_{Y,0} + \mu_2 x_{Y,1} \tag{5.16}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{Y-1,1} + (\lambda_2 - \theta\lambda_2)x_{Y,0} - (\lambda_2 - \theta\lambda_2 + N\mu_1 + \mu_2)x_{Y,1} + 2\mu_2 x_{Y,2} \tag{5.17}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{Y-1,2} + (\lambda_2 - \theta\lambda_2)x_{Y,1} - (\lambda_2 - \theta\lambda_2 + N\mu_1 + 2\mu_2)x_{Y,2} + 3\mu_2 x_{Y,3} \tag{5.18}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{Y-1,N} + (\lambda_2 - \theta\lambda_2)x_{Y,N-1} - (\lambda_2 - \theta\lambda_2 + N\mu_1 + N\mu_2)x_{Y,N} + N\mu_2 x_{Y,N+1} \tag{5.19}$$
$$0 = (\lambda_1 + \theta\lambda_2)x_{Y-1,Y} + (\lambda_2 - \theta\lambda_2)x_{Y,Y-1} - (N\mu_1 + N\mu_2)x_{Y,Y} \tag{5.20}$$

The set of linear equations (5.1) - (5.20), being finite, can be solved simultaneously using the standard (direct) approach, such as Gaussian elimination or the iterative approach, such as Jacobi or Gauss-Seidel methods. Using any of these approaches, state probabilities can be obtained, which can then be used to obtain the various performance measures of interest.

Although the problem that has been developed can be solved using the standard approach, other methods for obtaining solutions can be investigated, and two of these are considered in this chapter.

Firstly, by a careful consideration of the model developed, it is observed that the overall system is closely comparable to a two-class, non-priority queueing model in which customers are served using the first-come first-served discipline with different arrival rates and service rates. According to [129], such a system can be regarded as an M/G/1 queue, where G is a mixture of two exponential distributions (a hyper-exponential distribution) and packet arrivals are grouped into a single arrival stream. The analysis of this system has already been carried out in [129] and gives the following expressions for the queue lengths $L_q^{(1)}$, $L_q^{(2)}$ and L_q representing the queue length of queue 1, queue 2 and the overall queue respectively:

$$L_q^{(1)} = \frac{(\lambda_1 + \theta\lambda_2)(\rho_1/\mu_1 + \rho_2/\mu_2)}{1-\rho}, \tag{5.21}$$

$$L_q^{(2)} = \frac{(\lambda_2 - \theta\lambda_2)(\rho_1/\mu_1 + \rho_2/\mu_2)}{1-\rho}, \tag{5.22}$$

$$L_q = \frac{\lambda(\rho_1/\mu_1 + \rho_2/\mu_2)}{1-\rho}. \tag{5.23}$$

The second method employed in this chapter to obtain a solution to the developed RA queueing problem, and certainly the most important for this thesis, is using state reduction. State reduction is a technique for finding steady state probabilities for finite and even for infinite state Markov chains, developed by Grassmann, Taksar and Heyman, and generally called the GTH algorithm [130]. This approach is very important in that it is employed in studying the effect of, and finding an optimal value for the parameter θ. By varying θ, the arrival rates into each queue can be adjusted and the effects on the overall performance of the network investigated. After obtaining the equilibrium probabilities of the Markov chain through state reduction, the optimum value of the parameter θ is then obtained through Newton's method. It is important to note that the definition of θ is not necessarily limited to the fraction of the DT demands being moved from one queue to the other alone, as is being considered in this chapter. It could be defined to be any other factor, for instance a fraction of the higher priority demands. The important question to be addressed is how to determine the value of θ that will maximise the overall productivity of the network.

For easier depiction, let i represent the state (k,l). Also, for each $i \in X$, let r_i be the additional data rate (reward) achieved through the movement of θ data packets from queue 1 to queue 2. From any state $i \in X$, the rate of going to another state $j \in X$ is q_{ij}, where the q_{ij}s form the elements of the generator matrix Q. Implicitly, both r_i and q_{ij} depend on the parameter θ. Hence, to achieve an optimal total data rate (maximum throughput), an optimal value of θ has to be obtained such that the total additional data rates (total rewards) are maximised when the system is in equilibrium (steady state). This therefore defines the problem as that of determining the optimal value of θ when the network is in steady state. Let T be the total expected rewards (total additional data rate due to θ); the objective of the optimisation problem is given as:

$$\max T(\theta) = \sum_{i \in X} x_i r_i, \tag{5.24}$$

and the constraints are the limitations in both arrival and service rates of the network, as well as the time delays and the minimum data rate for an acceptable QoS.

Since θ is a variable entity, differential calculus can be employed in finding its optimal value. Let $T^{'}(\theta)$ and $T^{''}(\theta)$ be the respective first and second derivative of $T(\theta)$ with respect to θ (for ease of reference, $T(\theta)$, $T^{'}(\theta)$ and $T^{''}(\theta)$ are subsequently represented as T, $T^{'}$ and $T^{''}$ respectively). T is maximised by solving $T^{'} = 0$ for θ. To solve $T^{'} = 0$ using Newton's method, $T^{''}$ has to be obtained. Also, since the problem is a maximisation problem, $T^{''}$ has to be negative. To obtain $T^{'}$ and $T^{''}$, intermediate results from state reduction are used [131]. State reduction is first employed to obtain the steady state vector $\boldsymbol{x}$, the values of which are then used to obtain T and its derivatives.

Using the state reduction method to obtain $\boldsymbol{x}$ requires a relaxation of the steady state equation $\boldsymbol{xe} = \sum_{i \in X} x_i = 1$, so that the sum of x_i is no longer 1. A different solution is obtained, say, g_i and let its vector be $\boldsymbol{g}$. To differentiate g_i from x_i, g_i is referred to as the chance that there are i data packets in the network, while x_i is the probability that there are i data packets in the network at any given time. It can immediately be observed that the sum of g_i is not 1. Rather, a certain state, say state b, is given a chance $g_b = 1$ and the chances for all other states are calculated accordingly. Then, to find the probability x_i of being in state i, the chance g_i is divided by the sum of all chances $G = \sum_{i \in X} g_i$. Hence,

$$x_i = \frac{g_i}{\sum_{i \in X} g_i},$$

$$x_i = \frac{g_i}{G}, \tag{5.25}$$

or, $g_i = x_i G$. Differentiating twice with respect to θ, this becomes:

$$g_i^{'} = x_i^{'}G + x_i G^{'},$$
$$g_i^{''} = x_i^{''}G + 2x_i^{'}G^{'} + x_i G^{''},$$

hence,

$$x_i^{'} = \frac{g_i^{'} - x_i G^{'}}{G}, \tag{5.26}$$

$$x_i^{''} = \frac{g_i^{''} - 2x_i^{'}G^{'} - x_i G^{''}}{G}. \tag{5.27}$$

The values $g_i, g_i^{'}, g_i^{''}$ are obtained using the state reduction (or GTH) algorithm. The algorithm is provided in Appendix A. Once these values have been obtained, substituting into the set of equations above gives the corresponding $x_i, x_i^{'}, x_i^{''}$. Also, after obtaining $g_i, g_i^{'}, g_i^{''}$, obtaining the optimal value for θ is achieved by applying Newton's method. This is carried out as follows:
Assume that there are $M+1$ states numbered from 0 to M and M is finite. Recall $T = \sum_{i=0}^{M} x_i r_i$. But $x_i = \frac{g_i}{G}$. Substituting for x_i gives,

$$T = \frac{\sum_{i=0}^{M} g_i r_i}{G},$$
$$TG = \sum_{i=0}^{M} g_i r_i.$$

Taking first and second derivatives with respect to θ becomes,

$$T^{'}G + TG^{'} = \sum_{i=0}^{M} (g_i^{'} r_i + g_i r_i^{'}),$$

$$T^{''}G + 2T^{'}G^{'} + TG^{''} = \sum_{i=0}^{M}(g_i^{''}r_i + 2g_i^{'}r_i^{'} + g_i r_i^{''}).$$

The above submissions yield the following equations for T, $T^{'}$ and $T^{'}$:

$$T = \frac{\sum_{i=0}^{M} g_i r_i}{G}, \tag{5.28}$$

$$T^{'} = \frac{\sum_{i=0}^{M}(g_i^{'}r_i + r_i^{'}g_i) - TG^{'}}{G}, \tag{5.29}$$

$$T^{''} = \frac{\sum_{i=0}^{M}(g_i^{''}r_i + 2g_i^{'}r_i^{'} + r_i^{''}g_i) - 2T^{'}G^{'} - TG^{''}}{G}. \tag{5.30}$$

To obtain the optimal value of θ, an approximate value of θ is first chosen, say θ_m, and a new value θ_{m+1}, presumably a better approximation, is obtained according to Newton's method as follows:

$$\theta_{m+1} = \theta_m - \frac{T^{'}}{T^{''}}. \tag{5.31}$$

Substituting for $T^{'}$ and $T^{''}$ gives,

$$\theta_{m+1} = \theta_m - \frac{\sum_{i=0}^{M}(g_i^{'}r_i + r_i^{'}g_i) - TG^{'}}{\sum_{i=0}^{M}(g_i^{''}r_i + 2g_i^{'}r_i^{'} + r_i^{''}g_i) - 2T^{'}G^{'} - TG^{''}}. \tag{5.32}$$

In the above formulation, maximising θ is subject to $T^{''} < 0$. If not, then the sign of $T^{'}$ has to be considered. If $T^{'} > 0$, the rewards can be increased by increasing θ. If $T^{'} < 0$, the rewards can be increased by decreasing θ.

Performance measures employed to demonstrate the effect of θ on the network, as considered in this chapter, are the blocking probability and system throughput. Both of these measures are obtainable using the steady state probabilities. Other performance measures, such as the average number of packets in the queue or in the system and the average waiting time of a packet in the queue or before the packet's transmission is completed, are easily obtainable from the two measures evaluated but for the sake of brevity, they are not considered in this chapter. The blocking probability is the probability that a packet that arrives in the network is blocked or dropped because of its meeting a full system (i.e., all servers are fully engaged and the waiting queue is full) and is therefore not served. In steady state, the blocking probability P_B is defined as:

$$\text{Blocking probability}, P_B = \Pr\{\text{a packet arrives to meet a full system}\},$$
$$P_B = \Pr\{\text{system is full}\} = x_{Y,Y}. \tag{5.33}$$

The system throughput is defined as the number of total arrivals that are eventually served, i.e., the number of total arrivals that are not dropped or blocked. The throughput is, in effect, the effective arrival rate, represented as λ_e. The throughput is given as follows:

$$\text{Throughput} = \text{Effective arrival rate}, \lambda_e = \text{Total arrival rate}, \lambda \times \Pr\{\text{system is NOT full}\},$$
$$\text{Throughput} = \lambda(1 - P_B) = \lambda(1 - x_{Y,Y}). \tag{5.34}$$

5.5 RESULTS AND DISCUSSION

In this section, performance results of the developed model are presented. For ease of analysis, the network model, comprising two separate queues, is limited to $N = 2$ servers (subchannels) in each queue. A queue length of $Y = 2$ is equally employed in each queue. The model, parameters used (number of servers, queue length, etc.), as well as the results obtained are comparable to, and validated by the work in [131]. The model is however easily scalable, and the results obtained are fair representations of larger networks. For the first consideration (as presented in the first set of plots), SU demands in queue 1 are served at a data rate of 6 bits per symbol (64-QAM modulation) for each unit of time, while demands in queue 2 are served at a rate of 2 bits per symbol (4-QAM modulation) for each unit of time. In the second consideration, the service rates at both queues are reduced to study the effect of such reduction on the overall performance of the network. SU demands in queue 1 are therefore served at 2 bits per symbol (4-QAM modulation) for each time unit, while demands in queue 2 are served at a rate of 1 bit per symbol (BPSK modulation) for each time unit. In all analyses, the arrival rate to each of the queues is gradually increased from 1 to 10 bits per symbol per unit of time.

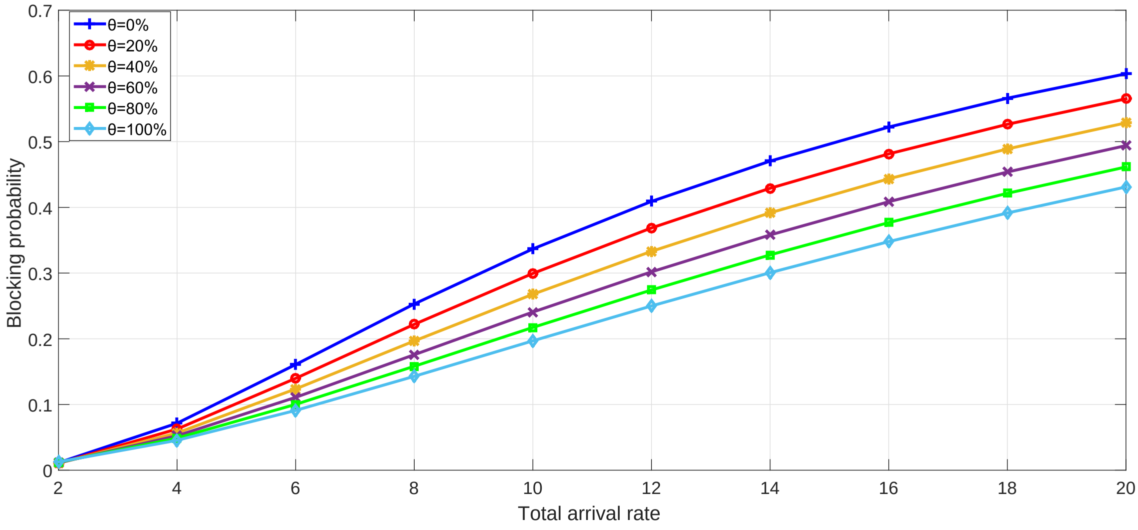

Figure 5.4. Blocking probability against total arrival rate for different θ values. Service rates are 6 bits/symbol/unit time in queue 1 and 2 bits/symbol/unit time in queue 2.

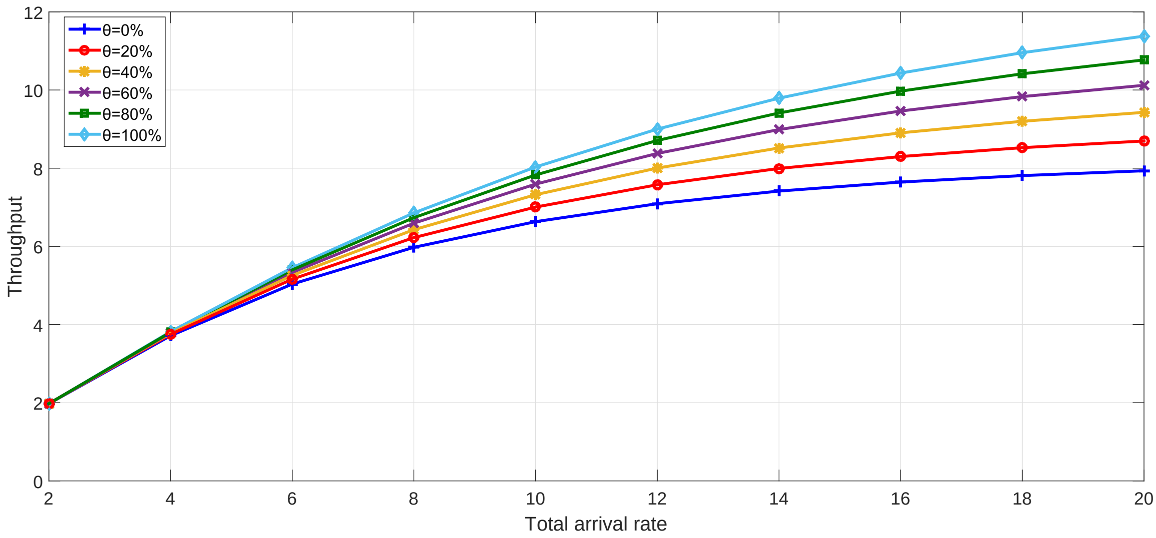

Figure 5.5. Throughput against total arrival rate for different θ values. Service rates are 6 bits/symbol/unit time in queue 1 and 2 bits/symbol/unit time in queue 2.

The value of θ is varied between $0-100\%$ of λ_2. Performance results of blocking probability and throughput of the heterogeneous CRN are presented and discussed.

Figs. 5.4 and 5.5 present the blocking probability and throughput performance measures respectively, when the service rate is at 6 bits per symbol per unit time for queue 1 and 2 bits per symbol per unit

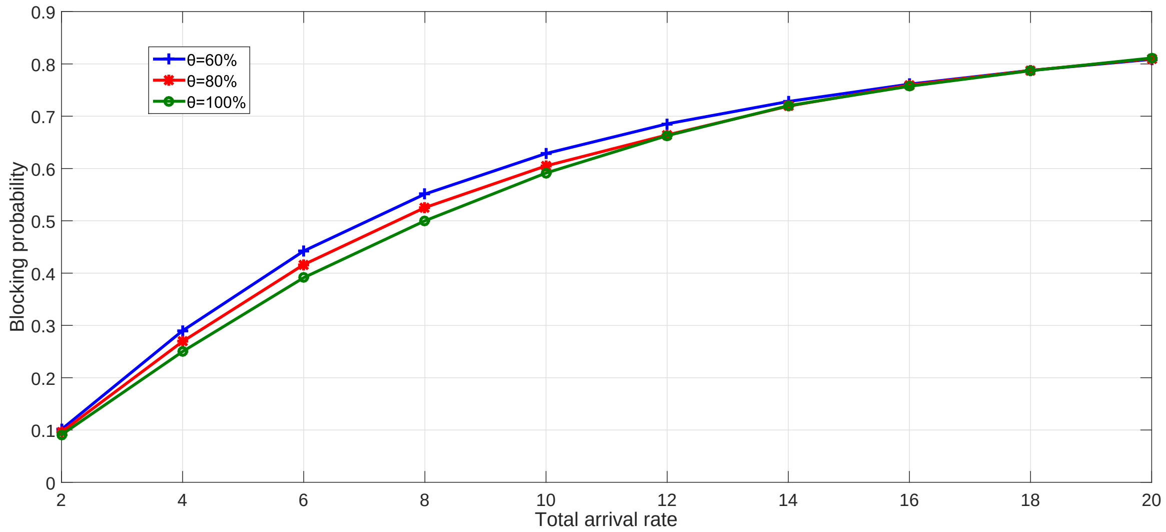

Figure 5.6. Blocking probability against total arrival rate for different θ values. Service rates are 2 bits/symbol/unit time in queue 1 and 1 bit/symbol/unit time in queue 2.

time for queue 2. These results are comparable to the ones presented in [131], thereby validating them. From the results, it can be observed that at low arrival rates, the blocking probability, which is the probability of finding the system full, is very low, implying that the system can effectively service almost all arrivals. The effective arrival rate, or throughput, is thus very high; in fact, close to the total number of arrivals in the system. Again, it can be observed that by gradually increasing arrivals, the effective arrival steadily increases, although at an increasing blocking probability. It therefore implies that by increasing the arrival rate, the system throughput can indeed be increased, albeit at a decreasing rate. Eventually, a highest possible value of the throughput is obtained because of the obvious limitation in server capacity. Also, the results show that by increasing θ, the blocking probability decreases while the throughput increases, signifying an improvement in the overall performance of the network.

In Figs. 5.6 and 5.7, the results of both blocking probability and throughput are likewise presented, only in this instance, the service rates of the servers are 2 bits per symbol per unit time in queue 1 and 1 bit per symbol per unit time in queue 2 respectively. The results show similar trends to those presented in Figs. 5.4 and 5.5, except that at some point, the improvement in performance due to an increase in θ is completely eliminated and rather, a gradual reduction in performance is observed. This is because, as more and more of the demands of queue 2 are moved to queue 1 in the hope of being served more quickly, a tipping point (which invariably corresponds to the optimum θ value) is reached after which, an additional increase in the value of queue 2 demands moved to queue 1 no

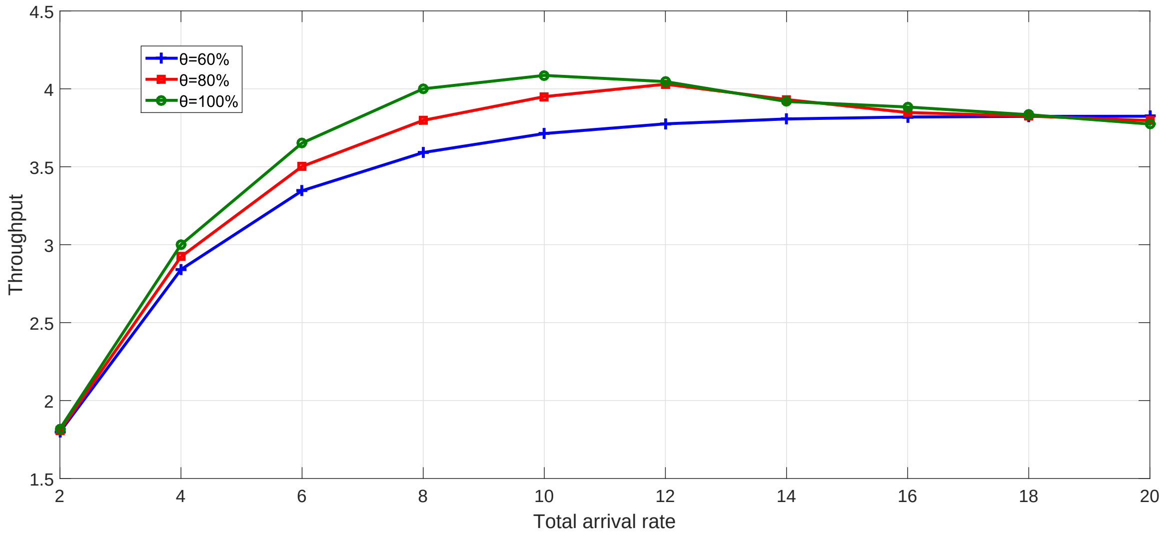

Figure 5.7. Throughput against total arrival rate for different θ values. Service rates are 2 bits/symbol/unit time in queue 1 and 1 bit/symbol/unit time in queue 2.

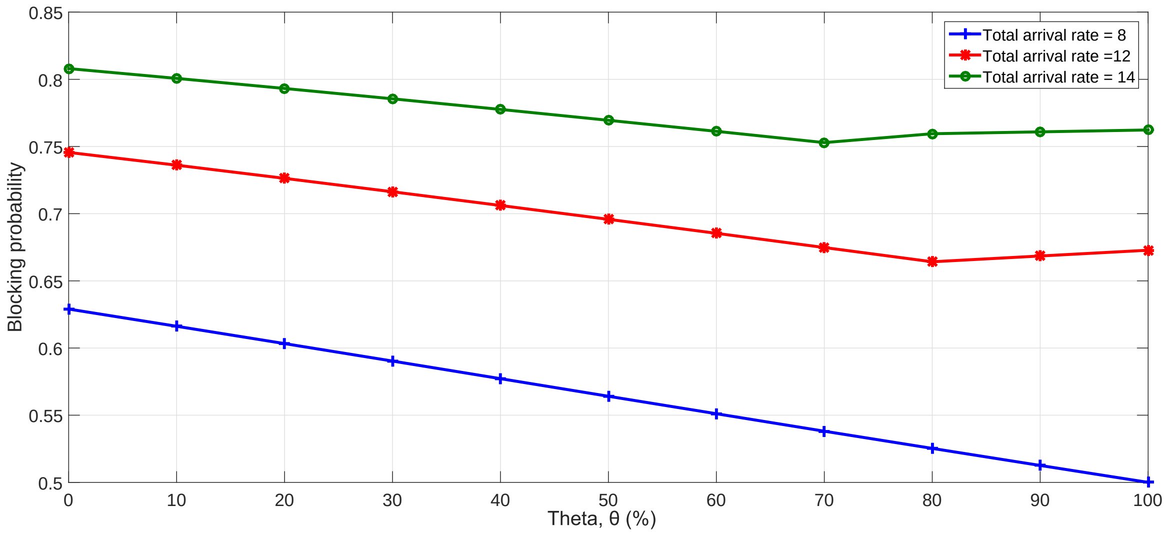

Figure 5.8. Blocking probability against increasing θ values for different total arrival rate. Service rates are 2 bits/symbol/unit time in queue 1 and 1 bit/symbol/unit time in queue 2.

longer improves the overall performance. Rather, the blocking probability increases more significantly, owing to the continuous increase in queue 1, resulting in a decrease in the network throughput as well as a poorer overall network.

It is very significant to observe, from a comparison of the results presented in Figs. 5.4 and 5.5, and

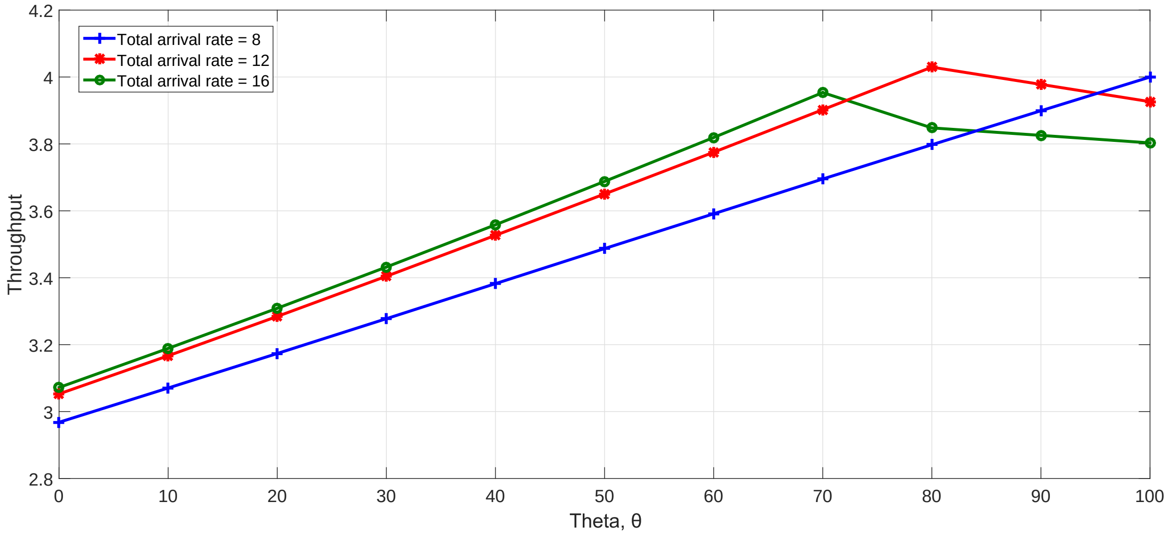

Figure 5.9. Throughput against increasing θ values for different total arrival rate. Service rates are 2 bits/symbol/unit time in queue 1 and 1 bit/symbol/unit time in queue 2.

Figs. 5.6 and 5.7 that, if all the demands of queue 2 are DT and queue 1 has a very high service rate (high enough to conveniently accommodate demands from both queues 1 and 2, as observable in Fig. 5.4), it would be practically unnecessary to have any service at all in queue 2. This is so because, in Figs. 5.4 and 5.5, even at $\theta = 100\%$, that is, after moving the entirety of the demands in queue 2 to queue 1 (with the assumption that all the demands are DT), the throughput performance had still not declined. In comparison with Figs. 5.6 and 5.7, where the service rate in queue 1 is relatively lower, after some optimum value of θ, an increase in its value results in a decline in the overall performance of the system because the capacity of queue 1 has been overstretched and therefore, both the blocking probability as well as the throughput performances begin to depreciate. It is therefore important for any given problem formulation always to find the optimum value of θ, and to allow only that fraction of queue 2 demands to be moved to queue 1 in order to achieve an overall best (optimum) performance for the network.

Figs. 5.8 and 5.9 describe the performance of the blocking probability and the throughput as a function of θ for different values of arrival. From the plots, it is possible to determine the value of the total rewards, and to observe its effects on the overall network performance. According to Fig. 5.9, the total rewards for any given value of θ is simply the difference between the throughput value at the given θ and the throughput value at $\theta = 0\%$, for any value of the arrival rates. Similarly, the optimal value of θ for each of the arrival rates considered can easily be observed. At the total arrival rate of 8 bits per symbol per unit time, an optimal value for θ was never realised because even at $\theta = 100\%$, improvement in performance was still being observed. However, for total arrival rates of 12 and 16 bits per symbol per unit time, an optimal value of θ is realised at the peaks of the plots, after which the performance of the network begins to depreciate. Finally, it should be noted that a change in either the arrival or service rates of the queues will shift the optimum value of θ, either to the right or left of the plots. Therefore, a universal optimum value of θ cannot necessarily be realised, only specific values for specifically developed problem formulations and network parameters (such as arrival rates, service rates) are feasible.

5.6 CONCLUSION

In this chapter, the limiting effects of time delay in RA for CRN has been addressed. In proffering a solution to this limitation, a queueing model was developed and studied. The model achieves RA optimisation in heterogeneous CRN with the consideration that different users can have buffered data in their queues waiting for transmission. In the model, by leveraging the different delay-tolerance profiles of user demands and the mobility of users, a queueing system is developed and analysed that achieves a greater overall capacity in the RA of the network, while still satisfying the varying demands of each of the users or user categories. In the RA model developed, the user demands are classified as either DS or DT, while a fraction θ of the DT demands in the queue farther from the SUBS is moved to the queue nearer to the SUBS for possible faster transmission. This is achievable because, the ring (or equivalently, the queue) nearer to the SUBS transmits at a higher rate, which implies that, its service rate (and thus, its capacity) is higher. The ring can therefore accommodate a larger number of SUs (and their demands) than the ring farther from the SUBS. By changing θ over a given range, its effect on the overall performance of the network is analysed and optimal values for θ, depending on prevalent arrival or/and service conditions, are realised. The results of the various analyses presented show that, with the developed model, an improvement in blocking probability and an optimal throughput can be effectively achieved for the heterogeneous CRN with buffered data.

CHAPTER 6 CONCLUSION

6.1 SUMMARY

The CRN, because of its amazing promise to help solve the spectrum scarcity quagmire, has attracted immense attention and recognition in recent times. With recent developments, CRN may soon take centre-stage in the wireless communication space as the ideal prototype for xG wireless communication and networking. However, for that to ever happen, indisputably, RA in CRN, which describes how CRN will be able to optimise the use of its scarce and limited resources in meeting the needs of its numerous and diverse users, is an integral component to make the CRN's promised possibilities a plausible and palpable reality. In light of this, thorough, in-depth investigations into the essentials and intricacies of RA for CRN is a necessity. Having first identified a knowledge gap in this regard in that not enough work on RA in CRN has been carried out and several limiting factors in the RA solutions have been neglected, the challenge to undertake the research study presented in this thesis emerged. The investigations conducted, as well as the findings presented in this thesis, thus form a cogent, concise and well-coordinated response to the many open-ended problems on RA in CRN, particularly the heterogeneous CRN.

The thesis has been presented in a very logical manner, following all the basic principles of technical reporting within the field of engineering, especially the electronic and telecommunications engineering domain. This concluding part of the thesis summaries the most essential ideas, contributions to knowledge and important findings, as already reported in the various chapters of the thesis write-up. Thereafter, recommendations are made for further improvements or possible future considerations.

Chapter one of the thesis provided a succinct introduction to the entire research work. It importantly established the necessary premises and built a strong pivot around which the entire research work eventually revolved. In the chapter, the actual problem definition was established, which was to find out what the limiting challenges to RA in CRN are, as well as to investigate and develop viable solutions to such problems. Hence, the chapter clearly presented the objectives that the research intended to achieve. The chapter concluded with a list of the contributions made by the extensive work carried out in the research, and gave a list of the publications that have resulted from the work.

In Chapter two, a comprehensive literature study on RA in CRN was conducted. The study delved into the body of knowledge on the subject matter, assessing, classifying and objectively critiquing the various solution models that have been put forth for solving RA problems in CRN. Furthermore, the chapter revealed the critical, limiting areas that hitherto had been neglected in past investigations on RA in CRN and showed the significance of such omissions or commissions to achieving optimality in CRN. The chapter concluded by giving possible ideas on how to address such limitations while providing solution models for RA in CRN.

Chapter three further investigated one of the main problems identified in chapter two - the problem of heterogeneity in the CRN's RA considerations. After painstakingly identifying and classifying the various heterogeneous interpretations applicable to CRN, the chapter developed models that incorporated these various classifications into their design and formulated optimisation problems based on these considerations. The developed problems, as expected, were all complex, NP-hard optimisations problems. However, by a careful study of their structures, it became possible to reformulate them as ILP problems, and to solve them using the BnB method for solving such ILP problems. The chapter also investigated the concept, implications and application of weights on the overall network performance. This was achieved by attaching a weight factor to the various categories of heterogeneous users and analysing the effects generated by such considerations.

In Chapter four, another important problem in RA for CRN, that is, the limitation due to the stringent constraint of low level of permissible interference to PUs, which has arguably been the most limiting factor to the productivity of CRN, was addressed through the use of cooperative diversity. With cooperation, it became possible to meaningfully engage almost all the subchannels in the SUs network in transmitting data, irrespective of their possible interference levels to PUs. The cooperative diversity mechanism employed, that is, the single best relay-selection scheme, helped in achieving a much greater productivity for the CRN at a slightly higher overhead, even when some of the PUs have extremely sensitive interference power requirements.

Chapter five essentially presents a solution model to another significant problem - the problem of time delays in RA for CRN. User classifications brought about the reality that certain categories of heterogeneous users can have a higher level of delay tolerance than some other categories of users. The chapter applied queueing theory in addressing the delay problem, therefore making it possible to optimise the capacities of RA in CRN. By showing that a fraction of the delay-tolerable users can be moved between queues, and also being able to obtain an optimal value for that fraction, a much greater improvement in the CRN's capabilities was realised.

In this concluding part of the thesis, the recommendations for further or/and future works on the research area are next provided.

6.2 RECOMMENDATIONS FOR FUTURE WORK

Although a whole lot of work has been carried out and several contributions made to the body of knowledge on CRN, it is significant to say, very evidently, that there is still some more work that can be done in this field. The following are therefore being recommended for further or future possible considerations by interested researchers in the field. The recommendations are made based on the identified problems that were addressed in chapters three, four and five of the thesis.

6.2.1 Recommendations based on heterogeneous considerations

- Of the numerous optimisation approaches described in chapter two, only a few of them have been explored in this thesis in investigating solutions to RA in CRN, while at the same time addressing the limiting problem of heterogeneity. Although each of the optimisation approaches has its own pros and cons, it would be an exciting research work, for instance, to investigate and compare methods like the Lagrangian duality or column generation with the ILP approach and heuristic that have been mainly employed in the thesis.

- An equally important point for consideration is that all the problem formulations investigated in the thesis have developed their RA problems in CRN using the underlay architecture. While the methods and solution approaches developed are indeed transferable to the overlay and hybrid architectures, an interesting research work would be to actually carry that out - to transfer the solution models that have been proffered for RA in heterogeneous CRN into the overlay or even hybrid architectures. This will inevitably add a few more constraints to the optimisation problems, and it would be interesting to see how the solutions are either in line with, or deviate from the current ones provided for the underlay architecture.

- In the HetNet consideration of heterogeneity in CRN, a good further work would be to actually develop the secondary networks as probably femtocells, picocells or any other variant of the small cells, with the primary network developed as a macrocell. While the analyses carried out in this thesis have been generalised for all HetNet considerations, a more specific formulation may possibly result in slight variations in the realisations and it would be a good idea to investigate how such variations can influence the CRN's performance, as well as their effects on the overall CRN networking.

6.2.2 Recommendations based on the use of cooperative diversity for mitigating interference to PUs

- Cooperative diversity was employed in chapter four to address the problem of limitations due to interference to PUs. Cooperative diversity is indeed a diverse and broad research area. However, only one cooperative technique, which was presumed to be the best technique for adoption, was developed and employed in the thesis. It is recommended to investigate several other types/mechanisms of cooperation/relaying so as to help in achieving and establishing the best cooperative diversity scheme/approach for RA in CRN. This would indeed be an exciting research focus.

- An important point for consideration while employing cooperation to mitigate the limitations in RA for heterogeneous CRN is to develop models whereby the chosen cooperator or relay can transmit at the same time its own data, alongside the data from the SU, in order to improve overall network capability. In the model developed and analysed in the thesis, it was assumed that the SU selected as the cooperator has no data of its own to transmit. While this is very possible, it may not happen all the time. Developing and analysing a model whereby the cooperator can transmit both its own data and those from the source SU simultaneously might give even better results, and is thereby recommended to be investigated.

6.2.3 Recommendations based on the use of queueing models for addressing delay problems

- For the queueing model developed to help solve the RA problem while considering time delay limitations, the network characterisation has been assumed to be in continuous-time, and therefore, a CTMC was developed and used in analysing and solving the problem. A probably more demanding but more realistic consideration of the problem, would be to assume that the CRN characterisation is in discrete-time. This would imply that, the queueing consideration would have to be a discrete-time Markov chain, and investigating it as such would indeed be an interesting challenge.

- Another possible future work would be to assume the queues generated in the RA problem for the buffered, heterogeneous CRN are infinite, as against the finite buffer assumption discussed in chapter five. If that is so, then it may be possible that the queues generated can be analysed using the matrix analytic approach of Neuts [132], and the solutions provided can be compared with those obtained using the GTH state reduction approach, as developed and employed in this work.

REFERENCES

[1] A. Goldsmith, *Wireless Communications*. New York, USA: Cambridge University Press, 2005.

[2] Statista, "Number of mobile phone users worldwide from 2013 to 2019," 2016, http://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/.

[3] P. Lange, "2012 africa - mobile broadband, data and mobile media market," Sept. 2012, http://www.budde.com.au/Research/2012-Africa-Mobile-Broadband-Data-and-Mobile-Media-Market.html.

[4] S. Haykin and P. Setoodeh, "Cognitive radio networks: The spectrum supply chain paradigm," *IEEE Trans. Cogn. Commun. Netw*, vol. 1, no. 1, pp. 3–28, Mar. 2015.

[5] F. C. Commission, "Report of spectrum efficiency working group," 2002, spectrum policy task force, Washington, DC, USA.

[6] ——, "Cognitive radio technologies proceeding," 2003, rep. ET Docket, no. 03-108.

[7] V. Valenta, R. Marsalek, G. Baudoin, M. Villegas, M. Suarez, and F. Robert, "Survey on spectrum utilization in europe: Measurements, analyses and observations," in *Proc. 5th Int. Conf. on CROWNCOM*, June 2010, pp. 1–5.

[8] C. Tran, R. Lu, A. Ramirez, C. Phillips, and S. Thai, "Dynamic spectrum access: Architectures and implications," in *Proc. IEEE MILCOM*, Nov. 2008, pp. 1–7.

[9] S. Filin, H. Harada, and M. Hasegawa, "Performance evaluation of dynamic spectrum assignment and access technologies," in *Proc. IEEE 19th Int. Symposium on PIMRC*, Sept. 2008, pp. 1–5.

[10] J. Pastircak, J. Gazda, and D. Kocur, "A survey on the spectrum trading in dynamic spectrum access networks," in *Proc. 56th Int. Symposium on ELMAR*, Sept. 2014, pp. 1–4.

[11] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radios," Ph.D. dissertation, KTH, Sweden, 2000.

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

[13] L. E. Doyle, *Essentials of Cognitive Radio*, ser. The Cambridge Wireless Essentials Series. New York, USA: Cambridge University Press, 2009.

[14] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *The Int. J. Comput. Telecommun. Netw.*, vol. 50, no. 13, pp. 2127 – 2159, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128606001009

[15] J. Lofberg, "YALMIP : a toolbox for modeling and optimization in MATLAB," in *Proc. IEEE Int. Symposium on Computer Aided Control Systems Design*, Sept. 2004, pp. 284–289.

[16] B. A. Fette, *Cognitive Radio Technology*, ser. Communications Engineering Series. Burlington, MA 01803, USA: Newness (Elsevier) Publications, 2006.

[17] B. Wang and K. Liu, "Advances in cognitive radio networks: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 5–23, Feb. 2011.

[18] Y.-C. Liang, K.-C. Chen, G. Li, and P. Mahonen, "Cognitive radio networking and communications: an overview," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3386–3407, Sept. 2011.

[19] X. Liu, Y. Zhang, Y. Li, Z. Zhang, and K. Long, "A survey of cognitive radio technologies and their optimization approaches," in *Proc. 8th Int. Conf. on CHINACOM*, Aug. 2013, pp. 973–978.

[20] B. Fette, "Fourteen years of cognitive radio development," in *Proc. IEEE MILCOM*, Nov 2013, pp. 1166–1175.

[21] K.-C. Chen, Y.-J. Peng, N. Prasad, Y.-C. Liang, and S. Sun, "Cognitive radio network architecture: Part I – general structure," in *Proc. 2nd Int. Conf. on UIMC*. New York, NY, USA: ACM, 2008, pp. 114–119. [Online]. Available: http://doi.acm.org/10.1145/1352793.1352817

[22] ——, "Cognitive radio network architecture: Part II – trusted network layer structure," in *Proc. 2nd Int. Conf. on UIMC*. New York, NY, USA: ACM, 2008, pp. 120–124. [Online]. Available: http://doi.acm.org/10.1145/1352793.1352818

[23] C. Xin and X. Cao, "A cognitive radio network architecture without control channel," in *Proc. IEEE GLOBECOM*, Nov. 2009, pp. 1–6.

[24] D. Xu, Q. Zhang, Y. Liu, Y. Xu, and P. Zhang, "An architecture for cognitive radio networks with cognition, self-organization and reconfiguration capabilities," in *Proc. IEEE VTC (Fall)*, Sept. 2012, pp. 1–5.

[25] A. Amanna and J. Reed, "Survey of cognitive radio architectures," in *Proc. IEEE SoutheastCon*, Mar. 2010, pp. 292–297.

[26] M. Monemi, M. Rasti, and E. Hossain, "Characterizing feasible interference region for underlay cognitive radio networks," in *Proc. IEEE ICC*, June 2015, pp. 7603–7608.

[27] W. Guo and X. Huang, "Maximizing throughput for overlaid cognitive radio networks," in *Proc. IEEE MILCOM*, Oct. 2009, pp. 1–7.

[28] W.-L. Chin and J.-M. Lee, "Spectrum sensing scheme for overlay cognitive radio networks," *IET Electron. Lett.*, vol. 51, no. 19, pp. 1552–1554, 2015.

[29] S. Senthuran, A. Anpalagan, and O. Das, "Throughput analysis of opportunistic access strategies in hybrid underlay overlay cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2024–2035, June 2012.

[30] J. Lai, E. Dutkiewicz, R. P. Liu, and R. Vesilo, "Comparison of cooperative spectrum sensing strategies in distributed cognitive radio networks," in *Proc. IEEE GLOBECOM*, Dec. 2012, pp. 1513–1518.

[31] M. Nabil, W. El-Sayed, and M. Elnainay, "A cooperative spectrum sensing scheme based on task assignment algorithm for cognitive radio networks," in *Proc. Int. Conf. on WCMC*, Aug. 2014, pp. 151–156.

[32] H. Xu and B. Li, "Efficient resource allocation with flexible channel cooperation in OFDMA cognitive radio networks," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.

[33] ——, "Resource allocation with flexible channel cooperation in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 5, pp. 957–970, May 2013.

[34] M. H. Hassan and M. Hossain, "Cooperative beamforming for cognitive radio systems with asynchronous interference to primary user," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5468–5479, Nov. 2013.

[35] Z. Mao and X. Wang, "Efficient optimal and suboptimal radio resource allocation in OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 440–445, Feb. 2008.

[36] C. Turgu and C. Toker, "A low complexity resource allocation algorithm for OFDMA systems," in *Proc. 15th IEEE Workshop on SSP*, Aug. 2009, pp. 689–692.

[37] C. Shi, Y. Wang, and P. Zhang, "Joint spectrum sensing and resource allocation for multi-band cognitive radio systems with heterogeneous services," in *Proc. IEEE GLOBECOM*, December 2012, pp. 1180–1185.

[38] X. Yu, T. Lv, P. Chang, and Y. Li, "Enhanced efficient optimal and suboptimal radio resource allocation in OFDMA system," in *Proc. 6th Int. Conf. on WiCOM*, Sept. 2010, pp. 1–4.

[39] S. Kim, B. G. Lee, and D. Park, "Energy-per-bit minimized radio resource allocation in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1862–1873, Apr. 2014.

[40] S. Bashar and Z. Ding, "Admission control and resource allocation in a heterogeneous OFDMA wireless network," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4200–4210, Aug. 2009.

[41] T. Villa, R. Merz, and R. Knopp, "Dynamic resource allocation in heterogeneous networks," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 1915–1920.

[42] E. B. Rodrigues and F. Casadevall, "Rate adaptive resource allocation with fairness control for OFDMA networks," in *Proc. 18th EW Conf.*, Apr. 2012, pp. 1–8.

[43] M. Fang and G. Song, "Adaptive resource allocation schemes for OFDMA systems with proportional rate constraint," in *Proc. Symposium on CIICT*, July 2012, pp. 106–110.

[44] S. Cicalo and V. Tralli, "Adaptive resource allocation with proportional rate constraints for uplink SC-FDMA systems," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1419–1422, Aug. 2014.

[45] H. Liming and X. Lin, "Margin adaptive resource allocation with long-term rate fairness considered in downlink OFDMA systems," in *Proc. IEEE EUROCON*, May 2009, pp. 1919–1923.

[46] N. Ul Hassan and M. Assaad, "Low complexity margin adaptive resource allocation in downlink MIMO-OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3365–3371, July 2009.

[47] M. Pischella and J.-C. Belfiore, "Distributed margin adaptive resource allocation in MIMO OFDMA networks," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2371–2380, Aug. 2010.

[48] B. S. Awoyemi, B. T. Maharaj, and A. S. Alfa, "QoS provisioning in heterogeneous cognitive radio networks through dynamic resource allocation," in *Proc. IEEE AFRICON*, Sept. 2015, pp. 1–6.

[49] J.-C. Liang and J.-C. Chen, "Resource allocation in cognitive radio relay networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 476–488, Mar. 2013.

[50] Y. Tachwali, F. Basma, and H. Refai, "Cognitive radio architecture for rapidly deployable heterogeneous wireless networks," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1426–1432, Aug. 2010.

[51] W. L. Winston and M. Venkataramanan, *Introduction to Mathematical Programming*, 4th ed. Pacific Grove, CA.; London: Thompson Brooks/Cole, 2003.

[52] P. Pedregal, *Introduction to Optimization*, ser. Texts in Applied Mathematics. New York, USA: Springer-Verlag, 2004.

[53] K. Edwin and H. Stanislaw, *An Introduction to Optimization*, 4th ed., ser. Wiley Series in Discrete Mathematics and Optimization. West Sussex UK: John Wiley and Sons, Inc., 2013.

[54] S. Boyd and L. Vandenberghe, *Convex Optimization*, ser. Berichte über verteilte messysteme. Cambridge University Press, 2004. [Online]. Available: https://books.google.co.za/books?id=mYm0bLd3fcoC

[55] M. G. Adian, H. Aghaeinia, and Y. Norouzi, "Optimal resource allocation for opportunistic spectrum access in heterogeneous MIMO cognitive radio networks," *Trans. Emerg. Telecommun. Technol.*, pp. n/a–n/a, 2014. [Online]. Available: http://dx.doi.org/10.1002/ett.2796

[56] M. G. Adian and H. Aghaeinia, "Optimal resource allocation in heterogeneous MIMO cognitive radio networks," *Wireless Personal Commun.*, vol. 76, no. 1, pp. 23–39, 2014. [Online]. Available: http://dx.doi.org/10.1007/s11277-013-1486-0

[57] M. Adian and H. Aghaeinia, “Optimal resource allocation for opportunistic spectrum access in multiple-input multiple-output-orthogonal frequency division multiplexing based cooperative cognitive radio networks,” *IET Signal Processing*, vol. 7, no. 7, pp. 549–557, Sept. 2013.

[58] ——, “Optimal and sub-optimal resource allocation in multiple-input multiple-output-orthogonal frequency division multiplexing-based multi-relay cooperative cognitive radio networks,” *IET Commun.*, vol. 8, no. 5, pp. 646–657, Mar. 2014.

[59] S. Wang, M. Ge, and C. Wang, “Efficient resource allocation for cognitive radio networks with cooperative relays,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2432–2441, Nov. 2013.

[60] S. Wang, Z.-H. Zhou, M. Ge, and C. Wang, “Resource allocation for heterogeneous cognitive radio networks with imperfect spectrum sensing,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 464–475, Mar. 2013.

[61] M. Ge and S. Wang, “On the resource allocation for multi-relay cognitive radio systems,” in *Proc. IEEE ICC*, June 2014, pp. 1591–1595.

[62] R. Xie, F. Yu, and H. Ji, “Dynamic resource allocation for heterogeneous services in cognitive radio networks with imperfect channel sensing,” *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 770–780, Feb. 2012.

[63] R. Xie, F. Yu, H. Ji, and Y. Li, “Energy-efficient resource allocation for heterogeneous cognitive radio networks with femtocells,” *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3910–3920, Nov. 2012.

[64] R. Xie, F. Yu, and H. Ji, “Spectrum sharing and resource allocation for energy-efficient heterogeneous cognitive radio networks with femtocells,” in *Proc. IEEE ICC*, June 2012, pp. 1661–1665.

[65] Y. Rahulamathavan, S. Lambotharan, C. Toker, and A. Gershman, “Suboptimal recursive optimisation framework for adaptive resource allocation in spectrum-sharing networks,” *IET Signal Process.*, vol. 6, no. 1, pp. 27–33, Feb. 2012.

[66] Y. Rahulamathavan, K. Cumanan, L. Musavian, and S. Lambotharan, "Optimal subcarrier and bit allocation techniques for cognitive radio networks using integer linear programming," in *Proc. 15th IEEE Workshop on SSP*, Aug. 2009, pp. 293–296.

[67] Y. Rahulamathavan, K. Cumanan, and S. Lambotharan, "Optimal resource allocation techniques for MIMO-OFDMA based cognitive radio networks using integer linear programming," in *Proc. 11th IEEE Int. Workshop on SPAWC*, June 2010, pp. 1–5.

[68] Y. Rahulamathavan, K. Cumanan, R. Krishna, and S. Lambotharan, "Adaptive subcarrier and bit allocation techniques for MIMO-OFDMA based uplink cognitive radio networks," in *Proc. 1st Int. Workshop on UKIWCWS*, Dec. 2009, pp. 1–5.

[69] A. Zafar, M.-S. Alouini, Y. Chen, and R. Radaydeh, "New resource allocation scheme for cognitive relay networks with opportunistic access," in *Proc. IEEE ICC*, June 2012, pp. 5603–5607.

[70] R. W. Floyd, "Nondeterministic algorithms," *J. ACM*, vol. 14, no. 4, pp. 636–644, Oct. 1967. [Online]. Available: http://doi.acm.org/10.1145/321420.321422

[71] O. Tripp, E. Koskinen, and M. Sagiv, "Turning nondeterminism into parallelism," *SIGPLAN Not.*, vol. 48, no. 10, pp. 589–604, Oct. 2013. [Online]. Available: http://doi.acm.org/10.1145/2544173.2509533

[72] G. Zhao, J. Li, K. Lee, and J. B. Song, "Optimal frequency-time allocation in cognitive radio wireless mesh networks," *IETE Technical Review*, vol. 28, no. 5, pp. 434–444, 2011. [Online]. Available: http://www.tandfonline.com/doi/abs/10.4103/0256-4602.85976

[73] A. El Shafie, A. Sultan, and T. Khattab, "Band allocation for cognitive radios with buffered primary and secondary users," in *Proc. IEEE WCNC*, Apr. 2014, pp. 1508–1513.

[74] F. Wang and W. Wang, "Robust beamforming and power control for multiuser cognitive radio network," in *Proc. IEEE GLOBECOM*, Dec. 2010, pp. 1–5.

[75] Z.-Q. Luo and W. Yu, “An introduction to convex optimization for communications and signal processing,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.

[76] M.-S. Cheon, S. Ahmed, and F. Al-Khayyal, “A branch-reduce-cut algorithm for the global optimization of probabilistically constrained linear programs,” *Mathematical Programming*, vol. 108, no. 2, pp. 617–634, 2006. [Online]. Available: http://dx.doi.org/10.1007/s10107-006-0725-5

[77] M. Z. Bocus, J. P. Coon, N. C. Canagarajah, J. P. McGeehan, S. M. D. Armour, and A. Doufexi, “Resource allocation for OFDMA-based cognitive radio networks with application to h.264 scalable video transmission,” *EURASIP J. Wireless Commun. Netw.*, vol. 2011, no. 1, p. 245673, 2011. [Online]. Available: http://jwcn.eurasipjournals.com/content/2011/1/245673

[78] L. Wang, W. Xu, Z. He, and J. Lin, “Algorithms for optimal resource allocation in heterogeneous cognitive radio networks,” in *Proc. 2nd Int. Conf. on PEITS*, vol. 2, Dec. 2009, pp. 396–400.

[79] S. Du, F. Huang, and S. Wang, “Power allocation for orthogonal frequency division multiplexing-based cognitive radio networks with cooperative relays,” *IET Commun.*, vol. 8, no. 6, pp. 921–929, Apr. 2014.

[80] W.-C. Pao and Y.-F. Chen, “Adaptive gradient-based methods for adaptive power allocation in OFDM-based cognitive radio networks,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 836–848, Feb. 2014.

[81] J. Zhang, Z. Zhang, H. Luo, and A. Huang, “A column generation approach for spectrum allocation in cognitive wireless mesh network,” in *Proc. IEEE GLOBECOM*, Nov. 2008, pp. 1–5.

[82] Z. He, S. Mao, and S. Kompella, “A decomposition approach to quality-driven multiuser video streaming in cellular cognitive radio networks,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 728–739, Jan. 2016.

[83] P. L. Vo, D. N. M. Dang, S. Lee, C. S. Hong, and Q. Le-Trung, “A coalitional

game approach for fractional cooperative caching in content-oriented networks," *Int. J. Comput. Telecomm. Netw.*, vol. 77, no. C, pp. 144–152, Feb. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2014.12.005

[84] L. Zheng and C. W. Tan, "Cognitive radio network duality and algorithms for utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 500–513, March 2013.

[85] Y. Shi and Y. Hou, "A distributed optimization algorithm for multi-hop cognitive radio networks," in *Proc. 27th IEEE Conf. on Computer Communications*, Apr. 2008.

[86] P. Li, S. Guo, W. Zhuang, and B. Ye, "On efficient resource allocation for cognitive and cooperative communications," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 264–273, Feb. 2014.

[87] F. Chen, W. Xu, Y. Guo, J. Lin, and M. Chen, "Resource allocation in OFDM-based heterogeneous cognitive radio networks with imperfect spectrum sensing and guaranteed QoS," in *Proc. 8th Int. Conf. on CHINACOM*, Aug. 2013, pp. 46–51.

[88] M. Hasegawa, H. Hirai, K. Nagano, H. Harada, and K. Aihara, "Optimization for centralized and decentralized cognitive radio networks," *Proc. IEEE*, vol. 102, no. 4, pp. 574–584, Apr. 2014.

[89] P. Mitran, L. B. Le, and C. Rosenberg, "Queue-aware resource allocation for downlink OFDMA cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3100–3111, Oct. 2010.

[90] E. Driouch, W. Ajib, and A. Ben Dhaou, "A greedy spectrum sharing algorithm for cognitive radio networks," in *Proc. ICNC*, Jan. 2012, pp. 1010–1014.

[91] T. Peng, W. Wang, Q. Lu, and W. Wang, "Subcarrier allocation based on water-filling level in OFDMA-based cognitive radio networks," in *Proc. Int. Conf. on WiCom*, Sept. 2007, pp. 196–199.

[92] Y. Liu, L. Liu, and C. Xu, "Spectrum underlay-based water-filling algorithm in cognitive radio networks," in *Proc. Int. Conf. on EICE*, Apr. 2011, pp. 2614–2617.

[93] R. Ujjwal, C. Rai, and N. Prakash, "Fair adaptive resource allocation algorithm for heterogeneous users in OFDMA system," in *Proc. Int. Conf. on SPIN*, Feb. 2014, pp. 402–406.

[94] A. Alshamrani, X. Shen, and L.-L. Xie, "QoS provisioning for heterogeneous services in cooperative cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 819–830, Apr. 2011.

[95] I. Boussaid, J. Lepagnot, and P. Siarry, "A survey on optimization metaheuristics," *Information Sciences*, vol. 237, pp. 82 – 117, 2013, prediction, Control and Diagnosis using Advanced Neural Computations. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025513001588

[96] Y. El Morabit, F. Mrabti, and E. Abarkan, "Spectrum allocation using genetic algorithm in cognitive radio networks," in *Proc. 3rd Int. Workshop on RAWSN*, May 2015, pp. 90–93.

[97] L. Zhu, Y. Xu, J. Chen, and Z. Li, "The design of scheduling algorithm for cognitive radio networks based on genetic algorithm," in *Proc. IEEE ICICT*, Feb. 2015, pp. 459–464.

[98] E. Meshkova, J. Riihijarvi, A. Achtzehn, and P. Mahonen, "Exploring simulated annealing and graphical models for optimization in cognitive wireless networks," in *Proc. IEEE GLOBECOM*, Nov. 2009, pp. 1–8.

[99] B. Ye, M. Nekovee, A. Pervez, and M. Ghavami, "TV white space channel allocation with simulated annealing as meta algorithm," in *Proc. 7th Int. Conf. on CROWNCOM*, June 2012, pp. 175–179.

[100] S. Motiian, M. Aghababaie, and H. Soltanian-Zadeh, "Particle swarm optimization (PSO) of power allocation in cognitive radio systems with interference constraints," in *Proc. 4th IEEE IC-BNMT*, Oct. 2011, pp. 558–562.

[101] S. Ashrafinia, U. Pareek, M. Naeem, and D. Lee, “Binary artificial Bee colony for cooperative relay communication in cognitive radio systems,” in *Proc. IEEE ICC*, June 2012, pp. 1550–1554.

[102] V. Jayaraj, J. Amalraj, and S. Hemalatha, “An analysis of genetic algorithm and tabu search algorithm for channel optimization in cognitive adhoc networks,” *Int. J. Comput. Sci. Mobile Computing*, vol. 3, no. 7, pp. 60–69, July 2014.

[103] R. Meng, Y. Ye, and N. gang Xie, “Multi-objective optimization design methods based on game theory,” in *Proc. 8th Conf. on WCICA*, July 2010, pp. 2220–2227.

[104] D. Nguyen and M. Krunz, “Heterogeneous spectrum sharing with rate demands in cognitive MIMO networks,” in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 3054–3059.

[105] M. Venkatesan and A. Kulkarni, “Soft computing based learning for cognitive radio,” *Int. J. Recent Trends in Eng. Technol.*, vol. 10, no. 1, p. 112, 2014.

[106] E. Shakshuki, M. Younas, A. Ahmed, G. Amel, S. Anis, and M. Abdellatif, “ANT 2012 and MobiWIS 2012 resource allocation for multi-user cognitive radio systems using multi-agent Q-learning,” *Procedia Computer Science*, vol. 10, pp. 46 – 53, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050912003675

[107] Y. Huang, J. Wang, and H. Jiang, “Modeling of learning inference and decision-making engine in cognitive radio,” in *Proc. 2nd Int. Conf. on NSWCTC*, vol. 2, Apr. 2010, pp. 258–261.

[108] G. V. Lakhekar and R. G. Roy, “A fuzzy neural approach for dynamic spectrum allocation in cognitive radio networks,” in *Proc. ICCPCT*, Mar. 2014, pp. 1455–1461.

[109] C.-W. Pyo, X. Zhang, C. Song, M.-T. Zhou, and H. Harada, “A new standard activity in IEEE 802.22 wireless regional area networks: Enhancement for broadband services and monitoring applications in TV whitespace,” in *Proc. 15th Int. Symposium on WPMC*, Sept. 2012, pp. 108–112.

[110] S. Landstrom, A. Furuskar, K. Johansson, L. Falconetti, and F. Kronestedt, "Heterogeneous networks (HetNets) - an approach to increasing cellular capacity and coverage," in *Proc. 15th Int. Symposium on WPMC*, Sept. 2012, pp. 108–112.

[111] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 22–30, June 2011.

[112] M. Kaplan and F. Buzluca, "A dynamic spectrum decision scheme for heterogeneous cognitive radio networks," in *Proc. 24th Int. Symposium on ISCIS*, Sept. 2009, pp. 697–702.

[113] M. Ma and D. H. K. Tsang, "Impact of channel heterogeneity on spectrum sharing in cognitive radio networks," in *Proc. IEEE ICC*, May 2008, pp. 2377–2382.

[114] V. Bhandari and N. H. Vaidya, "Heterogeneous multi-channel wireless networks: Routing and link layer protocols," *SIGMOBILE Mobile Computing and Communs. Review*, vol. 12, no. 1, pp. 43–45, Jan. 2008. [Online]. Available: http://doi.acm.org/10.1145/1374512.1374526

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

[116] K. Pretz, "Overcoming spectrum scarcity - cognitive radio networks might be one answer," 2012, http://theinstitute.ieee.org/technology-focus/technology-topic/overcoming-spectrum-scarcity.

[117] J. Li, T. Luo, and G. Yue, "Resource allocation scheme based on weighted power control in cognitive radio systems," in *Proc. ICCCAS*, vol. 1, Nov. 2013, pp. 178–182.

[118] J. Oh and W. Choi, "A hybrid cognitive radio system: A combination of underlay and overlay approaches," in *Proc. IEEE VTC (Fall)*, Sept. 2010, pp. 1–5.

[119] N. Hao and S.-J. Yoo, "Interference avoidance throughput optimization in cognitive radio ad hoc networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, 2012. [Online]. Available: http://dx.doi.org/10.1186/1687-1499-2012-295

[120] B. S. Awoyemi, B. T. Maharaj, and A. S. Alfa, "Resource allocation for heterogeneous cognitive radio networks," in *Proc. IEEE WCNC*, Mar. 2015, pp. 1759–1763.

[121] B. Awoyemi, *Performance Analysis of Cooperative Diversity in Land Mobile Satellite Systems*. University of KwaZulu-Natal, Durban, 2013. [Online]. Available: http://books.google.co.za/books?id=TGn6oAEACAAJ

[122] B. Awoyemi, T. Walingo, and F. Takawira, "Predictive relay-selection cooperative diversity in land mobile satellite systems," *Int. J. Satellite Commun. Netw.*, vol. 34, no. 2, pp. 277–294, 2016. [Online]. Available: http://dx.doi.org/10.1002/sat.1118

[123] J. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[124] B. Awoyemi, T. Walingo, and F. Takawira, "Relay selection cooperative diversity in land mobile satellite systems," in *Proc. IEEE AFRICON*, Sept. 2013, pp. 1–6.

[125] M. Pischella and D. Le Ruyet, "Cooperative allocation for underlay cognitive radio systems," in *Proc. 14th IEEE Workshop on SPAWC*, June 2013, pp. 245–249.

[126] Y. Zeng, Y.-C. Liang, A. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: Challenges and solutions," *EURASIP J. Advances in Signal Process.*, vol. 2010, no. 1, p. 381465, 2010. [Online]. Available: http://asp.eurasipjournals.com/content/2010/1/381465

[127] A. E. Shafie, A. Sultan, and T. Khattab, "Band allocation for cognitive radios with buffered primary and secondary users," *Computing Research Repository*, vol. abs/1401.0214, 2014.

[128] B. Zhao and V. Friderikos, "A queuing-based delay-tolerant scheme for energy efficiency over cognitive radio networks," in *Proc. IEEE GC Workshops*, Dec 2012, pp. 326–330.

[129] D. Gross, F. Shortie, M. Thompson, and M. Harris, *Fundamentals of Queueing Theory*, ser. Wiley Series in Probability and Statistics. New Jersey, USA: John Wiley and Sons, Inc., 2008.

[130] I. Sonin, "The state reduction and related algorithms and their applications to the study of markov chains, graph theory, and the optimal stopping problem," *Advances in Mathematics*, vol. 145, no. 2, pp. 159 – 188, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001870898918138

[131] W. K. Grassmann, "Optimizing steady state markov chains by state reduction," *European J. Operational Research*, vol. 89, no. 2, pp. 277 – 284, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0377221795002707

[132] M. Neuts, *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*, ser. Algorithmic Approach. Dover Publications, 1981. [Online]. Available: https://books.google.co.za/books?id=WPol7RVptz0C

ADDENDUM A STATE REDUCTION ALGORITHM

A.1 THE GRASSMAN-TAKSAR-HEYMAN (GTH) ALGORITHM

The following Grassmann-Taksar-Heyman (GTH) algorithm is used in obtaining the state probabilities for Markov chains.

The values for g_i, $g_i^{'}$ and $g_i^{''}$, required to obtain the steady state probabilities (and their derivatives), x_i, $x_i^{'}$ and $x_i^{''}$, can be obtained through state reduction as explained below.

Recall that $g_i = x_i G$

Let $\boldsymbol{g} = \boldsymbol{x}G$, hence,

$$\boldsymbol{g}Q = 0,$$

where Q is the generator matrix. Taking first and second derivatives with respect to θ, we have:

$$\boldsymbol{g}^{'}Q = -\boldsymbol{g}Q^{'},$$
$$\boldsymbol{g}^{''}Q = -2\boldsymbol{g}^{'}Q^{'} - \boldsymbol{g}Q^{''}.$$

The set of equations given above can be solved using state reduction to obtain $\boldsymbol{g}$, $\boldsymbol{g}^{'}$ and $\boldsymbol{g}^{''}$. The state reduction algorithm is as follows:

The states are numbered from 0 to M, giving a total of $M+1$ states. M is finite. For each state, we find a steady state equation using $\boldsymbol{g}G = 0$.

For state j, steady state equation is given as;

$$\sum_{i=0}^{M} g_i q_{ij} = 0,$$

where q_{ij} are the elements of the generator matrix Q. By using Gaussian elimination, equations $m+1, m+2, ..., M$ can be used to eliminate $g_{m+1}, g_{m+2}, ..., g_M$ from equations $0, 1, 2, ..., m$ so that a new set of equations is obtained. The new set of equations is represented by:

$$\sum_{i=0}^{m} g_i q_{ij}^m = 0.$$

Grassmann, Taksar and Heyman (GTH) did show that q_{ij}^m can indeed be interpreted as the transition rates of a continuous-time Markov chain, meaning that it is unnecessary to calculate the diagonal elements q_{ii}^m. Specifically, if s_m is defined as $-q_{mm}^m$, then,

$$s_m = -q_{mm}^m = \sum_{i=0}^{m} q_{mj}^m.$$

We can use normal elimination methods to find all q_{ij}^m recursively, starting with $q_{ij}^M = q_{ij}$ and then calculating $q_{ij}^{M-1}, q_{ij}^{M-2}, \ldots, q_{ij}^1$. This gives, for $i, j < m$,

$$q_{ij}^{m-1} = q_{ij}^m - \frac{q_{im}^m q_{mj}^m}{q_{mm}^m}.$$

Substituting for s_m becomes:

$$q_{ij}^{m-1} = q_{ij}^m + \frac{q_{im}^m q_{mj}^m}{s_m}.$$

A more recent approach to state reduction by Grassmann that gives a more convenient and quicker recursion is obtained by making,

$$q_{ij}^m = q_{ij}^{m+1} + \frac{q_{i\,m+1}^{m+1} q_{m+1\,j}^{m+1}}{s_{m+1}} = \left(q_{ij}^{m+2} + \frac{q_{i\,m+2}^{m+2} q_{m+2\,j}^{m+2}}{s_{m+2}} \right) + \frac{q_{i\,m+1}^{m+1} q_{m+1\,j}^{m+1}}{s_{m+1}}.$$

Hence,

$$q_{ij}^m = q_{ij} + \sum_{p=m+1}^{M} \frac{q_{ip}^p q_{pj}^p}{s_p}.$$

For $m > i$, this gives:

$$q_{im}^m = q_{im} + \sum_{p=m+1}^{M} \frac{q_{ip}^p q_{pn}^p}{s_p}.$$

For $m > j$, this gives:

$$q_{mj}^m = q_{mj} + \sum_{p=m+1}^{M} \frac{q_{np}^p q_{pj}^p}{s_p}.$$

b_{ij} is now defined as:

$$b_{ij} = \frac{q_{ij}^j}{s_j}, \quad i < j,$$
$$b_{ij} = q_{ij}^i, \quad i > j.$$

The b_{ij} can be calculated by row, starting with row M then continuing with row $M-1$ and so on. Once the b_{ij} are calculated, normal back-substitution can be used to obtain g_j thus:

Set $g_0 = 1$ and evaluate $g + i = 1, 2, \ldots, M$ as follows;

$$g_j = \sum_{i=0}^{j-1} g_i b_{ij} = b_{0j} + \sum_{i=1}^{j-1} g_i b_{ij}, \quad j > 0.$$

The equations for $g_i^{'}$ and $g_i^{''}$ do have the same structure as the equations for g_i. This can be verified by using the fact that $g_0 = 1$ and writing $\sum_{i=0}^{M} g_i q_{ij} = 0$ as follows;

$$q_{0j} + \sum_{i=1}^{M} g_i q_{ij} = 0, \quad j \geq 0.$$

Since $g_0 = 1$, $g_0^{'}$ must be 0. Also, the jth equation obtained from $\boldsymbol{g}^{'} Q = -\boldsymbol{g} Q^{'}$ has the jth element of $\boldsymbol{g} Q^{'}$ as its constant term. If this constant term is denoted as q_{0j}^{*}, then,

$$q_{0j}^{*} + \sum_{i=1}^{M} g_i^{'} q_{ij} = 0, \quad j \geq 0.$$

The last two equations are similar, except that all q_{0j} are replaced by q_{0j}^{*}. Consequently, if b_{ij}^{*} are the coefficients obtained by eliminating the $g_i^{'}$ from $\boldsymbol{g}^{'} Q = -\boldsymbol{g} Q^{'}$, $b_{ij}^{*} = b_{ij}$ except for $i = 0$. From the algorithm given above,

$$b_{0j}^{*} = \frac{\left(q_{0j}^{*} + \sum_{p=j+1}^{M} b_{ip} b_{pj} \right)}{s_j}, \quad j = M, M-1, ..., 1.$$

Again using the fact that $g_0^{'} = 0$,

$$g_j^{'} = b_{0j}^{*} + \sum_{i=1}^{j-1} g_i^{'} b_{ij}.$$

Similarly, if $q_{0j}^{**} = (2\boldsymbol{g}^{'} Q^{'} + \boldsymbol{g} Q^{''})_j$, b_{0j}^{**} can be obtained by a similar representation to b_{0j}^{*} and using back-substitution to find $g_j^{''}$ from the equation,

$$g_j^{''} = b_{0j}^{**} + \sum_{i=1}^{j-1} g_i^{''} b_{ij}.$$