Radio Resource Management for Wireless Mesh Networks Supporting Heterogeneous Traffic

by

Ho Ting Cheng

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Wireless mesh networking has emerged as a promising technology for future broadband wireless access, providing a viable and economical solution for both peer-to-peer applications and Internet access. The success of wireless mesh networks (WMNs) is highly contingent on effective radio resource management. In conventional wireless networks, system throughput is usually a common performance metric. However, next-generation broadband wireless access networks including WMNs are anticipated to support multimedia traffic (e.g., voice, video, and data traffic). With heterogeneous traffic, quality-of-service (QoS) provisioning and fairness support are also imperative. Recently, wireless mesh networking for suburban/rural residential areas has been attracting a plethora of attentions from industry and academia. With austere suburban and rural networking environments, multi-hop communications with decentralized resource allocation are preferred. In WMNs without powerful centralized control, simple yet effective resource allocation approaches are desired for the sake of system performance melioration. In this dissertation, we conduct a comprehensive research study on the topic of radio resource management for WMNs supporting multimedia traffic. In specific, this dissertation is intended to shed light on how to effectively and efficiently manage a WMN for suburban/rural residential areas, provide users with high-speed wireless access, support the QoS of multimedia applications, and improve spectrum utilization by means of novel radio resource allocation. As such, five important resource allocation problems for WMNs are addressed, and our research accomplishments are briefly outlined as follows:

Firstly, we propose a novel node clustering algorithm with effective subcarrier allocation for WMNs. The proposed node clustering algorithm is QoS-aware, and the subcarrier allocation is optimality-driven and can be performed in a decentralized manner. Simulation results show that, compared to a conventional conflict-graph approach, our proposed

approach effectively fosters frequency reuse, thereby improving system performance;

Secondly, we propose three approaches for joint power-frequency-time resource allocation. Simulation results show that all of the proposed approaches are effective in provisioning packet-level QoS over their conventional resource allocation counterparts. Our proposed approaches are of low complexity, leading to preferred candidates for practical implementation;

Thirdly, to further enhance system performance, we propose two low-complexity node cooperative resource allocation approaches for WMNs with partner selection/allocation. Simulation results show that, with beneficial node cooperation, both proposed approaches are promising in supporting QoS and elevating system throughput over their non-cooperative counterparts;

Fourthly, to further utilize the temporarily available radio spectrum, we propose a simple channel sensing order for unlicensed secondary users. By sensing the channels according to the descending order of their achievable rates, we prove that a secondary user should stop at the first sensed free channel for the sake of optimality; and

Lastly, we derive a unified optimization framework to effectively attain different degrees of performance tradeoff between throughput and fairness with QoS support. By introducing a bargaining floor, the optimal tradeoff curve between system throughput and fairness can be obtained by solving the proposed optimization problem iteratively.

Acknowledgements

First and foremost, I would like to express my utmost admiration and deepest gratitude to Professor Weihua Zhuang. As my Ph.D. advisor, Professor Zhuang is very inspirational and supportive. She has provided me with useful weekly guidance, constant encouragement, and constructive criticism over the course of my Ph.D. study. I shall never forget the informative discussions with her. Working with Professor Zhuang is and has always been a very enlightening and rewarding experience for me.

I would also like to express my appreciation to Professor Vojislav B. Mišić, Professor Liping Fu, Professor Liang-Liang Xie, and Professor Catherine H. Gebotys for serving on my thesis advisory committee. Thanks also go to Professor Zhou Wang who served as the delegate for my Ph.D. defense. I appreciate their time and effort devoted to reading my Ph.D. dissertation and providing me with their insightful comments and invaluable suggestions, thereby further improving the quality of my research work.

I am indebted to many individuals. In particular, I would like to thank Professor Xuemin (Sherman) Shen and Professor Jon W. Mark from the Center for Wireless Communications (CWC) for providing us with a top-notch research lab and a pleasant academic environment. I am also grateful to the members of the Broadband and Communications Research (BBCR) group for their assistance. Special thanks go to Dr. Ping Wang, Dr. Wei Song, Dr. Kuang-Hao (Stanley) Liu, Dr. Minghui Shi, Dr. Atef Abdrabou, Dr. Mohamad Awad, and Albert Wasef for their support, generosity, and friendship.

Thanks also go to the administrative staff, Ms. Wendy Boles, Ms. Lisa Hendel, Ms. Annette Dietrich, and Ms. Karen Schooley for their help, kindness, patience, and cooperation.

Finally, I would like to thank my parents for being an immense source of inspiration and support throughput my life.

This dissertation is dedicated to my parents

for their love, endless support, inspiration, and encouragement.

Contents

List of Tables **xviii**

List of Figures **xx**

1 Introduction **1**

1.1 Overview of Wireless Mesh Networks 1

1.2 Resource Management for Wireless Mesh Networks 4

1.3 Motivations and Objectives 8

1.4 Research Contributions 12

1.5 Outline of the Dissertation 15

2 System Model **17**

2.1 Network Architecture 17

2.2 Traffic Traversal over WMNs 20

2.2.1 Traffic Characterization 20

2.2.2 Traffic Model 21

2.2.3 Traffic Differentiation 22

2.3 Medium Access with QoS Provisioning 24
2.4 Performance Metrics 25
2.5 Chapter Summary 26

3 QoS-Aware Node Clustering and Subcarrier Allocation 27

3.1 Introduction 27
3.2 Related Work 29
3.3 Joint QoS-Aware Node Clustering and Tax-Based Subcarrier Allocation . . . 31
3.3.1 Neighbor Discovery 32
3.3.2 Initial Path Establishment 32
3.3.3 Traffic Load Estimation 33
3.3.4 Node Clustering Algorithm 35
3.3.5 Tax-Based Subcarrier Allocation 37
3.3.6 Complexity Analysis 40
3.4 Efficiency Evaluation by Game Theory 40
3.5 Performance Evaluation 42
3.5.1 Simulation Environment 42
3.5.2 QoS-Aware Node Clustering 44
3.5.3 Tax-Based Subcarrier Allocation 46
3.6 Discussion 56
3.7 Chapter Summary 56

4 Joint Power-Subcarrier-Time Resource Allocation with Effective QoS Provisioning 58

4.1 Introduction . . . 58

4.2 Related Work . . . 60

4.3 Joint Power-Subcarrier-Time Resource Allocation . . . 61

4.4 Proposed Resource Allocation Approaches . . . 64

4.4.1 KKT-Driven Resource Allocation . . . 64

4.4.2 GA-Based Resource Allocation . . . 66

4.4.3 Hybrid Resource Allocation: Combined-KKT-GA . . . 71

4.4.4 Complexity Analysis . . . 73

4.5 Performance Evaluation . . . 74

4.5.1 Simulation Environment . . . 74

4.5.2 Simulation Results . . . 76

4.6 Discussion . . . 85

4.7 Chapter Summary . . . 86

5 Performance Enhancement I: Non-Altruistic Node Cooperation 87

5.1 Introduction . . . 87

5.2 Related Work . . . 89

5.3 Non-Altruistic Node Cooperation with Regenerative Nodes . . . 92

5.4 QoS-Driven Non-Altruistic Node Cooperative Resource Allocation . . . 96

5.5 Proposed Resource Allocation Approaches with Node Cooperation . . . 100

5.5.1 KKT Interpretations . . . 100
5.5.2 Examples . . . 103
5.5.3 Proposed Approach with Centralized Control . . . 105
5.5.4 Proposed Approach with Distributed Control . . . 112
5.5.5 Complexity Analysis . . . 117
5.6 Efficiency Evaluation by Game Theory . . . 120
5.6.1 Nash Equilibrium . . . 120
5.6.2 Pareto Optimality . . . 121
5.7 Discussion . . . 122
5.8 Performance Evaluation . . . 124
5.8.1 Simulation Environment . . . 124
5.8.2 Simulation Results . . . 126
5.9 Chapter Summary . . . 138

6 Performance Enhancement II: Instinctive Channel Sensing for Cognitive Radio 140

6.1 Introduction . . . 140
6.2 Related Work . . . 142
6.3 Cognitive Radio Network Model . . . 143
6.4 Proposed Stopping Rule and Channel Exploitation . . . 145
6.4.1 Single-Secondary User Scenario ($M = 1$) . . . 145
6.4.2 Two-Secondary User Scenario ($M = 2$) . . . 148

6.4.3 Multi-Secondary User Scenario ($M > 2$) 150
6.5 Performance Evaluation 151
6.5.1 Simulation Environment 151
6.5.2 Simulation and Numerical Results 151
6.6 Discussion 165
6.7 Chapter Summary 167

7 Balancing Throughput and Fairness with QoS Assurance 168

7.1 Introduction 168
7.2 Related Work 169
7.3 Balancing Throughput and Fairness 171
7.3.1 Interference-Limited Network Model 171
7.3.2 Optimization Problem Formulation 173
7.4 Efficiency Evaluation by Game Theory 178
7.5 Numerical Results 178
7.5.1 Equal Weighting Factors 180
7.5.2 Different Weighting Factors 184
7.6 Discussion 188
7.7 Chapter Summary 189

8 Conclusions and Further Research 191

8.1 Conclusions 191
8.2 Further Research 194

Appendices **196**

A Supplementary **196**

A.1 Tax Interpretation of the KKT Conditions in Chapter 3 196

A.2 Upper Bound for Throughput in Chapter 4 198

A.3 Details on Non-Altruistic Node Cooperation in Chapter 5 199

A.3.1 AF Mode of Cooperation 200

A.3.2 DF Mode of Cooperation 201

A.3.3 Outage Performance 202

A.3.4 Relay Selection 205

A.3.5 Cooperation versus Non-Cooperation 207

A.3.6 Numerical Results 210

B Proofs of Propositions, Corollaries, Lemmas, and Theorems **212**

C List of Acronyms **229**

Bibliography **232**

List of Figures

1.1 An illustration of wireless mesh networking. 3

1.2 The organizational flowchart of this dissertation (where solid (dotted) lines represent formal and direct correlations (partial correlations) between two functional parties, and arrows indicate the flow of communication). 16

2.1 An illustration of a typical WMN for suburban/rural residential areas. 19

2.2 Frame structure. 19

3.1 An illustration of the network topology under consideration. 34

3.2 Traffic load estimation. 35

3.3 An illustration of a clustered WMN with 25 nodes. 45

3.4 Packet delays of the real-time traffic using the proposed path establishment scheme, the random path establishment scheme, and the max-rate path establishment scheme vs. the end-to-end rate for Node A and Node B (shown in Figure 3.3). 47

3.5 System throughputs of the proposed approach, the conflict-graph approach, the baseline approach, and the upper bound vs. the number of mesh routers (where $N = 1024$ and $B^{\max} = N/4$). 48

3.6 Frequency reuse ratios of the proposed approach, the conflict-graph approach, and the upper bound vs. the number of mesh routers (where $N = 1024$ and $B^{\max} = N/4$). 50

3.7 System throughputs of the proposed approach, the conflict-graph approach, the baseline approach, and the upper bound vs. the number of subcarriers (where the number of mesh routers is 100 and $B^{\max} = N/4$). 52

3.8 Frequency reuse ratios of the proposed approach, the conflict-graph approach, and the upper bound vs. the value of $B^{\max}$ (where the number of mesh routers is 100 and $N = 1024$). 53

3.9 System throughputs of the proposed approach, the conflict-graph approach, and the baseline approach vs. the value of $B^{\max}$ (where the number of mesh routers is 100 and $N = 1024$). 54

4.1 Subcarrier swapping. 67

4.2 A genetic representation of a set of feasible solutions to the optimization problem. 69

4.3 A pictorial illustration of the roulette wheel selection process. (All the individuals form a pie chart, where the size of a slice (i.e., the area where an individual occupies) is proportional to the fitness value of an individual. In this example, the 4^{th} individual is selected. The roulette wheel selection operator keeps selecting the individuals by literally spinning around the circle until the population size of the next generation is S.) 70

4.4 The flowchart of the proposed GA-based resource allocation scheme. 72

4.5 Comparison of the data throughput performance of the proposed approaches (GA-based, KKT-driven, and Combined-KKT-GA), the Shi's KKT-driven scheme, the Gao's scheme, and the random scheme vs. the number of links for background data traffic. 77

4.6 Comparison of the resource utilization of the three proposed approaches and the random scheme vs. the number of links for background data traffic. . . . 79

4.7 Comparison of the voice packet dropping rates of the proposed approaches (GA-based, KKT-driven, and Combined-KKT-GA), the Shi's KKT-driven scheme, the Gao's scheme, and the random scheme vs. the number of links for background data traffic. 80

4.8 Comparison of the system throughput of the proposed approaches (GA-based, KKT-driven, and Combined-KKT-GA), the Shi's KKT-driven scheme, the Gao's scheme, the random scheme, and the data capacity vs. the number of links for bursty data traffic. 82

4.9 Comparison of the resource utilization of the three proposed approaches and the random scheme vs. the number of links for bursty data traffic. 84

5.1 Examples of node cooperation in WMNs. 93

5.2 Schematic representation of relay-assisted transmissions. 94

5.3 Flowchart of the proposed four-phase centralized node cooperative resource allocation approach. 113

5.4 The phase structures of the proposed two-phase distributed node cooperative resource allocation approach. 115

5.5 Example of a typical node cooperation scenario in the transmission phase. . . 118

5.6 Flowchart of the proposed distributed approach. 119

5.7 Comparison of the throughput performance of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the number of mesh nodes (with $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, $T = 150$ms, and $\lambda = 50$ packets/second). 128

5.8 Comparison of the RG packet dropping rates of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the number of mesh nodes (with $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, $T = 150$ms, and $\lambda = 50$ packets/second). 130

5.9 Comparison of the throughput performance of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the polling time (with $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $\lambda = 50$ packets/second). 131

5.10 Comparison of the RG packet dropping rates of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the polling time (with $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $\lambda = 50$ packets/second). 133

5.11 Comparison of the throughput performance of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the number of RG nodes (with $M = 30$ mesh nodes, $T = 150$ms, and $\lambda = 50$ packets/second). 135

5.12 Comparison of the throughput performance of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the value of λ (with $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $T = 150$ms). 137

6.1 The slot structure and the channel sensing procedure for a secondary user. . 144

6.2 An illustration of the channel gains of two secondary users. 148

6.3 Reward performance of the proposed instinctive channel sensing in a single-secondary user scenario vs. the value of θ (where $N = 512$). 153

6.4 Channel utilization of the proposed instinctive channel sensing in a single-secondary user scenario vs. the value of θ (where $N = 512$). 154

6.5 Reward performance of the proposed instinctive channel sensing in a single-secondary user scenario vs. the value of N (where $\theta = 0.3$). 156

6.6 Channel utilization of the proposed instinctive channel sensing in a single-secondary user scenario vs. the value of N (where $\theta = 0.3$). 157

6.7 Comparison of the channel utilization of our proposed instinctive sensing approach, a random sensing approach, and the Jiang's sensing approach vs. the value of θ (where $N = 512$). 158

6.8 Simulation and analytical results for collision probability, P_c, in a multi-secondary user scenario vs. the number of secondary users, M, for different values of N (where $\theta = 0.9$). 162

6.9 Total reward performance of the proposed instinctive channel sensing vs. the value of M (where $\theta = 0.9$). 163

6.10 Resource utilization of the proposed instinctive channel sensing vs. the value of M (where $\theta = 0.9$). 164

7.1 The system throughput measure and the fairness measures vs. the value of J with equal weighting factors. 181

7.2 The relationship of system throughput and fairness with equal weighting factors. 183

7.3 The system throughput measure and the fairness measures vs. the value of J with different weighting factors. 185

7.4 The relationship of system throughput and fairness with different weighting factors. 187

A.1 Outage probabilities of cooperative transmissions using the DF cooperation mode and ordinary direct transmissions (where $E_{SD} = E_{RD}$, $a_S = 0.75$, $a_R = 0.25$, and $\sigma^2 = 10^{-3}$W). 211

List of Tables

3.1 Summary of important symbols used in this Chapter. 29

3.2 Relationship of the number of wireless nodes and the average number of clusterheads 44

4.1 Summary of important symbols used in this Chapter. 60

4.2 Summary of simulation parameters. 75

4.3 The characteristics of different resource allocation approaches. 85

5.1 Summary of important symbols used in this Chapter. 89

5.2 Relationship between the number of mesh nodes, M, and the node cooperation gain (NCG) (i.e., normalized throughput gain due to node cooperation) for the proposed approaches (with $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, $T = 150$ms, and $\lambda = 50$ packets/second) 129

5.3 Relationship between the value of T and the node cooperation gain (NCG) (with $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $\lambda = 50$ packets/second) 132

5.4 Relationship between the number of RG nodes, $|\mathcal{M}_1|$, and the node cooperation gain (NCG) (where $M = 30$ mesh nodes, $T = 150$ms, and $\lambda = 50$ packets/second) 136

5.5 Relationship between the value of λ and the node cooperation gain (NCG) (where $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $T = 150$ms) 138

6.1 Summary of important symbols used in this Chapter. 142

6.2 Simulation and analytical results for the probability of collision, P_c, in a two-secondary user scenario vs. the number of channels, N, for different values of θ. 160

7.1 Summary of important symbols used in this Chapter. 170

7.2 System parameters for the numerical analysis for the case of equal weighting factors 180

7.3 Numerical results for the case of equal weighting factors 182

7.4 System parameters for the numerical analysis for the case of different weighting factors 184

7.5 Numerical results for the case of different weighting factors 186

Chapter 1

Introduction

1.1 Overview of Wireless Mesh Networks

Wireless mesh networking has emerged as a promising solution for future broadband wireless access [4,11,18], supporting ubiquitous communications and mobile computing. Ubiquitous wireless access can be realized in various practical scenarios, namely home networking, office networking, and city networking [4,18]. This new networking paradigm provides not only a viable but also economical solution for both peer-to-peer applications and Internet access. Although the notion of mesh networking has been discussed extensively in wireline and optical networks [27], the preceding research mainly focuses on restoration of link failure and/or design of survivable and healing networks. On the other hand, with unique characteristics of wireless channels and scarcity of network resources, many new challenges on radio resource management for wireless mesh networks (WMNs) are raised such as capacity enhancement, interference control, quality-of-service (QoS) provisioning, effective medium access control (MAC), efficient packet scheduling, fairness assurance, etc. Therefore, directly applying existing resource management schemes for wireline and/or optical networks to WMNs can be inefficient and ineffective. In order to fully optimize system performance,

novel radio resource management approaches tailored for WMNs are imperative, which is the focus of this dissertation.

In general, WMNs consist of wireline gateways, mesh routers, and mesh clients, organized in a multi-tier hierarchical architecture [18], as shown in Figure 1.1. A mesh client network comprising several mesh clients can be formed in an *ad hoc* manner, and connected to one or more mesh routers. Mesh routers in fixed sites render a robust and reliable mesh backbone via wireless technology, providing relay service to the mesh client networks and other access networks such as cellular networks, wireless local area networks (WLANs), etc. This wireless mesh backbone provides a platform to integrate the wireless access networks, whereby a multi-mode mobile station with multiple air interfaces can roam freely among the access networks and subscribe to desired services. Wireline gateways are to connect the mesh backbone to the Internet backbone. One prominent attraction of this architecture stems from large-scale deployment, dynamic self-configuration, and self-management with high link reliability [4]. WMNs can be deployed in various practical scenarios with different purposes, thus referred to as multi-purpose WMNs [18]. Recently, the notion of wireless mesh networking has been attracting a plethora of attentions from academia and industry. In specific, wireless mesh networking for suburban/rural residential areas has been of great interest, for example, Mesh@Purdue [75], MIT Roofnet [78], Microsoft WMNs [76], Wray WMN [45], Nokia last mile access (LMA) [86], BelAir Networks [7], MeshDynamics [74], etc. Mesh routers can be set up at premises in the neighborhood, forming a resilient mesh backbone and providing an *all-wireless* environment for the target residential area. In contrast, the deployment of conventional urban community networking is mostly based on cable modems or digital subscriber lines (DSLs) connected to the Internet. With this conventional yet archaic networking structure, accommodating new terminals usually requires the addition of new switches and cables. An obvious pitfall is ascribed to its inflexibility. Suburban and rural areas can be underdeveloped and/or austere. Time-consuming operations such as

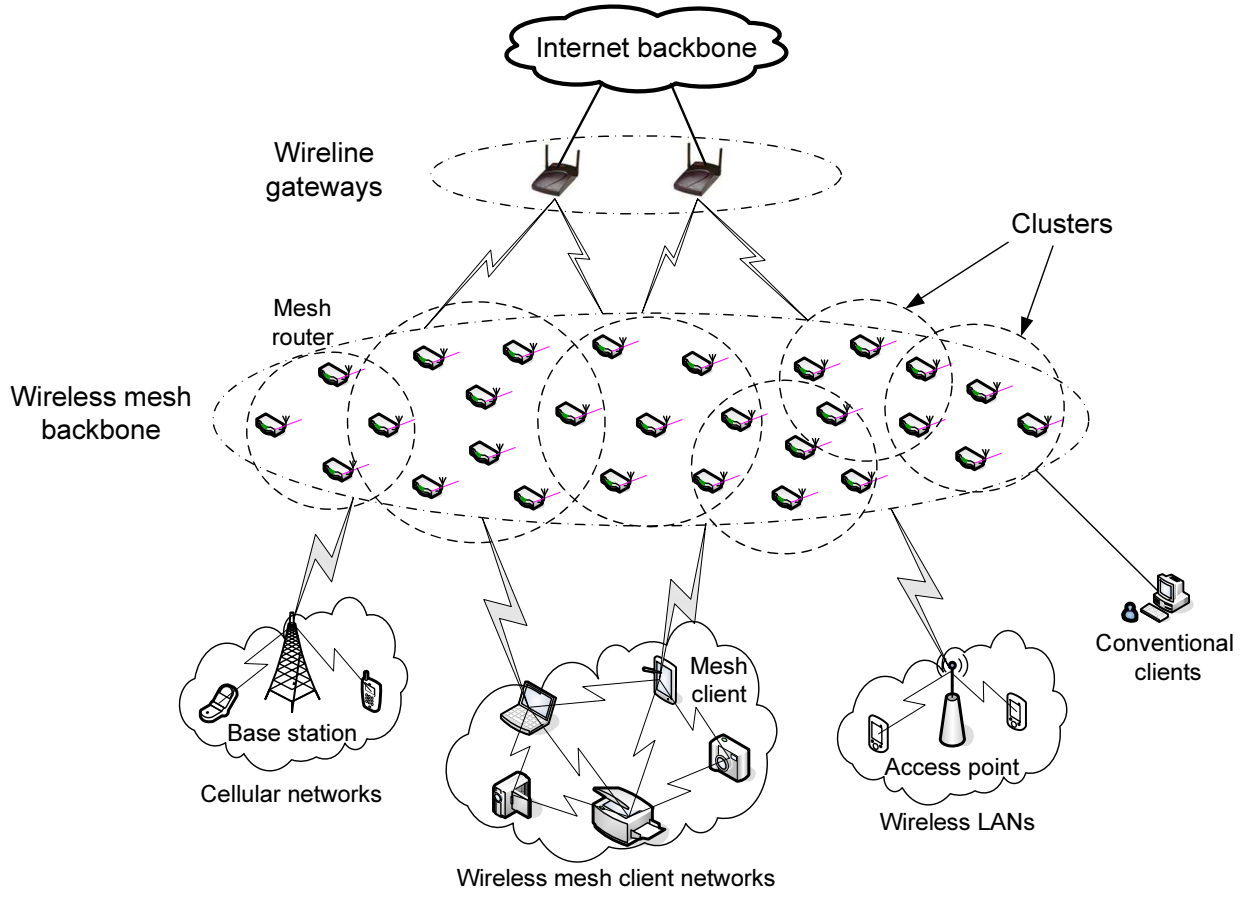


Figure 1.1: An illustration of wireless mesh networking.

land construction are usually required, and cable deployment in these austere environments becomes convoluted, leading to high setup and maintenance costs [86]. As a result, it is not favorable or even feasible to adopt such a traditional network design methodology for suburban/rural residential areas. In future suburban/rural area networking for broadband wireless access, a flexible, simple, cost-effective, and robust networking paradigm with the support of pervasive communications and computing is undoubtedly desired [52]. As such, wireless mesh networking offers a viable alternative with several advantages [18,86]: 1) large coverage; 2) scalability; 3) low provisioning cost; and 4) increased system capacity. As a consequence, WMNs are expected to be deployed extensively in suburban/rural residential areas in the foreseeable future.

1.2 Resource Management for Wireless Mesh Networks

In wireless communications, radio resource management is vital in controlling how scarce radio resources can be allocated, distributed, and utilized among all terminals in a system. Unlike wired links which have a constant link capacity, wireless links are relatively vulnerable due to fading over frequency and interference over time [73]. Effective resource management should be capable of mitigating the link impairments and exploiting the unique link features so as to optimize system-wise and user-wise performance. In order for WMNs to be deployed in suburban/rural residential areas successfully, we need to address certain challenges and issues with respect to resource management tailored for WMNs, outlined as follows.

- *Multi-channel communications* – In future broadband WMNs targeting high-speed communications, the notion of orthogonal frequency division multiplexing (OFDM) has been demonstrated as a promising modulation technology to support high data-rate transmissions with resistance to delay dispersion due to multi-path propagation. OFDM essentially converts a frequency-selective fading channel into a flat fading chan-

nel by employing multiple subcarriers [105], thereby mitigating the effect of inter-symbol interference (ISI). One pitfall of employing OFDM, however, is the need of accurate frequency synchronization so as to maintain perfect orthogonality among all OFDM subcarriers [105]. Mesh routers in WMNs, notwithstanding, are stationary. With no mobility and hence no Doppler effect, accurate frequency synchronization can be facilitated. As a result, OFDM-based transmissions can give better performance by mitigating the ISI problem in the high-speed WMNs. To cope with multimedia traffic, transmission rates can be varied by allocating a different number of subcarriers using orthogonal frequency division multiple access (OFDMA), realizing the notion of multi-channel communications. However, system performance depends upon the effectiveness and efficiency of channel allocation. Thus, simple yet effective channel allocation tailored for WMNs is essential.

- *Distributed multi-hop transmissions* – In urban metropolitan areas, cellular systems can be easily deployed. In contrast, underdeveloped terrains in suburban/rural residential areas discourage the setup of base stations, thereby favoring distributed control [86]. Besides, such adverse environments thwart the direct line-of-sight communications (i.e., 1-hop transmissions). Even if the maximum power constraint can be relaxed, a successful data transmission over a poor wireless link requires a very high transmit power. As such, more interference is induced, deterring the notion of *frequency reuse* and hence resulting in inefficient use of scarce network resources. On the other hand, mesh networking provides ease of deployment, and offers greater coverage of wireless access by means of multi-hop transmissions. In multi-hop WMNs, data generated from a source is to be forwarded to a destination by intermediate nodes via multiple hops. Thus, lower transmit power is needed for transmissions over each hop, and/or a higher transmission rate can be achieved. In addition, with the help of meshing, the isolation of nodes and the network breakdown can be avoided by providing alternative paths.

However, the way of packet delivery can affect interference control, whereby system performance can be degraded. Without centralized control, new distributed resource allocation strategies with efficient packet delivery are needed.

- *MAC-layer packet scheduling* – With the rising demand for multimedia applications such as videoconferencing, PowerPoint presentation streaming, voice over IP (VoIP) calls, and file transfer, MAC-layer service differentiation is vital in providing fine-grain QoS assurance. In particular, real-time traffic (e.g., VoIP calls) should be assigned high priority, while non-real-time traffic (e.g., emails) should be assigned low priority. To further support aforementioned multimedia services effectively, (per-class) packet scheduling with resource reservation is imperative for QoS provisioning. In WMNs with distributed control, however, resource reservation can be difficult. Without proper network coordination, packet collisions can occur, voiding the effectiveness of service differentiation and hence packet-level QoS provisioning. In addition, the overhead incurred by message exchanges can cause resource utilization reduction. Thus, new efficient MAC with effective packet scheduling tailored for WMNs is indispensable.

- *Effective frequency reuse* – The efficiency of a multi-channel WMN is key to the success of providing high-speed wireless access for diverse multimedia applications. With meager radio resources, efficient use of network resources is important. Apart from packet scheduling, the concept of frequency reuse can be exploited to utilize the resources more efficiently. Two transmissions can employ the same channel(s) if they are far away enough such that the co-channel interference level is below a required threshold [73]. With effective frequency reuse, system capacity can be further increased. To realize frequency reuse, interference control is necessary. Without powerful central controllers, distributed channel allocation for WMNs is required, and how to achieve maximal frequency reuse in WMNs by means of distributed channel allocation is certainly a challenging research issue.

- *Tradeoffs among resource utilization, QoS provisioning, and fairness*[1] – With limited resources, the goals of attaining high resource utilization, provisioning QoS, and maintaining fairness are conflicting with each other. For example, resources can be utilized efficiently if only the terminal with the best channel condition transmits, whereby the maximum throughput can be acquired [106]. Such an opportunistic transmission, however, gives rise to unfairness and plausibly violates the QoS requirements (e.g., delay bound of a voice packet) of some wireless terminal(s). If strict fairness is achieved according to some fairness criteria (such as max-min fairness and proportional fairness [83]), resource utilization can be quite low, and QoS support cannot be guaranteed [71]. Simply providing QoS assurance will result in poor resource utilization and unfairness, for some resources are usually reserved for high-priority terminals, leading to conservative resource allocation. Thus, finding a desired performance balance among the three aforesaid research goals is imperative.

- *Performance enhancement via advanced technologies* – Advanced communications technologies are commonly used to enhance system performance, particularly in the physical layer. For example, with a large array of antennas set up at a transmitter and/or a receiver, the multiple-input-multiple-output (MIMO) technology can be employed to not only tremendously boost the system capacity but also effectively attain a tradeoff between spatial multiplexing gain and diversity performance [104]. Given complete knowledge of channel state information, system performance can be further improved by signal beamforming with the help of directional antennas. With the directional antennas in place, the notion of frequency reuse can be fostered due to efficient interference control [104]. Recently, the concept of *network coding* has emerged and drawn a lot of attention [67]. By employing network coding-based routing in the network

[1]Due to the subjective nature, the notion of fairness cannot be uniquely and well defined. The simplest interpretation of it is the equitable share of resources according to a prearranged consensus or agreement [83].

layer, it can be shown that system throughput can be further increased, compared to conventional routing. To resemble MIMO, distributed *node cooperation* is also of great interest [64]. Beneficial node cooperation not only increases system throughput, but also improves transmission accuracy (i.e., higher diversity order can be achieved). To better utilize network resources, the notion of *cognitive radio*, where unlicensed secondary users can use the temporarily unoccupied spectrum licensed to primary users [38]. Thus, performance enhancement in WMNs can be realized by means of advanced communications technologies.

- *Simplicity versus optimality* – To attain the best system performance, optimal resource allocation is needed. However, obtaining globally optimal resource allocation solutions usually requires high computational complexity. In most cases, even powerful devices such as base stations cannot execute optimal resource allocation algorithms in a timely fashion. Therefore, applying those optimal resource allocation algorithms can be very inefficient and ineffective in practice. On the other hand, simple resource allocation algorithms can easily be implemented in practical systems. An obvious drawback of simple algorithms is that the system performance can be far from optimal. To strike a balance between simplicity and optimality, we should focus on devising efficient yet effective resource allocation strategies for decentralized WMNs with QoS support. In specific, our proposed approaches should be simple in terms of time complexity, and effective in terms of system performance improvement and QoS support assurance.

1.3 Motivations and Objectives

In this section, we present our motivations and research objectives.

- *Why WMN-specific node clustering?* In a wireless network with random channel access, the throughput of a wireless node generally decreases with the number of

nodes [36]. The implication is that a node should only communicate with nearby nodes, thereby favoring *clustering* [4,24]. In the literature, clustering is an effective way to manage a large wireless network [119]. Multi-level hierarchical clustering schemes are shown to achieve better system performance [6]. In fact, the notion of clustered WMNs has recently received an increasing attention from industry such as MeshAP-Pro and MeshBroschure. Although clustering has been researched in the context of sensor networks and mobile ad hoc networks (MANETs) for years, applying the existing clustering schemes to WMNs can be inefficient and ineffective due to different networking characteristics and design objectives. Thus, a new node clustering approach specifically tailored for QoS-sensitive WMNs is indispensable. On the other hand, in order to efficiently support multimedia services and increase system capacity, effective channel assignment and hence interference control are imperative to facilitate QoS provisioning and frequency reuse [21]. However, as mentioned in Section 1.2, austere suburban and rural environments discourage the notion of centralized control. With the help of node clustering, interference control and hence frequency reuse can be facilitated by channel allocation via clusterheads in a decentralized manner. In a cluster, collision-free scheduling is feasible, to satisfy various QoS demands. Resource allocation can then be carried out in a hybrid centralized-distributed fashion. Therefore, devising a node clustering scheme with effective channel allocation tailored for WMNs is imperative, taking into consideration the frequency reuse and QoS requirements of heterogeneous traffic.

- *Why joint power allocation, subcarrier allocation, and packet scheduling?* In OFDM-based WMNs, subcarrier allocation over the frequency domain is necessary to support high-speed multimedia applications (e.g., voice, video, and data) and grant diverse transmission rates. By assigning different subcarriers to different nodes, simultaneous transmissions can be fostered in a cooperative manner, plausibly increasing system

capacity. Since different frequency bands experience different fading characteristics, power allocation with respect to channel conditions is shown to be crucial for QoS provisioning in multi-channel communications. In order to effectively and efficiently support various services with QoS assurance in the MAC layer, bandwidth reservation and hence packet scheduling over the time domain are imperative. Therefore, joint power-frequency-time resource allocation is required. However, since this joint resource allocation problem is NP-hard, low-complexity algorithms are strongly desired. To further streamline packet-level QoS provisioning, service differentiation in the MAC layer is fundamental, which can be achieved via packet prioritization. In particular, higher-priority packets are served before lower-priority ones, realizing the notion of per-class packet scheduling. Therefore, efficient resource allocation algorithms with effective packet-level QoS provisioning for the joint power-frequency-time resource allocation problem are imperative.

- *Why node cooperation?* Increasing throughput is one of the key factors to the success of WMNs. Many resource allocation strategies have been proposed so as to provide a high-speed mesh backbone with QoS assurance. To further enhance the system performance, *cooperative diversity* or node cooperation can be employed to achieve a spatial diversity gain by way of a virtual antenna array formed by multiple wireless nodes in a distributed fashion [82]. In fact, node cooperation has been demonstrated promising in improving the spectral and power efficiency of wireless networks without additional complexity of multiple antennas [64]. The basic idea behind node cooperation rests on the observation that the signal transmitted by a source node can be overheard by other nodes in a wireless environment. The source and its *partner(s)* can jointly process and transmit their information, thereby creating a virtual antenna array and achieving a desired diversity-multiplexing tradeoff [8]. Although traditional co-located multi-antenna techniques are quite attractive for future broadband wireless access, setting up a large

antenna array at wireless terminals (e.g., wireless mesh routers) is impractical due to their size and power limitations. In contrast, node cooperation can provide a comparable spatial diversity gain without imposing extra hardware complexity on the devices. However, in the context of *non-altruistic* node cooperation, cooperative transmissions do not always outperform ordinary direct transmissions in terms of throughput [25]. Therefore, exploiting the merits of *beneficial* node cooperation is important for effective system performance enhancement.

- *Why channel sensing via cognitive radio?* In conventional wireless communication systems, spectrum resources are usually governed by license holders. This radio resource management allows licensed users to access the spectrum with no or minimal interference. However, recent studies show that many frequency bands in the radio spectrum are underutilized most of the time [103]. Due to the inefficiency of our current spectrum allocation, the notion of cognitive radio has emerged as an intelligent and promising solution, allowing dynamic spectrum access and hence alleviating the problem of low resource utilization. The success of cognitive radio is highly contingent upon the effectiveness of how (unlicensed) secondary users utilize the temporarily available spectrum bands that are licensed to primary users. In the context of WMNs, primary users are the registered subscribers who are guaranteed access to assigned radio resources, whereas secondary users are unregistered who can only access the spectrum when no primary users are active (i.e., no primary activity) in the neighborhood of interest. As such, in cognitive radio networks (CRNs), to guarantee the QoS of primary users, it is indispensable for secondary users to ensure that the spectrum is free of primary activities before transmitting their information. In the case of multi-channel networks, secondary users equipped with a simple transceiver have to sense the channels one at a time to determine which channel is available, if any. Therefore, an effective and efficient channel sensing order is crucial in both resource utilization

improvement and QoS support.

- *Why a unified framework for balancing throughput and fairness?* In traditional wireless networks, system throughput is a common performance metric [55]. However, next-generation wireless networks including WMNs are anticipated to support multimedia traffic (e.g., voice, video, and data traffic). With heterogeneous traffic, QoS provisioning and fairness support are also important. As mentioned in Section 1.2, with limited available radio resources, throughput maximization, QoS provisioning, and fairness assurance are conflicting goals [71], conducing to a natural tradeoff among these three performance measures. In particular, balancing system throughput and fairness for WMNs with QoS support and high resource efficiency is of great interest. However, little research work addresses the optimal relationship between throughput and fairness. In addition, there is no widely accepted unified framework to effectively attain different degrees of performance tradeoff between throughput and fairness with QoS support and efficient resource utilization. Therefore, a unified framework to effectively and efficiently balance throughput and fairness with QoS support is desired.

1.4 Research Contributions

In this research, we intend to address the following research question:

Given a suburban/rural residential area, how can we design a WMN so as to provide users with high-speed wireless access, support the QoS of multimedia applications, and improve radio spectrum utilization by means of effective and efficient resource allocation?

As such, this dissertation is to tackle the issues of radio resource allocation for WMNs supporting heterogeneous traffic in the avenues of QoS assurance, throughput melioration,

resource utilization improvement, and system performance balancing. The research contributions and significance of this dissertation are summarized as follows.

- **QoS-Aware Node Clustering and Subcarrier Allocation for WMNs** [24] – We study the joint problem of node clustering and subcarrier allocation with QoS assurance. By introducing an upper bound on the number of subcarriers allocated to a cluster, we propose a joint QoS-aware node clustering and subcarrier allocation approach tailored for WMNs. This novel resource management approach achieves Pareto optimality, demonstrating efficient use of network resources. Our proposed approach is of low complexity, leading to a viable candidate for practical implementation. Our findings reveal that how to allocate resources in a decentralized fashion can affect the solution space of a performance tradeoff between QoS provisioning and throughput maximization. This work is presented in Chapter 3.

- **Joint Power-Frequency-Time Resource Allocation with Effective QoS Provisioning for WMNs** [21,23] – We address the joint problem of power-frequency-time resource allocation for WMNs with QoS support. First, we prove that the joint resource allocation problem is an NP-hard problem. Then, we propose three resource allocation approaches, namely 1) a Karush-Kuhn-Tucker (KKT)-driven approach, 2) a genetic algorithm (GA)-based approach, and 3) a hybrid KKT-GA approach. Simulation results show that our newly proposed approaches outperform their resource allocation counterparts in terms of packet-level QoS provisioning. Our results show that call admission control is vital to guarantee the QoS support of ongoing multimedia calls in service. We also observe that admitting more multimedia calls can greatly reduce the system throughput. This work is presented in Chapter 4.

- **Performance Enhancement I for WMNs via Node Cooperation** [20] – We investigate the problem of non-altruistic node cooperative resource allocation for WMNs

with QoS support, taking subcarrier allocation, power allocation, partner selection/allocation, service differentiation, and packet scheduling into account. Two low-complexity node cooperative resource allocation approaches are proposed, namely four-phase centralized resource allocation and two-phase distributed resource allocation. Both proposed approaches are shown effective in both packet-level QoS provisioning and system performance enhancement. Simulation results also demonstrate that the proposed approaches are less vulnerable to the changes in the system parameters such as the accuracy of traffic load estimates. Further, our study reveals a critical principle that whether node cooperation is beneficial depends upon the nature of node cooperation, the mode of network operation, and the traffic pattern. This work is presented in Chapter 5.

- **Performance Enhancement II for WMNs via Cognitive Radio** [19] – We address the issue of how secondary users utilize unoccupied network resources by means of cognitive radio. By employing the theory of optimal stopping, we propose a simple channel sensing order for cognitive radio networks, referred to as instinctive channel sensing. By sensing the channels according to the descending order of their achievable rates, we can prove the optimality of our stopping rule that a secondary user should stop at the first sensed free channel. The probability of collision in a multi-secondary user scenario with respect to the proposed instinctive channel sensing is derived analytically and validated by extensive simulations. Further, we show that, as the number of secondary users increases, the system performance can be improved with the proposed channel sensing order. Our results also indicate that resource utilization can be maximized when the number of secondary users approaches the number of channels at the expense of a higher collision probability. This work is presented in Chapter 6.

- **Balancing Throughput and Fairness for WMNs with QoS Support** [22] – We derive a unified optimization framework to effectively balance throughput and fairness

with QoS support. With our proposed framework, we can procure the optimal relationship (i.e., tradeoff curve) of system throughput and fairness with QoS support. By introducing a bargaining floor, a relationship curve is obtained by solving the proposed optimization problem iteratively. Different degrees of performance tradeoff between throughput and fairness can be acquired by simply choosing appropriate values of the bargaining floor. Given an application of interest, the desired operating tradeoff can be determined. Verified by game theory, the resource utilization is efficient. Further, our numerical study shows that the tradeoff curve of system throughput against fairness is concave in shape, meaning that a unit decrease in system throughput leads to a larger marginal improvement in fairness performance. This work is presented in Chapter 7.

1.5 Outline of the Dissertation

This dissertation is organized as follows. The system model of this research is presented in Chapter 2. The joint problem of node clustering and subcarrier allocation is given in Chapter 3. Joint power-frequency-time resource allocation is presented in Chapter 4. Performance enhancement via non-altruistic node cooperation is described in Chapter 5. Performance enhancement via cognitive radio is given in Chapter 6. Chapter 7 presents an optimization framework for balancing throughput and fairness with QoS support. Finally, conclusions and further work are given in Chapter 8. To better illustrate the interplay among our research accomplishments, the organizational flowchart of this dissertation is depicted in Figure 1.2.

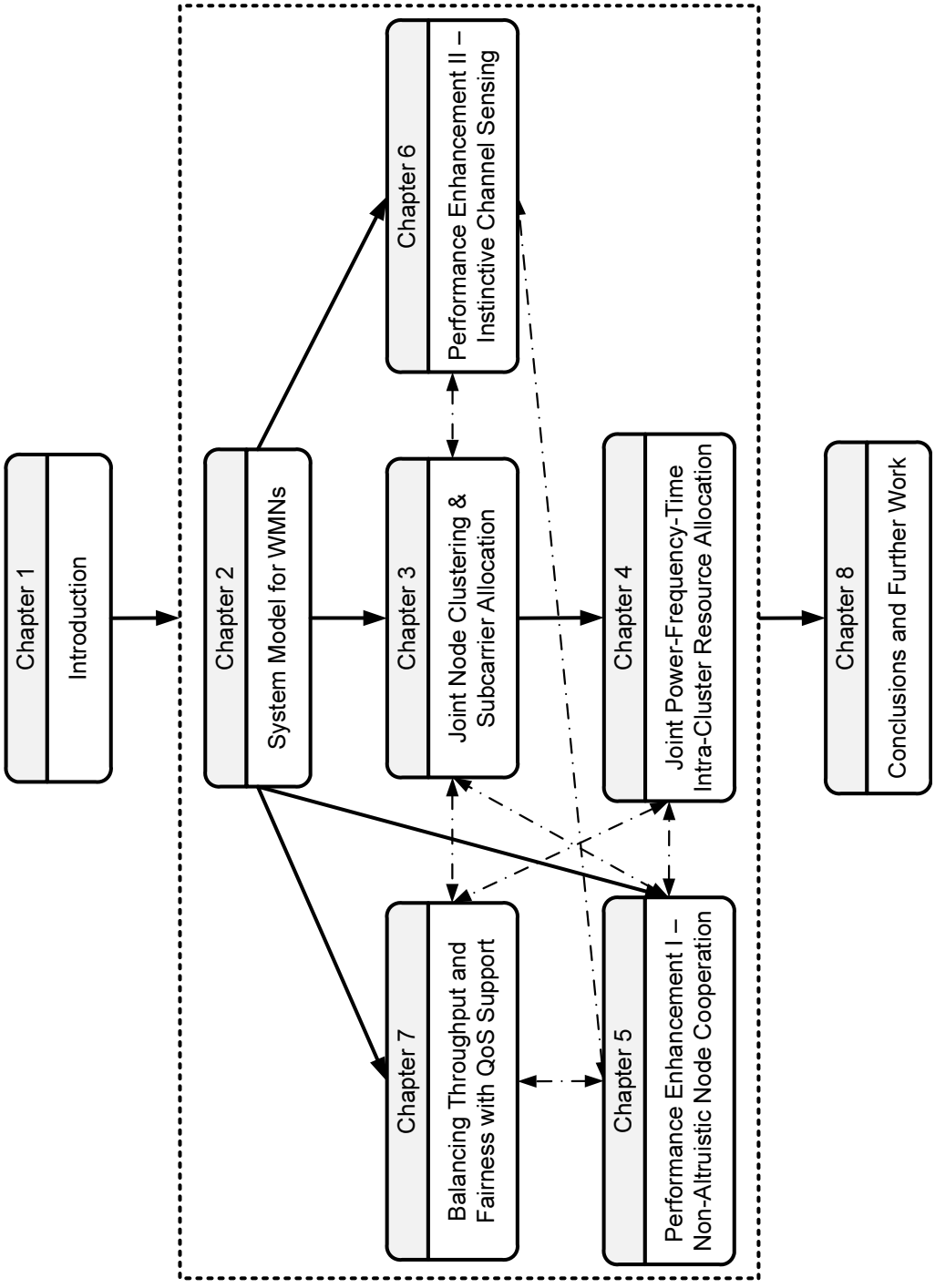


Figure 1.2: The organizational flowchart of this dissertation (where solid (dotted) lines represent formal and direct correlations (partial correlations) between two functional parties, and arrows indicate the flow of communication).

Chapter 2

System Model

2.1 Network Architecture

In this research, we consider an OFDM-based WMN for suburban/rural residential areas, consisting of a wireline gateway attached to the Internet backbone and a number of mesh routers and mesh clients scattered around. Some mesh routers are directly connected to a gateway, whereas the others are scattered around, rendering a hierarchical multi-hop network (see Figure 2.1). As discussed in Chapter 1, adverse suburban/rural residential areas favor the distributed control for the network operation. Besides, such austere networking environments thwart one-hop direct communications as opposed to multi-hop transmissions, providing ease of deployment and offering greater coverage of wireless access. In multi-hop WMNs, the links are relatively shorter, and hence the channel conditions are better. As a result, higher data rates can be attained at the cost of longer delays. In order to achieve load balancing and high throughput performance, gateways should be opportunistically placed in the WMN of interest (e.g., [31,66]); but how to place multiple gateways effectively in WMNs is beyond the scope of this work. In our research, mesh routers are assumed non-mobile and hence the channel gains can be estimated accurately (e.g., via pilot symbols [15]). We

consider that each mesh node is equipped with one transceiver with an omni-directional antenna, so that it cannot transmit and receive simultaneously. A mesh router can be a transmitter, relay, or receiver at different times. Mesh routers have no power-consumption constraints, but their maximum transmit power is bounded. To strive for a stable and scalable wireless network, node clustering is shown effective in providing networking stability and system throughput increase [119]. Besides, frequency reuse and QoS provisioning can be facilitated by means of node clustering. In this research, we consider that the WMN of interest is divided into several clusters according to our novel node clustering algorithm (proposed in Chapter 3). In specific, one node is selected as a clusterhead in a cluster. The main responsibility of the clusterhead is to facilitate interference control by effective channel allocation and perform packet scheduling for the active connections in the cluster. As a consequence, resource allocation can be carried out in a hybrid centralized-distributed fashion (i.e., centralized intra-cluster and distributed inter-cluster resource allocation). More details are to be discussed in Chapter 3. Further, we consider a synchronized WMN. Time is partitioned into frames, each of which is further divided into a beacon slot, a control slot, and L DATA slots. The structure of a frame is shown in Figure 2.2. In essence, a clusterhead collects the transmission requests from its clustermember(s) in the control slot, and announces the transmission schedule in the subsequent beacon. Call admission control (CAC) is assumed in place such that the QoS requirements of an admitted call (flow) can be satisfied. With the OFDM technology, each mesh router can choose a set of subcarriers for DATA transmissions and/or receptions, allowing simultaneous transmissions over different subcarriers in the WMN of interest.

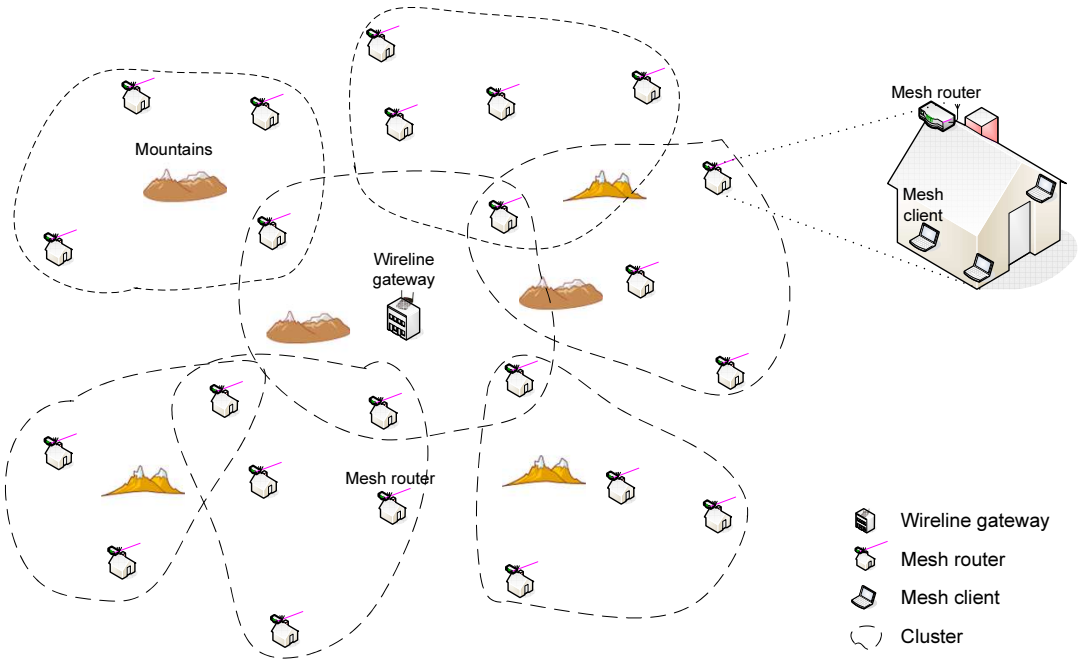


Figure 2.1: An illustration of a typical WMN for suburban/rural residential areas.

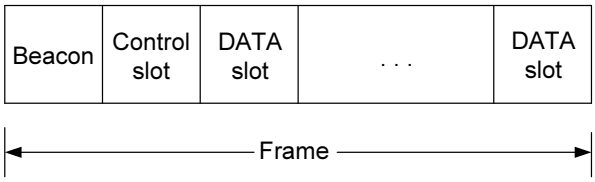


Figure 2.2: Frame structure.

2.2 Traffic Traversal over WMNs

2.2.1 Traffic Characterization

In typical WMNs, traffic characteristics (i.e., traffic load and traffic types) can vary greatly from one mesh router to another as different routers are placed in different locations (e.g., access/client networks). In the context of suburban/rural residential mesh networks, mesh routers mounted on the rooftops of residences comprise a wireless mesh backbone, while mesh clients associated with their closest mesh router(s) constitute various access networks. Depending on practical applications, traffic flows in WMNs can be classified into two types:

1) *Vertical traffic* – Traffic is to traverse from mesh clients to mesh routers to the gateway or vice versa (i.e., to/from the Internet); and

2) *Horizontal traffic* – Traffic is to traverse within the same WMN (i.e., to/from the mesh router(s)) in a peer-to-peer fashion.

For the vertical traffic traversal, in essence, mesh routers receive the packets generated from different traffic sources (i.e., wireless devices such as wireless phones, Internet Protocol Television (IPTV), and laptops) in mesh client networks (i.e., residential houses). Then, the mesh routers access the wireless medium and transmit the received packets to their neighboring mesh routers or the gateway. The transmission path of a traffic flow is determined by a routing protocol. In this research, we assume that there is a routing protocol in place so that over which link a packet is to traverse can be predetermined. Packets can then be transmitted to the next-hop mesh router(s) or dropped due to delay bound violation. Whether a packet transmission is successful over the wireless medium depends upon channel conditions (e.g., fading) and medium access strategies (e.g., MAC protocols). End-to-end packet delivery is considered complete when a packet generated from a source (e.g., a wireless device at a residence) is successfully transmitted to a destination (e.g., the Internet). On

the other hand, for the horizontal traffic traversal (i.e., peer-to-peer transmissions), packets generated from one residence are transmitted to a mesh router which forwards the packets to a mesh router at the target destination (i.e., the target residence) through multi-hop packet transmissions.

2.2.2 Traffic Model

In future broadband wireless access, WMNs are expected to support various kinds of multimedia applications. In suburban/rural residential areas, plausible applications include VoIP for phone conversations, video streaming for a home theater, file sharing, web browsing, etc. To realize all the aforementioned services in practice, we consider three types of traffic, namely voice, video, and data.

- *Voice traffic* – Voice traffic is similar to constant-bit-rate (CBR) class in asynchronous transfer mode (ATM) networks. It is real-time traffic with strict delay (and jitter) requirements, but can tolerate a certain bit error rate (BER). Voice packets are normally assigned high priority for transmission. In packet-switched wireless networks, compressed voice packets are preferred so as to avoid collisions (in case of contention-based MAC) and/or reduce interference (in case of interference-based MAC). A common voice codec designed for VoIP applications is G.729, which can be modeled as a two-sate ON-OFF model [2]. In the ON state, a fixed-size packet arrives at a constant rate, whereas in the OFF state, no packet is generated. The duration of an ON period and that of an OFF period both follow an exponential distribution. A desirable packet loss rate (PLR) is 1% or less for a good quality conversation [49].

- *Video traffic* – Like voice traffic, video traffic is real-time traffic, but its bandwidth requirement varies over time, depending on video contents, network congestion, physical channel conditions, etc. Based on different video coding schemes, transmission

rates can be dynamically adapted. Thus, different BER targets and video resolutions can be adjusted accordingly [65]. Video traffic is usually assigned high priority. A common video codec designed for video transmission (i.e., video streaming and video-conferencing) is H.323, which is an umbrella recommendation from the International Telecommunications Union (ITU) that defines protocols to provide audio-visual communication sessions without any QoS guarantee. However, QoS assurance for video applications is indispensable in future broadband wireless access. Typical video traffic can be modeled as a combination of multiple ON-OFF mini-sources [2].

- *Data traffic* – In general, data traffic can be viewed as non-real-time traffic. Data traffic is normally assigned the lowest priority, for it has neither delay requirements nor rate requirements. This non-real-time traffic can further be divided into two types: 1) bursty data traffic, where data packet arrivals are commonly modeled as a Poisson process; and 2) background data traffic, where data packets can be transmitted whenever possible. The size of a data packet usually follows a heavy-tailed distribution [91] (e.g., Weibull distribution). Since data packets are delay-insensitive, advanced channel coding, error control mechanisms, and/or TCP can be employed to ensure transmission accuracy (i.e., to satisfy BER requirements).

In practice, each mesh router relays an aggregate of mixed diverse traffic types generated from client mesh networks (i.e., residential houses). To resemble mixed traffic types in WMNs, we consider that a voice source, a video source, and a data source reside at every mesh node in our research.

2.2.3 Traffic Differentiation

Concerning the way of how packets are transmitted over WMNs, packet delivery can be categorized into three types [35]: 1) best-effort packet delivery; 2) per-class packet delivery;

and 3) per-flow packet delivery. For the best-effort packet delivery, incoming packets from different traffic flows are put in a single queue at a mesh router which then basically transmits the queued packets in a first-in-first-out (FIFO) manner. No scheduling is involved. One obvious drawback is that there is no service differentiation and hence no QoS provisioning. In order to differentiate various traffic types, per-class packet delivery is needed. The notion of the per-class packet delivery rests on the fact that a mesh router schedules and delivers the packets based on their *class*, by assigning different priorities to different traffic types. In per-class scheduling, each mesh router possesses multiple queues designated for different traffic classes. Incoming packets from different traffic flows are classified and put into different queues according to their classes. Then, a per-class scheduler is to allocate bandwidth for transmitting the packets based on their class QoS specifications. For example, real-time traffic such as voice is usually assigned a higher priority over non-real-time traffic such as background data traffic. In a nutshell, higher-priority (lower-priority) traffic is to be served first (later). However, individual traffic flow information is lost during the traffic classification at the queues. On the other hand, per-flow packet delivery is capable to provide fine-grain QoS guarantees, for each mesh router allots a separate queue for each incoming traffic flow. A per-flow packet scheduler then allocates network resources according to the QoS requirements of each individual flow. One major pitfall of using per-flow scheduling, however, is its poor scalability. Since the complexity of per-flow resource allocation increases greatly with the number of traversing flows, applying the methodology of per-flow resource allocation to large-scale WMNs is not practical or even feasible. For the sake of practical implementation, therefore, we mainly focus on per-class traffic differentiation for resource management with heterogeneous traffic in this research.

2.3 Medium Access with QoS Provisioning

To facilitate resource allocation and packet scheduling at the packet level, an efficient and effective MAC protocol[1] is imperative. In this research, we make use of a simple MAC protocol to illustrate our proposed (per-class) packet scheduling. First, each mesh node estimates its traffic load by averaging the rate requirement over a fixed estimation window (e.g., 100ms) on a regular basis. Second, in the control slot of an MAC frame (see Figure 2.2), a clusterhead polls its clustermembers and collects their traffic demands periodically. Third, the clusterhead runs our proposed resource allocation algorithm(s) and announces a resource allocation decision (such as subcarrier allocation and power allocation) in the next beacon slot. To provide fine-grain QoS support, bandwidth reservation is imperative. As such, those timeslots and channels/subcarriers allocated to a particular node (or link) are reserved for packet transmissions in the DATA slots until the next polling. To further streamline QoS provisioning and provide service differentiation, packet prioritization is imperative, where a higher-priority packet is served before a lower-priority one, as discussed in Section 2.2.3. We conceive that the priority of real-time traffic (packets) is related to the performance of their packet dropping rates. Notice that, in this research, we consider that packet dropping is merely due to delay bound violation, under the assumption of accurate transmission at the physical layer. The higher the packet dropping rate that a real-time traffic flow suffers from, the higher the priority of the packets associated with that flow. After gathering the transmission requests, the clusterhead grants the requests of those higher-priority packets first. In other words, our QoS provisioning strategy for the real-time traffic is contingent on the packet dropping rate.

[1]Notice that optimal MAC protocol design for WMNs is desired [18]; addressing this issue, however, is beyond the scope of this dissertation.

2.4 Performance Metrics

To gauge the effectiveness and efficiency of our proposed resource allocation strategies, certain performance measurements are necessary. In this research, we consider the following performance metrics.

- *Throughput* – The throughput is defined as the number of successful bits transmitted over a link per second. This performance measure is applied to both the non-real-time and real-time traffic.
- *System throughput* – The system throughput is defined as the sum of the throughputs obtained over all the active links.
- *System capacity* – The system capacity is defined as the maximum achievable system throughput in the whole WMN of interest.
- *Channel utilization* – The channel utilization is defined as the ratio of the throughput obtained over a link to the maximum transmission rate of a channel.
- *Resource utilization* – The resource utilization[2] is defined as the fraction of system throughput contributed by all the active links.
- *Packet delay* – The packet delay is defined as the interval between the instant that a packet arrives at a source node and the instant the packet is transmitted successfully to its destination. This performance measure is applied to the real-time traffic.
- *Packet dropping rate* – The packet dropping rate is defined as the fraction of discarded packets due to the delay bound violation. This performance measure is applied to the real-time traffic.

[2]In this research, signalling overhead is not taken into account when we evaluate resource utilization.

- *Frequency reuse ratio* – The frequency reuse ratio is defined as the average number of occurrences that a channel (or subcarrier) is used simultaneously per DATA slot.

In this research, we conduct both numerical analysis and computer simulations to evaluate system performance, where our simulation programs are written using Matlab.

2.5 Chapter Summary

In this Chapter, we present the system model under consideration in this research. In specific, we focus on a multi-hop multi-channel clustered WMN for suburban/rural residential areas. We consider both vertical traffic and horizontal traffic traversing the WMN of interest, where diverse traffic types are taken into account. To strike a desired balance between time complexity and QoS support, per-class traffic differentiation is used in our resource management. Our QoS provisioning strategy at the packet level is also described. Further, the performance measurements employed in this research are defined.

Chapter 3

QoS-Aware Node Clustering and Subcarrier Allocation

3.1 Introduction

As discussed in Section 1.3, node clustering is a viable approach to improve system performance and facilitate QoS support. Although clustering has been researched in the context of sensor networks and mobile ad hoc networks (MANETs) for years, applying the existing clustering schemes to WMNs may not be efficient or effective due to different networking characteristics and design objectives. The goal for establishing WMNs is to provide ubiquitous communications to users and render an efficient mesh backbone with QoS support [18], while the primary purpose for deploying sensor networks is to offer environmental monitoring (e.g., temperature, pressure, etc) and/or surveillance (e.g., military field surveillance) [118]. Mobility and energy efficiency are the major concerns in MANETs [119], where the nodes are mobile and have power constraints, but there are no such limitations in WMNs. Further, the deployment of WMNs is relatively permanent, giving rise to the need of high efficiency of WMNs with QoS support. With the help of clusterheads, interference control and hence

frequency reuse can be facilitated by channel allocation in a decentralized manner. In this Chapter, we address the issues of node clustering and subcarrier allocation for WMNs with QoS support. The contributions and significance of this research work are three-fold [24].

- First, we study the joint problem of node clustering and subcarrier allocation for WMNs. By introducing an upper bound on the number of subcarriers allocated to a cluster, we propose a node clustering algorithm with effective subcarrier allocation. The proposed node clustering algorithm is QoS-aware, and the subcarrier allocation is optimality-driven and can be performed in a decentralized manner.

- Second, our proposed resource allocation approach outperforms two counterparts, namely a baseline approach and a conflict-graph approach. With increased frequency reuse, our proposed approach achieves high system throughput, and provides a good performance compromise between packet delay and end-to-end transmission rate for real-time traffic. Simulation results show that, by adjusting the value of the upper bound of subcarriers allocated to clusters carefully, we can achieve an improved system performance. The proposed scheme is also shown to be Pareto optimal, making efficient use of scarce radio resources.

- Third, our results not only confirm the fact that QoS provisioning and throughput maximization are two conflicting performance metrics [22], but also reveal that how to allocate resources in a decentralized wireless network can affect the solution space of a performance tradeoff between QoS provisioning and throughput maximization. This important revelation gives rise to the need of a new framework for distributed QoS-aware resource allocation for WMNs.

The remainder of this Chapter is organized as follows. Related work is given in Section 3.2. The proposed joint QoS-aware node clustering and tax-based subcarrier allocation is presented in Section 3.3. Efficiency of the proposed resource allocation approach from the

Table 3.1: Summary of important symbols used in this Chapter.

Symbol	Definition
$U_m(\mathbf{c}, \mathbf{p}, R_m^d)$	utility function the m^{th} clusterhead
R_m^d	QoS demand of the m^{th} clusterhead
$S_{m,n}^l(\cdot)$	payoff function of the m^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot
$r_{m,n}^l$	achievable data rate of the m^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot
$c_{m,n}^l$	indicator of allocating the n^{th} subcarrier to the m^{th} clusterhead on the l^{th} timeslot
$p_{m,n}^l$	transmit power of the m^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot
$T_{km,n}^l$	tax paid by the m^{th} clusterhead for generating interference to the k^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot
M	number of clusterheads
N	number of available subcarriers
L	number of timeslots (i.e., DATA slots) in a frame
$B^{\max}$	upper bound of the number of allocated subcarriers in a cluster

perspective of game theory is addressed in Section 3.4. Performance evaluation is given in Section 3.5. A brief discussion is provided in Section 3.6. Finally, we will summarize this Chapter in Section 3.7. A summary of important symbols used in this Chapter is given in Table 3.1 for easy reference.

3.2 Related Work

Node clustering is an effective way to maintain network stability and scalability, where changes in cluster membership only introduce the information update locally (i.e., in the corresponding clusters) rather than globally (i.e., the entire network), thereby lessening the overhead of message exchanges [119]. In [68], an adaptive clustering algorithm is proposed to provide guaranteed QoS to real-time multimedia traffic in a decentralized manner. Similar work with power control is also presented in the recent literature (e.g., [37,116]). In the aforementioned work, inter-cluster communications are facilitated by means of code divi-

sion multiple access (CDMA). However, channel allocation and interference control are not taken into consideration, thereby hindering system throughput improvement. Besides, the preceding work focuses on small-scale ad hoc networks, where the issue of frequency reuse is not addressed properly. Other node clustering schemes based on different system metrics (i.e., node degree and cluster size) are proposed for MANETs [119]. Most of the existing schemes aim at maximizing network connectivity and minimizing energy consumption; however, packet-level QoS provisioning is often neglected. Thus, applying those node clustering algorithms directly to WMNs can be ineffective or inefficient to support different multimedia applications (e.g., voice, video, and data) with diverse QoS requirements. In addition, increasing throughput is the key to the success of providing a robust wireless mesh backbone in large-scale WMNs. Therefore, the notion of frequency reuse and QoS provisioning should be taken into account in designing a node clustering algorithm for WMNs.

In the literature, there exists a wide range of channel assignment schemes for wireless networks [5,13,30,46,57,62,80]. The objective of most existing techniques is to achieve optimal channel allocation [13,30,57,80]. In the context of cellular systems, channel allocation is relatively straightforward, thanks to the robust hexagonal cellular structure [57]. For fixed channel allocation, a channel set is assigned to every cell. In order to deal with the traffic load variations, channel re-assignment and negotiation (e.g., channel borrowing [53]) can be performed with the help of mobile switching centers. To further attain high channel efficiency, dynamic channel allocation schemes are proposed (e.g., dynamic channel selection [92]), where centralized control is usually needed to ensure the effectiveness of frequency reuse at the cost of computational complexity. Since the mesh routers in a large-scale WMN are likely scattered around, the clustered WMN is not expected to be as structured as its aforesaid counterpart. In addition, mobile switching centers are not always available in WMNs, especially in suburban and rural residential areas. Therefore, simply applying existing channel allocation schemes designed for the cellular systems to large-scale WMNs can plausibly

degrade the system performance. On the other hand, centralized algorithms are devised to maximize the number of transmission links in the network and balance the traffic load among different channels in [30] and [13], respectively. For the sake of practical implementation, distributed schemes based on an interference conflict graph are proposed [5,46,57,62]. With the set of vertices representing the transmission links in a wireless network and the set of edges representing the transmission conflicts, channel allocation is performed in such a way that the adjacent links (or vertices) in an interference conflict graph cannot use the same channel(s) for packet transmissions. However, most of the existing approaches are not QoS-aware. In addition, packet scheduling deduced by the interference conflict graph may not guarantee collision-free transmissions in a large-scale WMN, for the aggregate interference coming from the transmissions outside the neighborhood can be very large. As such, the schedule deduced by an interference conflict graph can discourage feasible concurrent transmissions and/or reduce the transmission rates of adjacent nodes. *TiMesh* is proposed to avoid a *ripple effect* [80], but hampers the notion of frequency reuse, underutilizing the scarce radio resources. Thus, combining topology control with distributed channel allocation tailored for WMNs is necessary.

In this Chapter, we propose a QoS-aware node clustering algorithm with effective tax-based subcarrier allocation tailored for WMNs. Our novel resource allocation approach achieves Pareto optimality and outperforms the approach using an interference conflict graph.

3.3 Joint QoS-Aware Node Clustering and Tax-Based Subcarrier Allocation

In this section, a QoS-aware node clustering algorithm with tax-based subcarrier allocation tailored for WMNs is proposed. As increasing throughput in a mesh backbone is vital, the objective of the joint node clustering and subcarrier allocation problem is chosen to maxi-

mize the system throughput. Nonetheless, the node clustering problem and the subcarrier allocation problem are coupled. To reduce computational complexity, our approach is to solve the node clustering problem and to allocate subcarriers in succession alternatively, to be described in Section 3.3.4.

Considering various aspects such as system capacity, QoS provisioning, the burden on clusterheads, packet delivery delay, and the austere suburban/rural environment, we formulate the node clustering problem by setting an upper bound on the number of subcarriers allocated to a cluster. Denote $B^{\max}$ as the upper bound of the number of allocated subcarriers in a cluster. The bound caps the maximum traffic load in a cluster and hence the burden on a clusterhead, controls a cluster size to a certain extent, and facilitates packet scheduling assisted by clusterheads. In fact, as we will discuss in Section 3.5, choosing the value of this upper bound carefully can further improve system performance. Our proposed approach for the joint problem of node clustering and subcarrier allocation includes: 1) neighbor discovery; 2) initial path establishment; 3) traffic load estimation; 4) node clustering algorithm; and 5) subcarrier allocation.

3.3.1 Neighbor Discovery

A mesh router can discover its neighbor(s) via any routing protocol. In this research, a neighbor is identified if the channel gain between the node of interest and the neighboring node is above a certain threshold. As mesh routers are considered stationary, all the possible paths of a mesh router to the gateway can be predetermined.

3.3.2 Initial Path Establishment

Establishing an initial path is imperative to facilitate traffic load estimation and hence QoS-aware node clustering. Denote p_m^e as the e^{th} link along the p^{th} shortest path of the m^{th}

node to the gateway. The initial path of a mesh router to the gateway is determined by the following condition. Choose the path p^* that can maximize the end-to-end rate of the m^{th} node.

$$p^* = \arg\max_{p}\{\min_{\forall e \in p}\{\widetilde{R}_m(p_m^e)\}\}, \forall m \tag{3.1}$$

where $\widetilde{R}_m(y) = \log_2\left(1 + \widetilde{G}_m(y)P_m^{\max}/\eta\right)$ can be viewed as a data rate obtained at the m^{th} node on the y^{th} link with $\widetilde{G}_m(y)$ being the (constant) channel gain over the available radio spectrum on the y^{th} link, $P_m^{\max}$ the maximum power constraint of the m^{th} node, and η the noise power. The effect of co-channel interference is discussed in Section 3.3.5. The objective of (3.1) is to find the shortest path of a node to the gateway which maximizes the minimum of all link rates (i.e., maximize the end-to-end rate). This initial path establishment criterion ends in a tree architecture for a WMN (see Figure 3.1). For simplicity, the shortest paths are used as an example to demonstrate node clustering and subcarrier allocation.

3.3.3 Traffic Load Estimation

To facilitate resource allocation at the MAC layer, traffic load estimation is necessary [18]. Each node estimates the traffic load of real-time traffic by averaging the rate requirement over a fixed observation window (e.g., 100ms). The sum of the local traffic load estimate of a node and the relay load received is forwarded to its next-hop neighbor (toward the gateway) determined by its initial established path. Let X_m be the local traffic load estimate at the m^{th} node, and Y_m be the relay load received by the m^{th} node. Figure 3.2 illustrates how the traffic load estimation works. Then, $Y_u = X_w + Y_w$, where the u^{th} node is the next-hop neighbor (toward the gateway) of the w^{th} node. We assume that there is no local traffic generation at the gateway. Such traffic load estimation is also shown to be crucial for supporting QoS in intra-cluster resource allocation [21]. Details of intra-cluster resource allocation are given in Chapter 4.

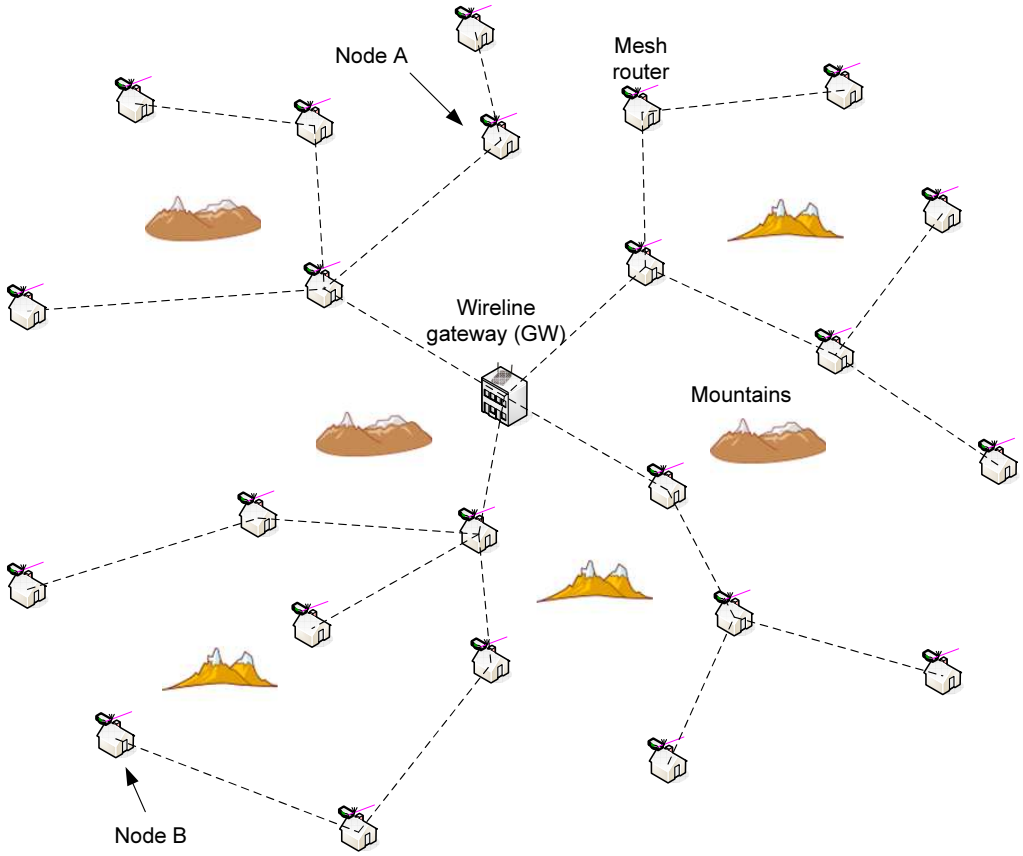


Figure 3.1: An illustration of the network topology under consideration.

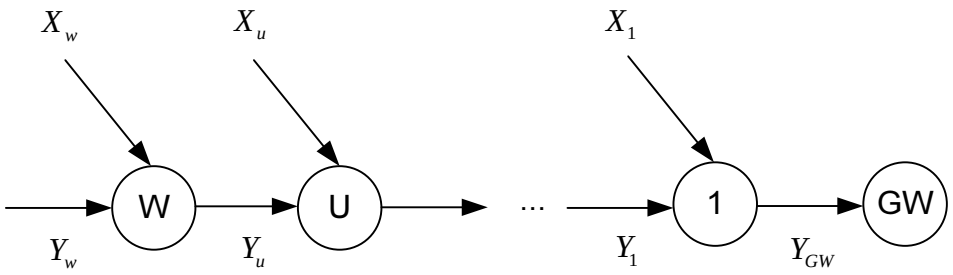


Figure 3.2: Traffic load estimation.

3.3.4 Node Clustering Algorithm

We assume that clusterheads can operate in a dual-power mode [118], where higher power levels are reserved for inter-cluster signaling[1] and lower power levels are dedicated to intra-cluster communications. Define an I-tier cluster as a cluster formed at the I^{th} clustering level from the gateway. The proposed node clustering algorithm is described as follows.

Step 1: All nodes are set to be unassigned (i.e., neither clusterheads nor clustermembers). The gateway is set to be the default clusterhead. Set I=1.

Step 2: The clusterhead of interest selects one of its 1-hop unassigned nodes[2] and collects its QoS requirement (i.e., traffic load demand). If no neighbors can be selected, go to Step 5.

Step 3: The clusterhead of interest chooses the best available subcarrier(s) for the selected

[1]In case direct inter-cluster transmissions fail, we assume that cooperative communications [15,16] are in place such that the inter-cluster signaling is always feasible.

[2]For simplicity, we only consider the case where a clusterhead selects its 1-hop neighbors as its clustermembers in this work.

node based on the subcarrier allocation criterion given in (3.5), to be discussed in Section 3.3.5. If the QoS requirement of the selected node cannot be satisfied, go to Step 5.

Step 4: If the total number of subcarriers acquired is less than or equal to $B^{\max}$ in the cluster of interest, the selected node becomes a clustermember of that cluster, the chosen subcarriers are recorded in a table stored at the clusterhead of interest, and go back to Step 2.

Step 5: An I-tier cluster is created. The clusterhead of interest keeps selecting the best available subcarrier(s), if feasible, until the total number of chosen subcarriers is $B^{\max}$ so as to further improve both the total throughput and interference tolerance of that cluster. Repeat Steps 2-4 with another I-tier clusterhead, if any, until no more clusters can be formed at the I-tier.

Step 6: The set of $(I+1)$-tier clusterhead(s) is chosen by the J-tier clusterhead(s) by means of Black-Burst jamming [18][3], where $J = 1, 2, \ldots, I$. The J-tier clusterhead(s) is(are) to signal its(their) unassigned neighbors, and the closest unassigned neighbor(s) then transmit its(their) Black-Burst jamming signal. Note that the length of a Black-Burst jamming signal is a decreasing function of the smallest number of hops from a node to the gateway. Therefore, with the gateway as a coordinator, the unassigned node(s) with the longest Black-Burst jamming signal (i.e., the smallest number of hops to the gateway) *win(s)* the contention and is(are) chosen to be the $(I+1)$-tier clusterhead(s), and the order of their cluster formation is randomly assigned. If a clusterhead can be elected, set $I=I+1$ and go back to Step 2.

Step 7: Any unassigned node joins its closest clusterhead(s). Any node that is a 1-hop

[3]The notion of Black-Burst jamming was originally proposed for channel contention and service differentiation in the MAC layer [18]. In our work, we use this approach to select a set of clusterheads.

neighbor of a clusterhead automatically becomes a clustermember of that cluster. Any clustermember that is a 1-hop neighbor of more than one clusterhead and any clusterhead that is a clustermember of another clusterhead can be viewed as a clustergateway. Notice that a node can have multiple roles in the network.

Step 8: If there are any subcarriers unallocated, the remaining subcarriers are allocated according to the subcarrier allocation criterion given in (3.5) in sequence, starting from the first formed clusterhead (i.e., the gateway) to the last formed clusterhead, until all subcarriers are employed. The value of $B^{\max}$ is adjusted accordingly.

Clusters closer to the gateway have higher priority, whereas the ones farther away from the gateway have lower priority. The rationale is that traffic bottlenecks are usually found at (and near) the gateway and hence those clusters close to the gateway are assigned higher priority in the proposed resource allocation approach. Besides, a clustered WMN will consist of 1- and/or 2-clusters[4], and different clusters may have different sizes. Clusterheads are usually located in the middle of their cluster. Further, clusterheads in the dual-power mode form a connected graph.

3.3.5 Tax-Based Subcarrier Allocation

Without any effective subcarrier allocation, it is plausible that large co-channel interference is introduced by unfavorable simultaneous transmissions. Consider the case where clusterheads do not exchange any information and myopically maximize the aggregate throughput in their clusters. Since there is no penalty for a cluster to use all the (chosen) subcarriers for its intra-cluster communications, the resultant co-channel interference generated to other clusters can be very large, decreasing overall system throughput. The above solution can be far from

[4]A j-cluster is defined as a subset of nodes which are mutually reachable by a path of length at most j-hops for a positive integer j.

optimal, and hence an effective resource allocation approach tailored for large WMNs is vital. In the following, we view a clusterhead as the representative of its cluster.

We propose a novel QoS-aware subcarrier allocation scheme motivated by Karush-Kuhn-Tucker (KKT) optimality conditions [9] (see Appendix-A.1). Let M denote the number of clusterheads, N the number of available subcarriers, and L the number of timeslots (i.e., DATA slots) in a frame. Consider the following payoff function of the m^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot.

$$S^l_{m,n}\left(c^l_{m,n}, p^l_{m,n}\right) = c^l_{m,n} r^l_{m,n} - p^l_{m,n} \sum_{k \neq m}^{M} T^l_{km,n} \tag{3.2}$$

where $r^l_{m,n} = \log_2(1 + \gamma^l_{m,n})$ represents the achievable data rate of the m^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot with the received signal-to-interference-plus-noise ratio (SINR) $\gamma^l_{m,n}$, $c^l_{m,n}$ is the indicator of allocating the n^{th} subcarrier to the m^{th} clusterhead on the l^{th} timeslot, i.e., $c^l_{m,n} \in \{0, 1\}$, $p^l_{m,n}$ is the transmit power, and $T^l_{km,n}$ is *tax* paid by the m^{th} clusterhead for generating interference to the k^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot, defined as

$$\begin{aligned} T^l_{km,n} &= -\frac{\partial U_k(\mathbf{c}, \mathbf{p}, R^d_k)}{\partial p^l_{m,n}} \\ &= \left(\frac{\partial U_k(\mathbf{c}, \mathbf{p}, R^d_k)}{\partial \gamma^l_{k,n}}\right) \cdot \left(-\frac{\partial \gamma^l_{k,n}}{\partial P^l_{I,k,n}}\right) \cdot \left(\frac{\partial P^l_{I,k,n}}{\partial p^l_{m,n}}\right) \\ &= \varepsilon^l_{k,n} \cdot I^l_{k,n} \cdot c^l_{m,n} G_{km,n} \end{aligned} \tag{3.3}$$

where $U_k(\mathbf{c}, \mathbf{p}, R^d_k)$ is the utility function of the k^{th} clusterhead with R^d_k being the QoS demand of the k^{th} clusterhead, $P^l_{I,k,n} = \sum_{i \neq k} c^l_{i,n} G_{ki,n} p^l_{i,n}$ being the received interference power level of the k^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot, $\mathbf{c} = [c^l_{k,n}]_{M \times N \times L}$, and $\mathbf{p} = [p^l_{k,n}]_{M \times N \times L}$. The SINR of the m^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot is given by $\gamma^l_{m,n} = \frac{\varphi G_{mm,n} p^l_{m,n}}{\sigma \sum_{k \neq m} c^l_{k,n} G_{mk,n} p^l_{k,n} + \eta}$, where $G_{mk,n}$ is the channel gain from the k^{th} clusterhead to the m^{th} clusterhead over the n^{th} subcarrier, φ a BER measure, and σ the cross-correlation factor between any two signals, i.e., $\sigma \in [0, 1]$. The utility function of the k^{th} clusterhead is

defined as

$$U_k(\mathbf{c}, \mathbf{p}, R_k^d) = \begin{cases} \sum_l^L \sum_n^N c_{k,n}^l r_{k,n}^l - R_k^d, & \sum_l^L \sum_n^N c_{k,n}^l r_{k,n}^l - R_k^d \geq \epsilon \\ 0, & \sum_l^L \sum_n^N c_{k,n}^l r_{k,n}^l - R_k^d < \epsilon \end{cases} \tag{3.4}$$

where $0 < \epsilon \ll 1$. We assume that for $0 < \sum_l^L \sum_n^N c_{k,n}^l r_{k,n}^l - R_k^d < \epsilon$, $\frac{\partial U_k(\mathbf{c},\mathbf{p},R_k^d)}{\partial \gamma_{k,n}^l} \rightarrow \infty, \forall k$. Notice that $T_{km,n}^l$ can be interpreted as the marginal decrease in the utility obtained by the k^{th} clusterhead per unit increase in the transmit power of the m^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot. Thus, $T_{km,n}^l$ is always non-negative. Similarly, $\varepsilon_{k,n}^l$ represents the sensitivity of utility obtained by the k^{th} clusterhead per unit change in the received SINR of the k^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot, and $I_{k,n}^l$ represents the marginal decrease in the received SINR of the k^{th} clusterhead per unit increase in the received interference power level over the n^{th} subcarrier on the l^{th} timeslot. With the tax interpretation of the KKT conditions, each clusterhead, therefore, essentially maximizes the difference between its throughput obtained minus its *lump-sum tax* paid to the other clusterheads in the mesh backbone due to the induced interference. Each clusterhead is to optimize $c_{m,n}^l$ and $p_{m,n}^l$ such that its own payoff function is maximized. Notice that $\mathbf{c}$ and $\mathbf{p}$ are the optimization variables. The proposed subcarrier allocation strategy is more suitable and applicable than a traditional coloring approach to WMNs, mesh routers of which are not evenly distributed. More importantly, with the tax information, subcarrier allocation and hence frequency reuse can be carried out in a decentralized manner.

In this work, we only focus on the subcarrier allocation by fixing power allocation (i.e., uniform power distribution), though subcarrier allocation and power allocation should be jointly considered for the sake of optimality. Thus, the criterion of subcarrier allocation can be deduced as follows. For the m^{th} clusterhead on the l^{th} timeslot, choose n^* such that

$$n^* = \arg\max_n \left\{ S_{m,n}^l \left(c_{m,n}^l \right) \right\} \tag{3.5}$$

and set $c_{m,n^*}^l = 1$. Note that an information exchange among clusterheads is triggered

whenever there is any change in subcarrier allocation.

3.3.6 Complexity Analysis

Consider that, with message signaling among clusterheads, the clusterheads have complete knowledge of their lump-sum tax. Since each clusterhead behaves individually in sequence, the time complexity of the proposed subcarrier allocation is on the order of $O(NL)$. Thus, the time complexity of the proposed node clustering algorithm given in Section 3.3.4 is on the order of $O(MNLB^{\max})$. With the help of effective *data structure* (e.g., binary tree implementation [112]), the complexity can be further reduced to the order of $O(MLB^{\max} \log_2 N)$.

3.4 Efficiency Evaluation by Game Theory

In this section, we show that the subcarrier allocation solution obtained from our proposed approach achieves efficient use of network resources. In game theory, efficient resource utilization is determined by the concept of *Pareto optimality*[5] [88]. Modeled by a round-robin game, our proposed QoS-aware subcarrier allocation approach (or game) also attains a *Nash equilibrium*[6] (NE) [88] in the case where all available subcarriers are active and all the clusters are heavily loaded.

[5]An action profile $\mathbf{b}^* = (b_1^*, b_2^*, ..., b_M^*)$ is said to be Pareto optimal if and only if there exists no other action profile $\tilde{\mathbf{b}}$ such that for some m, $Y_m(\tilde{\mathbf{b}}) > Y_m(\mathbf{b}^*)$ and $Y_n(\tilde{\mathbf{b}}) \geq Y_n(\mathbf{b}^*)$, for $n \neq m$, where $Y_m(\cdot)$ is a utility function of user m in the context of game theory. In words, an action profile (or resource allocation) is Pareto optimal if there exists no other action profile that makes some user(s) better off without making the other user(s) worse off.

[6]An action profile $\mathbf{b}^* = (b_1^*, b_2^*, ..., b_M^*)$ attains a Nash equilibrium (NE) if no unilateral deviation in strategy by any single user is profitable, i.e., $Z_m(b_m^*, b_{-m}^*) \geq Z_m(\tilde{b}_m, b_{-m}^*), \forall m$, where $\tilde{b}_m$ is another strategy other than b_m^* of user m, b_{-m} is an action profile of all users except for user m, and $Z_m(\cdot)$ is a payoff function of user m in the context of game theory. In words, an action profile (or resource allocation) attains an NE if no player can do better by unilaterally changing its strategy.

Proposition 1 *Suppose all available subcarriers are chosen by the clusterheads at least once on every timeslot (i.e.,* $\sum_m c^l_{m,n} \geq 1, \forall n, l$*). The subcarrier allocation solution obtained from our proposed approach is Pareto optimal.*

The proposed node clustering algorithm ensures that all the available subcarriers are to be selected at least once. Thus, from the perspective of game theory, the proposed QoS-aware subcarrier allocation approach attains Pareto optimality, and hence the resources are efficiently utilized.

Proposition 2 *Suppose all available subcarriers are active. If all the clusters are heavily loaded (i.e.,* $0 < \sum_l^L \sum_n^N c^l_{m,n} r^l_{m,n} - R^d_m < \epsilon, \forall m$*), frequency reuse is prohibited.*

Proposition 3 *Suppose all available subcarriers are active and all the clusters are heavily loaded. Modeled by a round-robin game played by the clusterheads, the proposed subcarrier allocation solution attains an NE.*

Corollary 1 *If an NE is attained, all available subcarriers are in use.*

When all the subcarriers are in use and all the clusters are fully loaded, an NE is attained (i.e., Proposition 3), leading to a stable clustered networking structure. However, under the same aforesaid conditions, the notion of frequency reuse is discouraged due to zero tolerance to any additional co-channel interference (i.e., Proposition 2), thereby reducing system throughput. The implication is that, to foster frequency reuse and hence increase system throughput, clusters should not be heavily loaded so as to allow certain interference margin. Since throughput improvement is indispensable in a mesh backbone, frequency reuse should be the first issue to be addressed in developing a resource allocation algorithm. On the other hand, in order to procure a desired balance between QoS provisioning and system throughput maximization, CAC should be in place, which limits the number of

calls admitted into the system and the number of clustermembers affiliated to a cluster. Our evaluation reinforces the fact that QoS provisioning and throughput maximization are conflicting with each other [22]. More importantly, our study also reveals a crucial principle that how to allocate resources to the wireless nodes in a decentralized fashion is critical in determining the solution space of a performance tradeoff between QoS provisioning and throughput maximization. For instance, a clusterhead should not select too many nodes as its clustermembers; otherwise, the effectiveness of frequency reuse and hence the system throughput can be reduced. Nonetheless, as long as all the subcarriers are active, our proposed subcarrier allocation approach achieves Pareto optimality, making efficient use of network resources.

3.5 Performance Evaluation

3.5.1 Simulation Environment

We consider a WMN with nodes randomly distributed over the network coverage area, with a node density of 1 node per 1000m^2. Parameters for performance evaluation are chosen as follows: $\varphi = 1$, $\sigma = 1$, $\eta \sim N(0, 10^{-12}\text{W})$, $P_m^{\max} = 1\text{W}$, $\forall_m$, and $L = 4$. The maximum transmission rate of each subcarrier is 100kb/s. We adopt the path loss model suggested in [44]:

$$PL = A + 10\alpha \log_{10} \frac{d}{d_0} + s, \quad d > d_0 \tag{3.6}$$

where $A = 20 \log_{10}(4\pi d_0/\lambda)$ with λ being the subcarrier wavelength in meters, α is the path loss exponent with $\alpha = a - bh_b + c/h_b$ with h_b being the height of an antenna in meters (i.e., 10m $\leq h_b \leq$ 80m), and a, b, and c being constants dependent on the terrain category, d_0 is the reference distance (e.g., 100m), and s is the shadow fading term capturing the effect of shadowing, which follows a log-normal distribution [44]. In this research, we choose the

terrain category A (i.e., hilly/moderate-to-heavy tree density) investigated in [44], where $\lambda = 0.1579$m, $a = 4.6$, $b = 0.0075\text{m}^{-1}$, $c = 12.6$m, $h_b = 10$m, and $\sigma_s = 10.6$dB. We employ the traffic models for voice, video, and data described in Section 2.2.2: Voice traffic is generated according to a two-state ON-OFF model. In the ON state, a fixed-size packet arrives at a constant rate, whereas in the OFF state, no packet is generated. The duration of an ON period and that of an OFF period both follow an exponential distribution. For simplicity, video traffic is characterized by a two-state ON-OFF model, where different incoming packets have different packet sizes, generating a variable-rate traffic in the ON state [39]. Data traffic is the background best-effort traffic and available anytime, which is assigned the lowest priority and does not have any rate requirements. Data packets can be transmitted whenever there are available resources. We consider that there are a voice source, a video source, and a data source residing at every node. We use the MAC protocol described in Section 2.3 for both intra-cluster and inter-cluster communications. The duration of a timeslot is 5ms, and hence the duration of a frame is $5(2+L)$ms. For intra-cluster MAC, each clusterhead collects the transmission requests and traffic demands from its clustermembers by polling them periodically in the control slot, and announces the resource allocation in the subsequent beacon. The polling is done in every 100ms. Similarly, for inter-cluster MAC, a clusterhead collects the transmission requests from its neighboring clusterhead(s) in the control slot, and announces the transmission schedule in the subsequent beacon. The timeslots and subcarriers allocated to a particular node are reserved for packet transmissions of real-time traffic in the DATA slots until the next polling. Moreover, for real-time traffic (i.e., voice and video), the higher the packet dropping rate that a traffic flow experiences, the higher the priority of the packets associated with that flow. A clusterhead grants the requests of those higher-priority packets first, facilitating QoS provisioning. Both the polling and the beacon packet transmissions are assumed error-free. We also carry out an offline CAC, where we consider that the maximum number of admitted voice calls and that of admitted video

Table 3.2: Relationship of the number of wireless nodes and the average number of clusterheads

Number of wireless nodes	25	50	100	200	400
Average number of clusterheads	8.1	17.2	33.7	67.1	98.3

calls are 13 and 12, respectively. More simulation details for voice and video traffic can be found in Chapter 4. Here, we focus on the impact of our proposed node clustering with tax-based subcarrier allocation on the system performance. We perform the simulations for 10,000 runs and average the results, where each simulation run sustains 5,000 frames. The performance measurements used in our simulations are 1) system throughput, 2) frequency reuse ratio, and 3) packet delay.

3.5.2 QoS-Aware Node Clustering

We first study the performance of our proposed QoS-aware node clustering algorithm with $N = 1024$ and $B^{\max} = N/4$. After discovering the neighbors, every node establishes an initial shortest path to the gateway according to the condition given in (3.1). Given an established initial path, each node can estimate the traffic load. The proposed node clustering algorithm is then carried out. Figure 3.3 illustrates a simulation result for the cluster structure of a WMN with 25 nodes. Since a clusterhead selects some of its 1-hop neighbors as its clustermembers in this work, the WMN only consists of 1- and 2-clusters. The relationship of the number of nodes and that of clusterheads is given in Table 3.2.

We then compare the system performance in terms of packet delay and throughput of the proposed scheme for the initial path establishment (named *proposed*) with the scheme for random shortest path establishment (named *random*) and the scheme for path establishment that gives the maximum end-to-end rate among all possible paths (named *max-rate*).

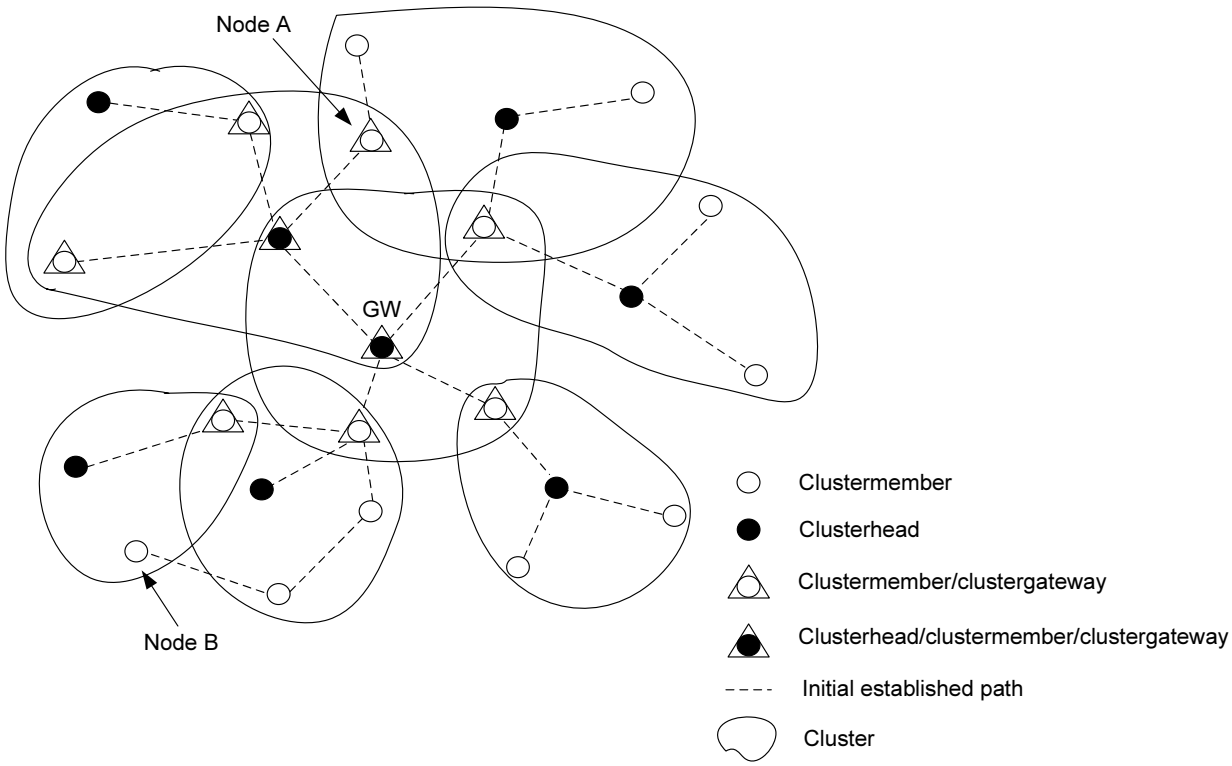


Figure 3.3: An illustration of a clustered WMN with 25 nodes.

The delay-throughput performance comparison of the three schemes for Node A and Node B (shown in Figure 3.3) is depicted in Figure 3.4. Notice that the same node clustering algorithm with subcarrier allocation is employed for the three schemes. As seen, using the max-rate scheme for path establishment, both voice packets at Node A and video packets at Node B attain the largest end-to-end data rate (or throughput) at the cost of packet delay. Although both the proposed and random schemes result in the smallest packet delays, the proposed approach achieves higher end-to-end rates. We observe that the improvement of the proposed approach over the random approach, however, is not substantial because routing is not taken into consideration. The improvement is expected to be larger if node clustering, subcarrier allocation, and routing are jointly considered; addressing this issue is left for further work, though. Nonetheless, the proposed condition for initial path establishment results in a good tradeoff between the packet delay and the end-to-end rate. Other system performance (e.g., packet dropping rates) in regards to intra-cluster resource allocation are reported in Chapter 4 (and in [23]).

3.5.3 Tax-Based Subcarrier Allocation

Here, we evaluate the performance of the proposed tax-based subcarrier allocation in terms of system throughput and frequency reuse ratio in a clustered WMN with $N = 1024$ and $B^{\max} = N/4$. For comparison, we consider a baseline approach where there is no frequency reuse and an approach using an interference conflict graph [46]. Notice that, for the approach using an interference conflict graph, the adjacent links (or vertices) in an interference conflict graph cannot use the same subcarrier(s). In the simulations, the same node clustering algorithm and the same value of $B^{\max}$ are applied to all these approaches. To further validate our simulation results, an upper bound of system throughput performance is also plotted for reference, which is obtained by an exhaustive search. Notice that this upper bound is the maximum achievable system throughput achieved by the clusters without considering QoS

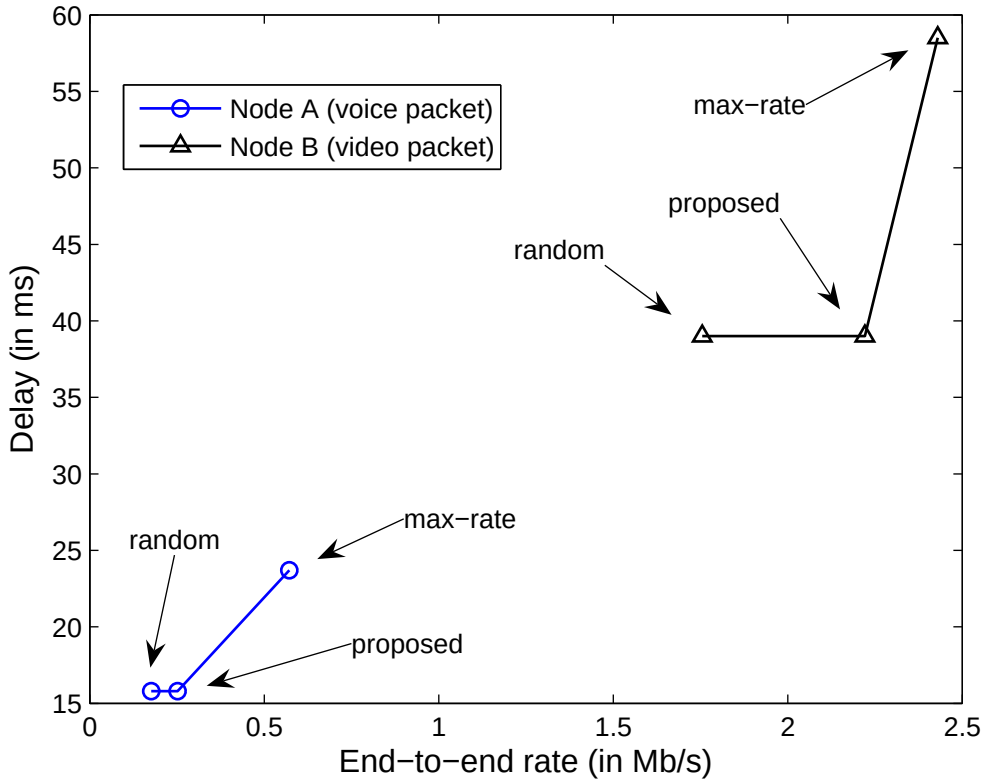


Figure 3.4: Packet delays of the real-time traffic using the proposed path establishment scheme, the random path establishment scheme, and the max-rate path establishment scheme vs. the end-to-end rate for Node A and Node B (shown in Figure 3.3).

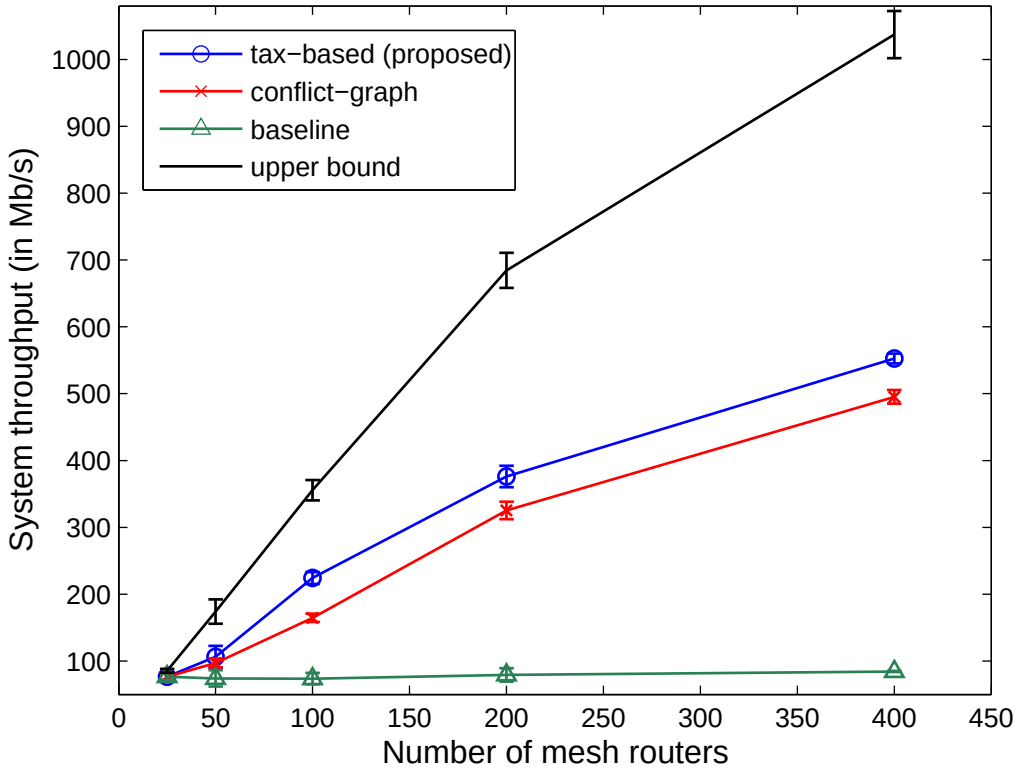


Figure 3.5: System throughputs of the proposed approach, the conflict-graph approach, the baseline approach, and the upper bound vs. the number of mesh routers (where $N = 1024$ and $B^{\max} = N/4$).

constraints (i.e., best-effort traffic only).

Effect of the number of mesh routers, M

Figure 3.5 shows the system throughput versus the number of mesh routers. The standard deviations of the results are also plotted for reference. The system throughput curves for all three approaches are very close when the number of mesh routers is small, which is due to the fact that the network size is small and very few subcarriers can be reused in a

small WMN. As the number of mesh routers increases (i.e., a larger WMN), our proposed approach clearly achieves a higher system throughput than the other two approaches, thanks to increased frequency reuse. The rationale of our proposed approach being superior to the conflict-graph approach stems from the fact that the interference conflict graph merely yields a condition that adjacent vertices cannot use the same subcarrier(s), thereby suppressing the potential and favorable concurrent transmissions. In contrast, the tax-based subcarrier allocation fosters frequency reuse to a greater extent. We also observe that, on average, our proposed approach conduces to a higher frequency reuse ratio than the conflict-graph approach (see Figure 3.6). Notice that the difference between the throughput obtained by the tax-based approach and that by the conflict-graph approach becomes more substantial when the number of subcarriers increases, as discussed in Section 3.5.3. Without frequency reuse, the system throughput of the baseline approach is almost the same against the number of mesh routers, for each subcarrier can only be allocated once, leading to the worst throughput performance. The upper bounds for the system throughput obtained and frequency reuse ratio are plotted for reference in Figure 3.5 and Figure 3.6, respectively. As observed, there is an obvious performance gap between the system throughput obtained from the proposed approach and the upper bound. The performance disparity is ascribed to the fact that QoS provisioning and throughput improvement are conflicting with each other [22]. By taking QoS provisioning into account, some of the resources are reserved for the real-time traffic in the proposed approach, resulting in lower system throughput. On the other hand, our results show that frequency reuse is crucial for increasing the system throughput. As such, the proposed tax-based subcarrier allocation can better utilize the network resources, thereby giving rise to a radio spectrum efficient WMN.

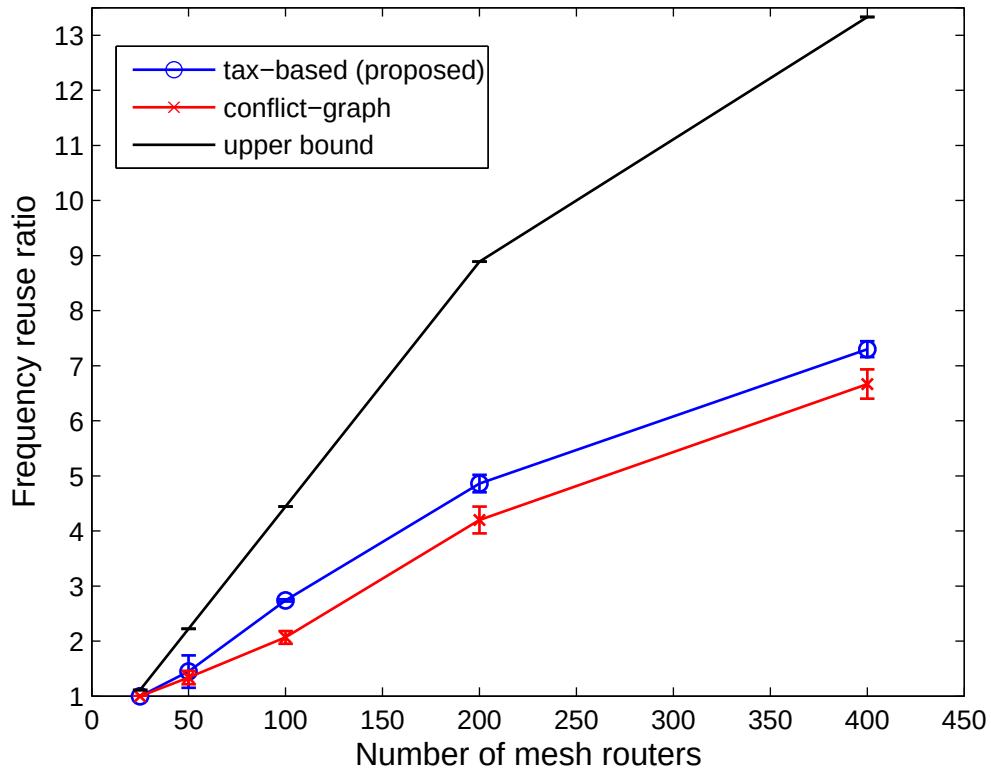


Figure 3.6: Frequency reuse ratios of the proposed approach, the conflict-graph approach, and the upper bound vs. the number of mesh routers (where $N = 1024$ and $B^{\max} = N/4$).

Effect of the number of subcarriers, N

We consider the effect of the number of available subcarriers on the system performance by fixing the number of mesh routers to be 100 and $B^{\max} = N/4$. Figure 3.7 shows the system throughput of all approaches versus the number of subcarriers. Since the number of subcarriers increases, the system throughput of all the approaches increases. As mentioned previously, the tax-based approach performs the best due to the increased frequency reuse. The upper bound of the system throughput is also plotted for reference. We also observe that the frequency reuse ratio achieved by the proposed algorithm is more or less the same with the number of subcarriers (i.e., at the level of 2.8). It shows that the frequency reuse ratio of the proposed approach is almost independent of the number of subcarriers available in the system.

Effect of $B^{\max}$

We investigate the system performance with different values of $B^{\max}$ by setting $N = 1024$ and the number of mesh routers to be 100. Figures 3.8 and 3.9 depict the frequency reuse ratio and the system throughput versus the value of $B^{\max}$, respectively. Regarding the proposed approach, when $B^{\max}$ is small, each cluster gets only a handful of subcarriers. Hence, the chance of reusing the subcarriers is smaller, resulting in a smaller frequency reuse ratio and system throughput. When $B^{\max}$ becomes larger, more subcarriers can be reused and so the system throughput increases greatly. However, when $B^{\max}$ reaches a certain value, both the frequency reuse ratio and system throughput start to drop. The reason is that too many subcarriers are allocated to a cluster, which causes some neighboring clusters to choose other available subcarriers and hence reduces the effectiveness of frequency reuse. In fact, our results are comparable to the differences between fixed channel allocation (FCA) and dynamic channel allocation (DCA) in the context of cellular systems. Our proposed subcarrier allocation behaves like FCA when $B^{\max}$ is small and DCA when $B^{\max}$ is large.

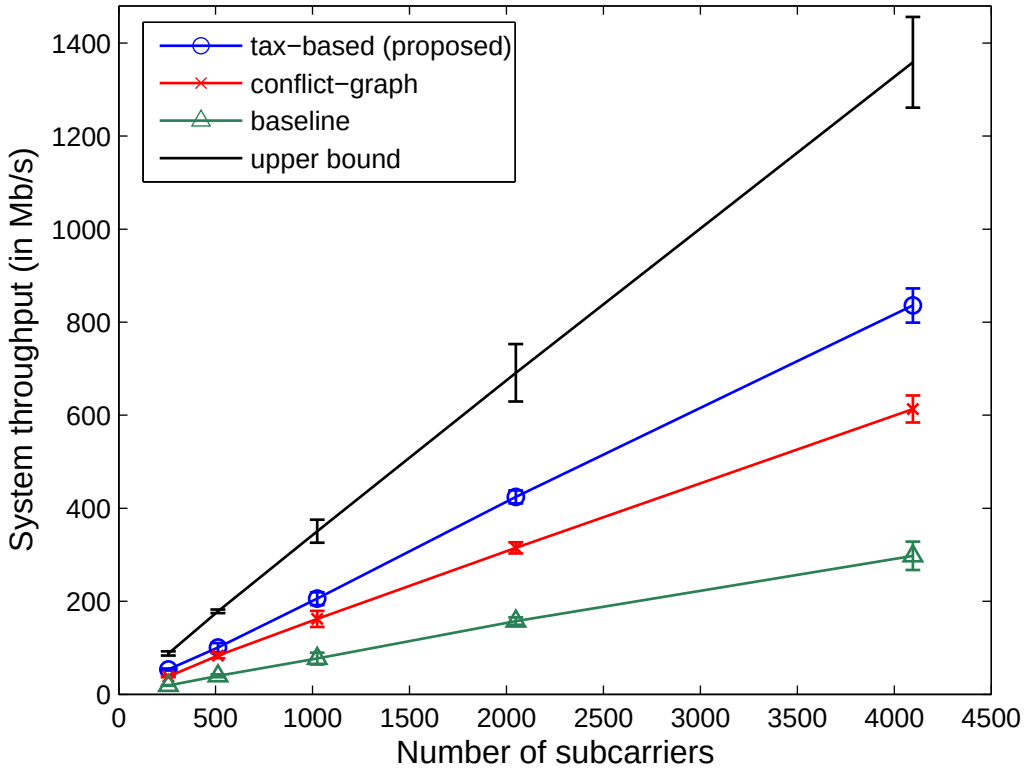


Figure 3.7: System throughputs of the proposed approach, the conflict-graph approach, the baseline approach, and the upper bound vs. the number of subcarriers (where the number of mesh routers is 100 and $B^{\max} = N/4$).

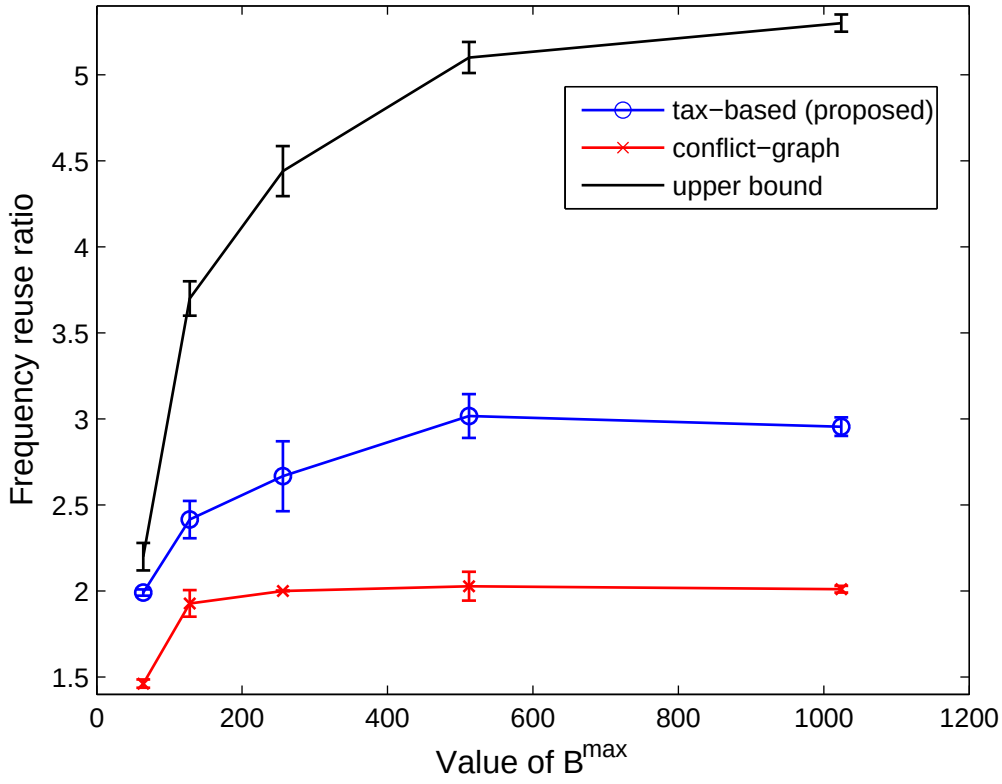


Figure 3.8: Frequency reuse ratios of the proposed approach, the conflict-graph approach, and the upper bound vs. the value of $B^{\max}$ (where the number of mesh routers is 100 and $N = 1024$).

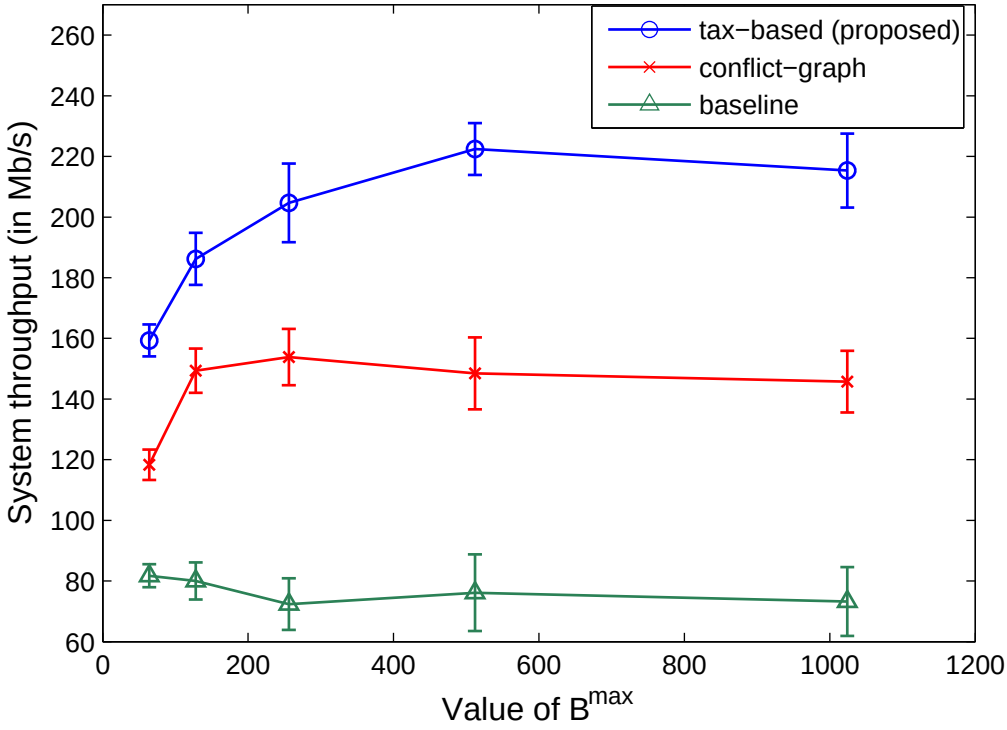


Figure 3.9: System throughputs of the proposed approach, the conflict-graph approach, and the baseline approach vs. the value of $B^{\max}$ (where the number of mesh routers is 100 and $N = 1024$).

It is well-known that, in general, DCA exhibits superior performance compared to FCA in the case of non-uniform traffic distribution and/or low traffic load [57]. However, the performance of DCA degrades under heavy load conditions (e.g., large $B^{\max}$). Hybrid channel allocation (HCA) is shown to be the best performance compromise [57]. In our work, the traffic load is not evenly distributed and hence the system throughput improves when $B^{\max}$ increases. On the other hand, the traffic load in a cluster also increases with $B^{\max}$, as more clustermembers can be selected. When the traffic load reaches a certain threshold, the system throughput starts to decrease. Therefore, the value of $B^{\max}$ should be chosen carefully in order to achieve high system throughput. Concerning the system throughput obtained by the conflict-graph approach, it first goes up sharply as more subcarriers can be reused. As $B^{\max}$ increases, however, the system throughput of the conflict-graph approach decreases, although the frequency reuse ratio remains roughly the same from $B^{\max} = 128$ onward. That decline is based on the (large) aggregate interference generated from the transmissions outside the neighborhood. This phenomenon reasserts our argument mentioned in Section 3.2 that the resource allocation solution deduced by a conflict-graph approach can reduce the system throughput in a large WMN. Thus, for the same frequency reuse ratio, in general, the system throughput obtained by the conflict-graph approach is less than that obtained by the proposed tax-based approach. An upper bound is also plotted for reference in Figure 3.8. Since more subcarriers can be chosen in a cluster, the curve goes up with the value of $B^{\max}$, the idea of which is similar to the notion of *multi-user diversity*. On a different note, we observe that, when $B^{\max}$ is very small or very large, some clusters are "starved or almost starved", where no or very few subcarriers are allocated to those (lower priority) clusters. The value of $B^{\max}$, therefore, should be carefully chosen to ensure that a clustered WMN can operate efficiently.

3.6 Discussion

To further enhance the system performance for WMNs, network-layer routing and MAC-layer packet scheduling should be jointly taken into consideration. In the case of a single gateway, routing is imperative to avoid traffic congestion at (and near) the gateway by way of load balancing, plausibly achieving an improved performance. However, obtaining optimal solutions for the joint problem of routing and MAC-layer resource allocation is, in general, computationally expensive. As such, devising low-complexity resource allocation strategies is needed. Another effective way to attain a desired traffic load distribution is to deploy multiple gateways, leading to the issue of gateway deployment. With efficient gateway deployment, co-channel interference can be reduced greatly and, therefore, system capacity can be further increased. In practice, to lower the costs of gateway deployment, the number of gateways deployed should be kept minimal but large enough to provide users satisfactory (perceived) QoS. Further, the mobility of gateways can be employed to cater to the variations of traffic loads and/or channel conditions. Under what conditions can the mobility of gateways be beneficial remains an open research problem. On the other hand, we consider that a clusterhead selects its 1-hop neighbors only in this work. To further reduce the signalling overhead among clusters, we should allow clusterheads to choose any of their nearby neighbors (e.g., 2-hop neighbors) so long as the QoS requirements of their clustermembers can be met at the cost of packet delay. Attaining a desired balance between complexity and system performance is important yet challenging. Addressing the aforesaid research issues, however, is left for further work.

3.7 Chapter Summary

In this Chapter, we propose a novel resource allocation scheme for the problem of joint node clustering and subcarrier allocation in WMNs with QoS assurance. The proposed node clus-

tering algorithm is QoS-aware, and the proposed tax-based subcarrier allocation is shown to effectively enhance frequency reuse and increase the system throughput. Our approach is also shown to provide a good performance balance between packet delay and end-to-end data rate for real-time traffic, leading to a viable candidate for practical implementation. The proposed resource allocation solution is Pareto optimal and hence utilizes network resources efficiently. In addition, our analysis reveals that how to allocate resources in a decentralized fashion affects a performance balance between QoS provisioning and throughput maximization. Simulation results show that our proposed approach outperforms a baseline approach and a conflict-graph approach. Our findings also demonstrate that the value of $B^{\max}$ should be determined carefully in order to further enhance the system performance.

Chapter 4

Joint Power-Subcarrier-Time Resource Allocation with Effective QoS Provisioning

4.1 Introduction

As mentioned in Chapter 1, joint power-subcarrier-time resource allocation is imperative due to the necessity of packet scheduling for QoS provisioning, multi-channel communications, and opportunistic power allocation. With the help of clusterheads, collision-free scheduling is feasible within a cluster, facilitating QoS provisioning and system throughput improvement. After a set of subcarriers are allocated to a cluster according to our tax-based subcarrier allocation proposed in Chapter 3, we focus on intra-cluster resource allocation in this Chapter. In specific, we address the issues of power allocation, subcarrier allocation, packet scheduling, and QoS support. The contributions and significance of this research work are three-fold [21,23].

- First, we study the problem of power allocation, subcarrier allocation, and packet scheduling for WMNs. We prove that the joint power-frequency-time resource allocation problem is NP-hard. Plus, we propose three QoS-aware intra-cluster packet-level resource allocation approaches, namely a Karush-Kuhn-Tucker (KKT)-driven, a genetic algorithm (GA)-based approach, and a hybrid KKT-GA approach.

- Second, we compare our proposed approaches with several counterparts suggested in the literature. Our results show that all the newly proposed approaches are demonstrated effective in QoS provisioning. In specific, the GA-based approach outperforms all the other approaches in terms of throughput, while the KKT-driven approach outperforms its counterparts of the same time complexity. Our hybrid approach is shown to achieve a desired balance between system performance and time complexity, resulting in a preferred candidate for practical implementation. We also analytically derive the bounds for the throughputs obtained by real-time and non-real-time traffic, serving as performance benchmarks.

- Third, our study shows that effective CAC not only can guarantee the QoS support of admitted multimedia calls in the system, but also can sustain the throughput performance. Besides, we observe that, in WMNs with background (bursty) data traffic, admitting more multimedia calls can decrease (increase) the system throughput.

The remainder of this Chapter is organized as follows. Related work is given in Section 4.2. The problem formulation of joint power-frequency-time resource allocation is presented in Section 4.3. Three resource allocation approaches are proposed in Section 4.4. Performance evaluation is given in Section 4.5. A brief discussion is provided in Section 4.6. Finally, a summary is drawn in Section 4.7. A summary of important symbols used in this Chapter is given in Table 4.1 for easy reference.

Table 4.1: Summary of important symbols used in this Chapter.

Symbol	Definition
$U_m(R_m(\mathbf{c},\mathbf{p}))$	utility function of the m^{th} link
$R_m(\mathbf{c},\mathbf{p})$	actual (aggregate) transmission rate of the m^{th} link over a frame
R_m^d	instantaneous transmission rate demand of the m^{th} link in the current frame
$c_{m,n}^l$	indicator of allocating the n^{th} subcarrier to the m^{th} link on the l^{th} timeslot
$p_{m,n}^l$	transmit power over the n^{th} subcarrier of the m^{th} link's transmitter on the l^{th} timeslot
$G_{mk,n}^l$	channel gain from the k^{th} link's transmitter to the m^{th} link's receiver over the n^{th} subcarrier on the l^{th} timeslot
I_n	interference power at the n^{th} subcarrier
$g_{mk,n}^l$	effective channel gain, i.e., $g_{mk,n}^l = \varphi G_{mk,n}^l/(I_n+\eta)$
η	background noise power
φ	BER measure
$P_m^{\max}$	maximum power constraint of the m^{th} link's transmitter
M	number of active links in a cluster
N	number of subcarriers allocated to a cluster
L	number of timeslots in a frame
T	number of iterations
S	population size
FF_i	fitness function of the i^{th} individual

4.2 Related Work

The problem of subcarrier-bit-power allocation for OFDM systems has been researched extensively with respect to the physical layer [33,54,85,98,102,114,117]. Due to its computational hardness, the joint optimization problem is usually decoupled into several subproblems [117]. In [117], the objective function is to maximize the total achievable rate. However, with a Hungarian approach, the complexity is at least on the order of $O(N^3)$, where N is the number of subcarriers. QoS demands and fairness constraints are taken into account in [54], where heuristic schemes are proposed for convex optimization problems (e.g., the signal-to-noise ratio (SNR) maximization problem). In [33], a heuristic resource allocation scheme is proposed for the uplink OFDMA systems, taking fairness and time-varying fading into consideration. In [102], a two-level resource allocation scheme for downlink OFDMA systems

is proposed to achieve Nash bargaining fairness. A Lagrangian-based approach is employed in [114] to solve the total transmit power minimization problem, but the high complexity of the preceding algorithm impedes its practical implementation. In order to reduce the computational complexity, KKT interpretations are popularly employed in designing resource allocation algorithms (e.g., [85,98] for utility maximization in uplink OFDMA/WiMAX systems). In the previous work, packet scheduling, however, is not addressed properly. Packet-level QoS provisioning is often neglected in the MAC layer. Thus, applying these existing resource allocation schemes directly to the WMNs with heterogeneous (packet-level) QoS requirements can be ineffective or inefficient. In addition, those schemes derived from the theory of convex optimization may not be efficient or even applicable to non-convex optimization problems.

In this Chapter, we devise three efficient yet effective strategies to solve the joint power-subcarrier-time resource allocation problem for WMNs. Our proposed approaches are demonstrated promising in packet-level QoS provisioning and system throughput improvement.

4.3 Joint Power-Subcarrier-Time Resource Allocation

In this section, we consider a WMN divided into a number of clusters (see Figure 2.1). In a cluster, one node is selected as a clusterhead according to our proposed node clustering algorithm discussed in Chapter 3. As mentioned in Section 2.1, time is partitioned into frames. Each frame is further divided into a beacon slot, a control slot, and a number of DATA slots. The beacon is used to provide timing and cluster information, and broadcast scheduling decisions for the DATA slots. In the control slot, the clusterhead collects the requests from its clustermembers, and announces the resource allocation in the subsequent beacon.

There are various system constraints associated with the joint power-subcarrier-time

resource allocation problem. The sum of the transmit power of each node on the allocated subcarriers is bounded by a maximum power level:

$$p_{m,n}^{l} \geq 0, \forall_{m,n,l} \tag{4.1}$$

$$\sum_{n=1}^{N} p_{m,n}^{l} = P_{m}^{\max}, \forall_{m,l} \tag{4.2}$$

where $p_{m,n}^{l}$ is the transmit power over the n^{th} subcarrier of the m^{th} link's transmitter on the l^{th} timeslot and $P_{m}^{\max}$ is the maximum power constraint of the m^{th} link's transmitter. Concerning subcarrier allocation, we consider the case where each subcarrier can only be allocated to one transmission link in a cluster (i.e., exclusive subcarrier allocation). In this work, time sharing of subcarriers in a DATA slot is not considered. The constraints with respect to subcarrier allocation can be formulated as follows:

$$\sum_{m=1}^{M} c_{m,n}^{l} = 1, \forall_{n,l} \tag{4.3}$$

$$c_{m,n}^{l} \in \{0, 1\}, \forall_{m,n,l} \tag{4.4}$$

where $c_{m,n}^{l}$ is the indicator of allocating the n^{th} subcarrier to the m^{th} link on the l^{th} timeslot. Different traffic types require different packet transmission rates. In our problem formulation, we take the instantaneous rate requirements of different traffic types (e.g., voice, video, and data) in the current frame, if any, into account:

$$R_{m}(\mathbf{c}, \mathbf{p}) \geq R_{m}^{d}, \forall_{m} \tag{4.5}$$

where R_{m}^{d} is the instantaneous transmission rate demand of the m^{th} link in the current frame and $R_{m}(\mathbf{c}, \mathbf{p}) = \sum_{l=1}^{L} \sum_{n=1}^{N} c_{m,n}^{l} \log_2 \left(1 + g_{mm,n}^{l} p_{m,n}^{l}\right)$ represents the actual aggregate transmission rate of the m^{th} link over a frame with $\mathbf{c} = [c_{m,n}^{l}]_{M \times N \times L}$, $\mathbf{p} = [p_{m,n}^{l}]_{M \times N \times L}$, and $g_{mk,n}^{l} = \varphi G_{mk,n}/(I_n + \eta)$ with $G_{mk,n}$ being the channel gain from the k^{th} link's transmitter to the m^{th} link's receiver over the n^{th} subcarrier, I_n the interference power at the n^{th} subcarrier, φ a BER measure, and η the background noise power. Notice that QoS provisioning is

to be handled by MAC-layer packet scheduling (as discussed in Section 2.3). With the above constraints, we employ the well-known utility maximization framework to abstract the objective. The objective function can be generalized to optimize system throughput, fairness (see the examples in Chapter 5), or tradeoffs among several system performance metrics [22] (e.g., a tradeoff between throughput and fairness, to be discussed in Chapter 7). Here, the objective function is chosen to maximize the system throughput.

Problem Formulation: Let M, N, and L denote the number of active links in a cluster, the number of subcarriers available in a cluster, and the number of timeslots (i.e., DATA slots) in a frame, respectively. Consider the following intra-cluster resource allocation optimization problem (ICRAOP)

$$\max_{\mathbf{c},\mathbf{p}} \left\{ \sum_{m=1}^{M} U_m(R_m(\mathbf{c},\mathbf{p})) \right\} \tag{4.6}$$

$$\text{subject to } R_m(\mathbf{c},\mathbf{p}) \geq R_m^d, \forall_m \tag{4.7}$$

$$p_{m,n}^l \geq 0, \forall_{m,n,l} \tag{4.8}$$

$$\sum_{n=1}^{N} p_{m,n}^l = P_m^{\max}, \forall_{m,l} \tag{4.9}$$

$$\sum_{m=1}^{M} c_{m,n}^l = 1, \forall_{n,l} \tag{4.10}$$

$$c_{m,n}^l \in \{0,1\}, \forall_{m,n,l} \tag{4.11}$$

where $U_m(.)$ is the utility function of the m^{th} link. By reducing the well-known NP-complete *number partitioning problem* [34] to the ICRAOP, it can be proved that the ICRAOP is NP-hard.

Proposition 4 *The ICRAOP is an NP-hard problem.*

4.4 Proposed Resource Allocation Approaches

To solve the ICRAOP, we propose three resource allocation approaches: 1) KKT-driven resource allocation; 2) GA-based resource allocation; and 3) hybrid resource allocation.

4.4.1 KKT-Driven Resource Allocation

In general, solving the ICRAOP requires exponential time complexity [9]. To make the problem more tractable, the most commonly used technique is to relax the integer constraints (4.11) into

$$c_{m,n}^{l} \geq 0, \forall m, n, l. \tag{4.12}$$

To investigate the (necessary) conditions for optimality of the solution, consider the following KKT conditions of the relaxed problem.

$$-\left(U_m'(R_m(\mathbf{c},\mathbf{p}))+\beta_m\right)\log_2\left(1+g_{mm,n}^{l}p_{m,n}^{l}\right)-\alpha_{m,n}^{l}+\lambda_n^{l}=0,\ \forall m,n,l \tag{4.13}$$

$$-\left(U_m'(R_m(\mathbf{c},\mathbf{p}))+\beta_m\right)\left(c_{m,n}^{l}\frac{g_{mm,n}^{l}}{1+g_{mm,n}^{l}p_{m,n}^{l}}\right)-\gamma_{m,n}^{l}+\mu_m^{l}=0,\ \forall m,n,l \tag{4.14}$$

$$R_m(\mathbf{c},\mathbf{p}) \geq R_m^{d}, \forall m \tag{4.15}$$

$$\sum_{m=1}^{M} c_{m,n}^{l} = 1, \forall n, l \tag{4.16}$$

$$\sum_{n=1}^{N} p_{m,n}^{l} = P_m^{\max}, \forall m, l \tag{4.17}$$

$$\beta_m\left(R_m^{d}-R_m(\mathbf{c},\mathbf{p})\right)=0, \forall m \tag{4.18}$$

$$\gamma_{m,n}^{l}p_{m,n}^{l}=0, \forall m, n, l \tag{4.19}$$

$$\alpha_{m,n}^{l}c_{m,n}^{l}=0, \forall m, n, l \tag{4.20}$$

$$c_{m,n}^{l}, p_{m,n}^{l}, \alpha_{m,n}^{l}, \gamma_{m,n}^{l} \geq 0, \forall m, n, l \tag{4.21}$$

$$\beta_m \geq 0, \forall m \tag{4.22}$$

where β_m, $\gamma^l_{m,n}$, μ^l_m, λ^l_n, and $\alpha^l_{m,n}$ are the Lagrange multipliers for the constraints (4.7), (4.8) (4.9), (4.10), and (4.12), respectively. Combining (4.13) and (4.20), the criterion of subcarrier allocation can be deduced as follows. For each timeslot l and each subcarrier n, choose m^* such that

$$m^* = \arg\max_m \left\{ U'_m(R_m(\mathbf{c},\mathbf{p})) \log_2 \left(1 + g^l_{mm,n} p^l_{m,n}\right) \right\} \tag{4.23}$$

and set $c^l_{m^*,n} = 1$. For a given subcarrier assignment, the optimal power allocation can be obtained via *utility-based water-filling* [100]. However, the criteria of optimal subcarrier allocation and optimal power allocation are coupled. To reduce the computational complexity, our resource allocation first fixes the power allocation, invokes subcarrier allocation based on the criterion given in (4.23), reallocates the subcarriers until all the system constraints are met, and then performs utility-based water-filling for power allocation. The detailed procedure is given below.

Step 1: For each link m, equally distribute the transmit power over all N subcarriers on each timeslot l (i.e., $p^l_{m,n} = P^{\max}_m/N$, $\forall l$).

Step 2: For each timeslot l and each subcarrier n, allocate the n^{th} subcarrier according to (4.23). Repeat until all subcarriers are allocated over all timeslots in the next frame.

Step 3: Since the initially obtained resource allocation solution may not satisfy all the constraints in (4.7), subcarrier reallocation is needed. Subcarrier reallocation should result in only a small decrease in the total utility while satisfying all the QoS constraints. Let M_u be the set of unsatisfied links. Consider the following cost function of reallocating a subcarrier to the j^{th} link instead of the originally assigned i^{th} link

$$\epsilon_{j,n} = \frac{1}{\Delta U_i(R_i(\mathbf{c},\mathbf{p})) - \Delta U_j(R_j(\mathbf{c},\mathbf{p}))} \frac{r^l_{j,n}}{R^d_j - R_j(\mathbf{c},\mathbf{p})}, \forall j \in M_u. \tag{4.24}$$

The cost $\epsilon_{j,n}$, a product of two terms, represents the *likelihood* of reallocating the n^{th} subcarrier to the j^{th} link instead of the originally assigned i^{th} link. The first term

measures the decrease in the total utility for subcarrier swapping. The smaller the change in the total utility, the larger the value of the first term so as to maintain the objective function as large as possible. The second term indicates the tendency of satisfying the QoS requirements of the j^{th} link if the n^{th} subcarrier is reallocated to it. The larger the second term, the more likely the QoS requirement of the j^{th} link can be met. The cost function is essentially the combined effects of both terms. Therefore, choose the subcarrier with the maximum cost, i.e., $j^* = \arg\max_{j \in M_u} \epsilon_{j,n}$. If the QoS constraint (4.7) for the i^{th} link is not violated, perform subcarrier swapping by setting $c^l_{j^*,n} = 1$ and $c^l_{i,n} = 0$. Repeat this step until all the subcarriers and the timeslots have been searched. The flowchart of the subcarrier swapping is depicted in Figure 4.1.

Step 4: For each timeslot l, perform utility-based water-filling for power allocation on each link.

4.4.2 GA-Based Resource Allocation

GA-based methodology is commonly used as a search algorithm based on mechanics of natural selection and natural genetics, aiming at near-optimal solutions [79]. A GA-based resource allocation scheme starts with an initial population of individuals characterized by *chromosomes*, and then improves the population through evolution. In each generation, three operations are carried out one by one to yield a new population: 1) selection; 2) crossover; 3) and mutation. In the selection process, a proportion of the existing population is selected to breed a new generation. A selection operator determines better individuals of the current generation stochastically. Those individuals who are not selected will vanish in the next generation. A crossover operator is to generate new individuals from the ones chosen by the selection operator. The crossover operation between two individuals is performed by swapping some of their genes to produce two new individuals. A mutation operator probabilistically changes some arbitrary gene(s) of a chromosome to a new value to form

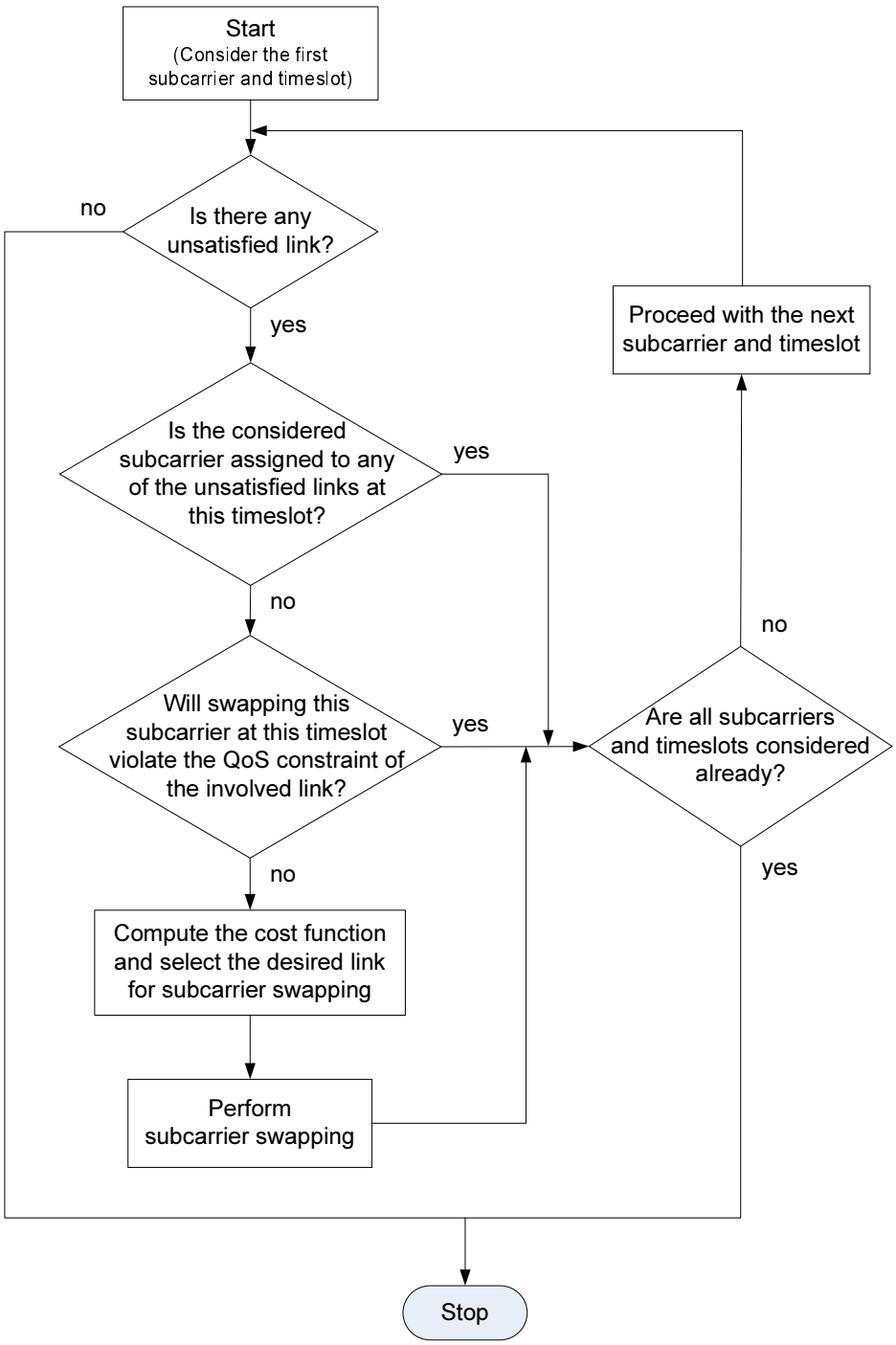


Figure 4.1: Subcarrier swapping.

a new individual. The mutation operation helps increase the diversity of the population by creating new individuals which may be quite different from the existing population. Compared to a (sub)gradient method [9], GAs can give a better solution by avoiding the local optimum(s) via crossover and/or mutation. One pitfall of GAs, however, is its relatively high computational cost.

A chromosome length of MNL is chosen for our GA-based resource allocation scheme. Each individual represents a joint power-frequency-time resource allocation solution. Each chromosome is a string of elements with binary values representing a subcarrier allocation solution over the time and the frequency domains. To form an initial population, all subcarriers in a frame are allocated randomly to the links in such a way that every individual is a feasible solution to the ICRAOP. A genetic representation of a set of feasible solutions to the ICRAOP is depicted in Figure 4.2. The fitness function is chosen to be the objective function given in (4.6). The fitness function of the i^{th} individual is given by $FF_i = \sum_{m=1}^{M} U_m(R_m(\mathbf{c}, \mathbf{p})), \forall i$, where $\mathbf{c}$ can be obtained by decoding the chromosome of the i^{th} individual. Optimal power allocation is employed in the fitness evaluation of an individual. Thus, the larger the fitness value, the fitter the individual, and the better the solution to the ICRAOP. For the selection operation, we consider a well-known roulette wheel selection operator [79], where an individual is selected with a probability proportional to its fitness value. A pictorial illustration of the roulette wheel selection process is shown in Figure 4.3. Due to its random nature, the roulette wheel selection operator allows weaker individuals to survive and potentially become stronger chromosomes (i.e., better solutions) during the course of evolution (i.e., crossover and mutation). Note that the population size is kept constant in every generation so an individual in the existing population may be selected more than once. For the sake of genetic stability, genes for crossover are chosen according to a uniform distribution, whereas only one gene in a chromosome is randomly chosen for mutation. Any individual who violates any of the constraints in the ICRAOP will be eliminated in the next generation. The

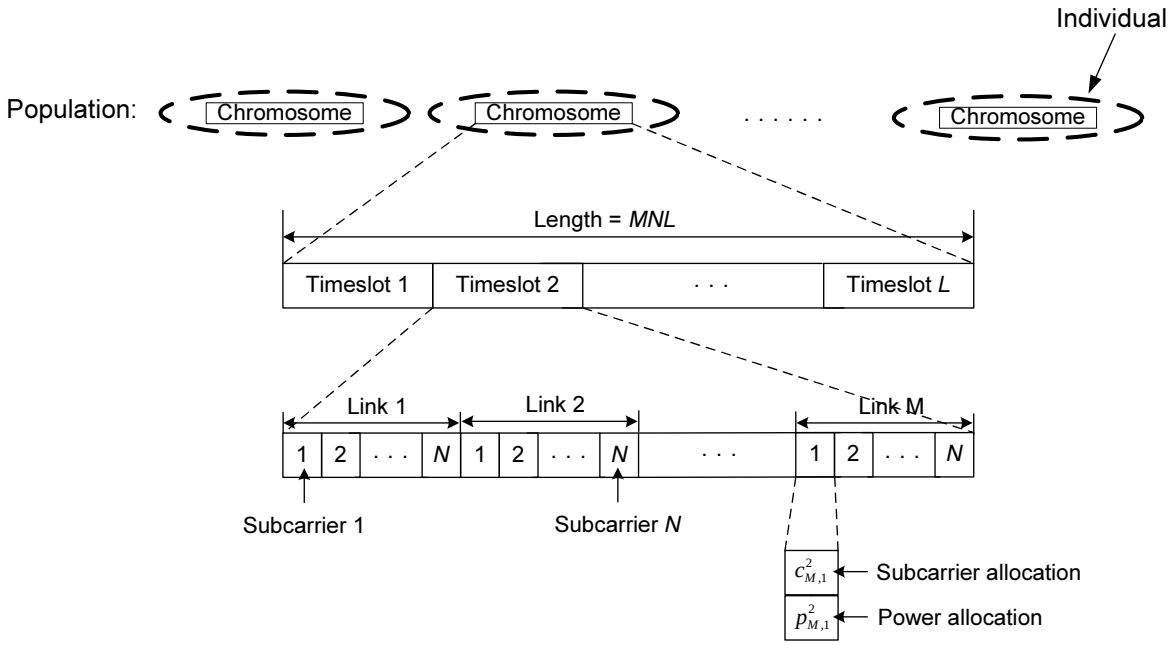


Figure 4.2: A genetic representation of a set of feasible solutions to the optimization problem.

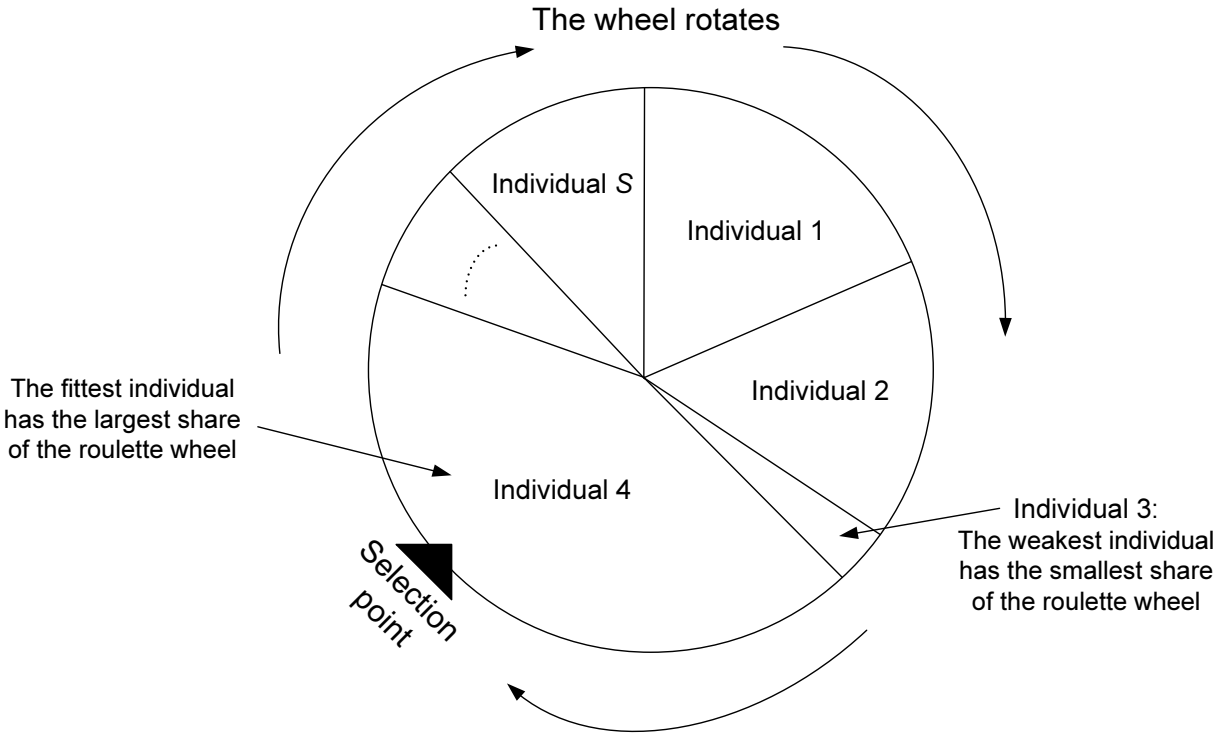


Figure 4.3: A pictorial illustration of the roulette wheel selection process. (All the individuals form a pie chart, where the size of a slice (i.e., the area where an individual occupies) is proportional to the fitness value of an individual. In this example, the 4^{th} individual is selected. The roulette wheel selection operator keeps selecting the individuals by literally spinning around the circle until the population size of the next generation is S.)

flowchart of the proposed GA-based resource allocation scheme is shown in Figure 4.4. The recommended crossover and mutation probabilities are 0.7 and 0.01, respectively [79]. Based on our observations, the preferred value for the maximum number of iterations, T, and that for the population size, S, are $25,000$ and 100, respectively.

4.4.3 Hybrid Resource Allocation: Combined-KKT-GA

Since the GA-based approach considers all three resource dimensions simultaneously, it can plausibly achieve better system performance than the KKT-driven approach. However, in practice, running the GA-based resource allocation algorithm with $S = 100$ and $T = 25,000$ on a low-cost off-the-shelf mesh router is not preferred due to its high computational complexity. To strive for a desired balance between system performance and complexity, we propose a hybrid scheme named *Combined-KKT-GA* by combining the merits of two aforementioned approaches. By studying the performance comparison of the proposed KKT-driven scheme and the proposed GA-based scheme with $S = 100$, we observe two main trends: 1) If the GA-based scheme performs better than the KKT-driven one after a few iterations (e.g., 10 iterations), then the former one will give a better resource allocation solution eventually; 2) If the GA-based scheme cannot prevail after a small number of iterations, it will be very likely to take several thousands of iterations to win out. Since the KKT-driven resource allocation solution will likely be locally optimal only, the rationale for the above observations is mainly due to the quality of the initial population generated in the GA-based algorithm (i.e., the initial set of the feasible solutions) and the structure of the feasible region of our optimization problem (i.e., the solution space governed by the system constraints). If the initial population is residing at a favorable *neighborhood* of the solution space where the path to reaching the global optimum is "hassle-free", the GA-based scheme can achieve a better solution in a few iterations with ease. In contrast, if the initial population is residing at a disadvantageous neighborhood of the solution space and/or the solution space contains many various local

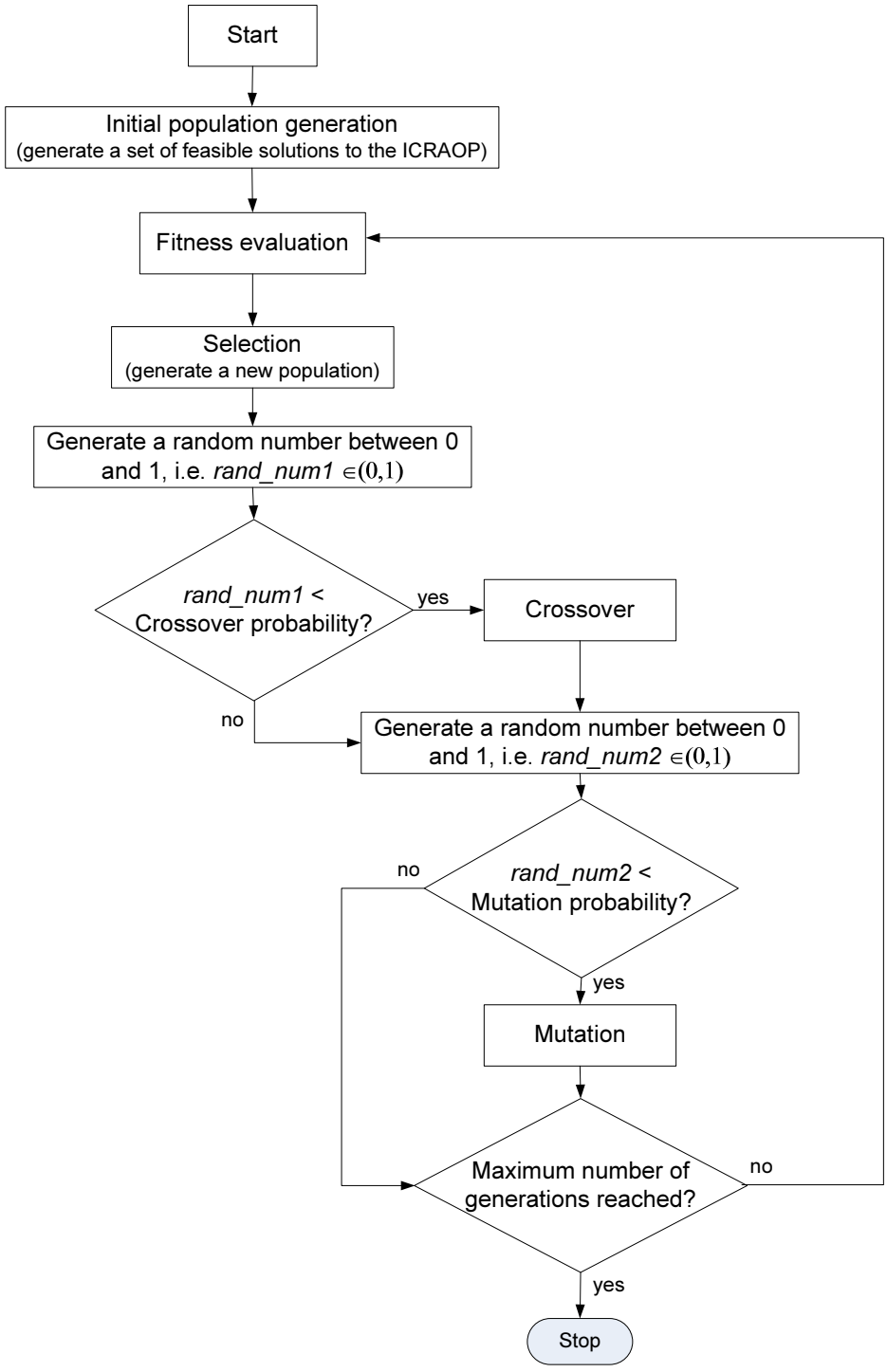


Figure 4.4: The flowchart of the proposed GA-based resource allocation scheme.

optimums in between the global optimum and the starting point of search, the GA-based scheme is expected to undergo lots of iterations before reaching the globally optimal point. Since efficiency is one of the most important goals in resource allocation for future WMNs, a hybrid resource allocation approach is proposed and outlined as follows.

Step 1: Obtain a resource allocation solution using the KKT-driven approach.

Step 2: Obtain a resource allocation solution using the GA-based approach with $S = 100$ and $T = 10$.

Step 3: Choose the better solution out of the above two solutions.

Therefore, on average, the performance of this hybrid approach, Combined-KKT-GA, is expected to be lower bounded and upper bounded by that of our KKT-driven scheme and that of our GA-based scheme, respectively. Note that the proposed hybrid approach resembles the notion of *SoftMAC*[1] [84]. In our case, the solution of the proposed Combined-KKT-GA approach is highly contingent on the quality of the initially generated population in the GA-based algorithm and the solution space of the optimization problem.

4.4.4 Complexity Analysis

Regarding the practicality of our proposed resource allocation algorithms, we study their time complexities. The complexities of the proposed KKT-driven approach, the proposed GA-based approach, and the Combined-KKT-GA approach are on the order of $O(LMN)$, $O(TSLMN)$, and $O(CLMN)$, respectively, where $C \ll TS$. As such, our KKT-driven and Combined-KKT-GA approaches are of low complexity, giving rise to viable candidates for practical implementation in WMNs with low-cost off-the-shelf mesh nodes. In the presence

[1]SoftMAC selects the best MAC protocol from a protocol pool based on various system parameters such as channel conditions, packet dropping, and packet delay.

of powerful devices with strong computational power, our GA-based approach can also be employed in practice.

4.5 Performance Evaluation

4.5.1 Simulation Environment

We consider a cluster with a number of wireless nodes randomly located in a 1km x 1km coverage area. We assume that the routing is predetermined so that the transmission source and destination pair of an incoming packet is known in advance. Simulation parameters are chosen as follows: $\eta \sim N(0, 10^{-12}\text{W})$, $I_n \sim N(0, 10^{-10}\text{W})$, $\forall_n$, $P_m^{\max} = 10\text{mW}$, $\forall_m$, $U_m(R_m(\mathbf{c}, \mathbf{p})) = R_m(\mathbf{c}, \mathbf{p})$, $\forall_m$, $N = 100$, and $L = 4$. The maximum transmission rate of each subcarrier is set to be 200kb/s. We adopt the channel model suggested in [44] (i.e., hilly/moderate-to-heavy tree density). We consider three types of traffic, namely voice, video, and data described in Section 2.2.2. Voice traffic and video traffic are generated according to a two-state ON-OFF model. For voice traffic, in the ON state, a fixed-size packet arrives at a constant rate, whereas in the OFF state, no packet is generated. The duration of an ON period and that of an OFF period both follow an exponential distribution. For video traffic, different incoming packets in the ON state have different packet sizes, generating a variable-rate traffic [39]. The parameters of the voice and video traffic models are summarized in Table 4.2. Here, we consider two data traffic models: 1) background data traffic, where data packets are available anytime; and 2) bursty data traffic, where data packet arrivals follow a Poisson process with mean rate 50 packets/second, where the packet size follows a Weibull distribution (i.e., Weibull(2,2)). Data traffic is assigned the lowest priority and does not have any rate requirements. To better utilize network resources, both the data rate and the size of a packet are varied based on a wireless channel condition by setting the transmission durations of all packets to be the same (i.e., the timeslots) [61]. To

Table 4.2: Summary of simulation parameters.

	Parameter	Value
Voice traffic	mean ON period	1.00s
	mean OFF period	1.35s
	packet inter-arrival time	20ms
	constant data rate	32kb/s
	delay bound	20ms
	packet-loss-rate requirement	1%
	bit-error-rate (BER) requirement	10^{-3}
Video traffic	mean ON period	1.47s
	mean OFF period	1.92s
	packet (frame) inter-arrival time	5ms
	minimum data rate	256kb/s
	delay bound	75ms
	packet (frame)-loss-rate requirement	1%
	bit-error-rate (BER) requirement	10^{-5}

resemble mixed traffic types, we consider a voice source, a video source, and a data source residing at every node. There are 1 beacon slot, 1 control slot, and 4 DATA slot(s) in a frame. The duration of a timeslot is 5ms. The polling is done in every 100ms. Both the polling and the beacon packet transmissions are assumed to be error-free. An offline CAC is performed, and we noticed that the system can admit 13 voice calls and 12 video calls. For medium access, we employ the packet-level QoS provisioning strategy given in Section 2.3. We perform the simulations for 10,000 runs and average the results, where each simulation run sustains 10,000 frames. The performance measurements used in our simulations are 1) (system) throughput, 2) resource utilization, and 3) packet dropping rate.

4.5.2 Simulation Results

We carry out extensive computer simulations to compare our three proposed approaches with a random scheme (in [21]), the Shi's KKT-driven scheme (in [98]), and the Gao's scheme (in [33]). In the random scheme, each node selects a number of subcarriers to meet its current traffic demand. Subcarrier selection is random. Transmit power is uniformly distributed over the selected subcarriers. If a subcarrier is chosen by more than one node, we assume that a collision occurs, and the subcarrier is wasted and does not contribute to the actual packet transmission rate. Resource reservation is not taken into account in this random scheme. Since every node behaves independently, the complexity of this random scheme is on the order of $O(N)$. The Shi's KKT-driven scheme is similar to our proposed KKT-driven approach but with different subcarrier reallocation strategies. The former aims at maximizing the system throughput (without considering QoS support), whereas the latter focuses on QoS satisfaction of the admitted calls. The complexity of the Shi's KKT-driven scheme is on the order of $O(LM^2N)$. Although the Gao's scheme makes use of optimal power allocation, its subcarrier allocation is only suboptimal. The complexity of the Gao's scheme is the same as our proposed KKT-driven approach, which is on the order of $O(LMN)$.

Here, we study the impact of the number of links on the system performance for all six approaches. In our performance evaluation, we further consider two cases: 1) with background data traffic; and 2) with bursty data traffic. We also analytically derive the upper bound for the throughputs obtained by all the traffic for reference (see Appendix-A.2).

Case 1: with Background Data Traffic

Figure 4.5 shows the throughput performance for background data traffic versus the number of links. The standard deviations of the results are also plotted for reference. Clearly,

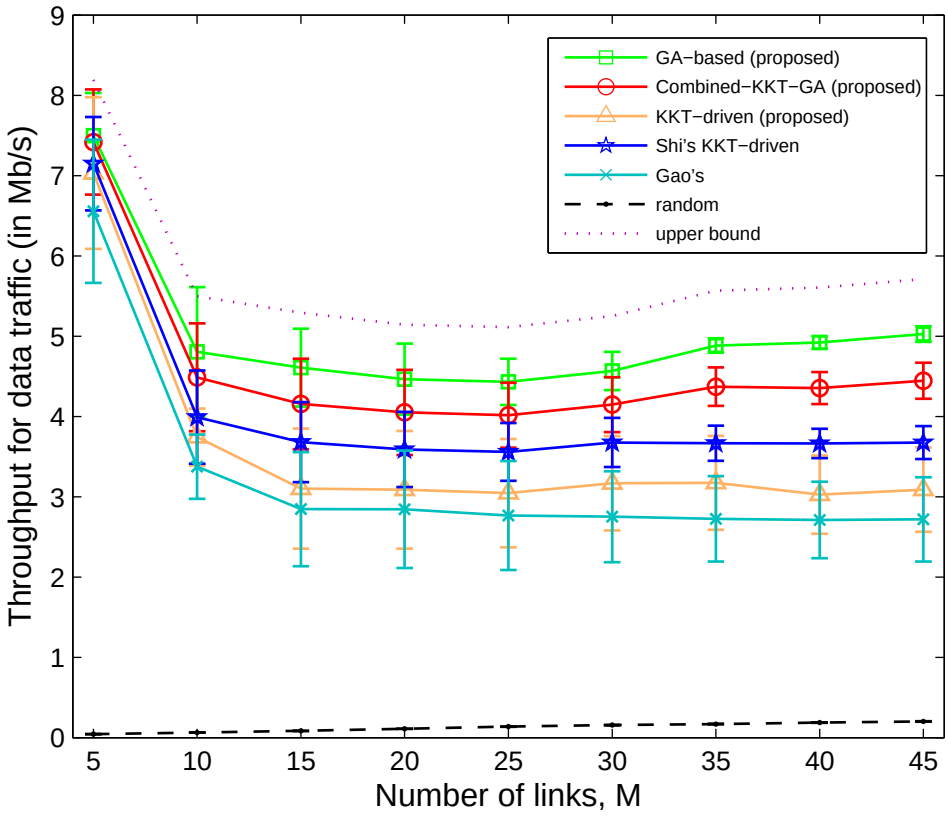


Figure 4.5: Comparison of the data throughput performance of the proposed approaches (GA-based, KKT-driven, and Combined-KKT-GA), the Shi's KKT-driven scheme, the Gao's scheme, and the random scheme vs. the number of links for background data traffic.

the proposed GA-based approach outperforms the other five approaches. This result is justifiable because our GA-based approach essentially takes all the three resource dimensions (i.e., power, subcarrier, and time) into consideration simultaneously. As expected, the throughput performance of Combined-KKT-GA is somewhat between that of the proposed KKT-driven approach and that of the proposed GA-based approach. We observe that the Shi's KKT-driven approach is better than our proposed KKT-driven approach because of its throughput-oriented subcarrier reallocation. Despite optimal power allocation in place, the Gao's scheme employs suboptimal subcarrier allocation and hence performs worse than our KKT-driven approach. Notice that the time complexities of the Gao's scheme and our proposed KKT-driven one are on the same order. On the other hand, the proposed Combined-KKT-GA approach achieves higher throughput than the Shi's KKT-driven, the Gao's, and the random approaches. Since more network resources are reserved for the real-time traffic with the number of links, the total throughputs of the background data traffic in our proposed approaches drop. Due to the CAC in place, no more than 13 voice sources and 12 video sources can be admitted into the system. Hence, from $M = 15$ onward, the total throughputs level off. The slight fluctuation of the curves is partly because of the random arrivals of voice and video packets. The upper bound for the throughput obtained by the data traffic is also plotted for reference. For resource utilization, we observe that the Combined-KKT-GA approach utilizes the resources better than the random scheme (which has less than 10% efficiency), the Gao's approach (about 45% from $M = 15$ onward), the proposed KKT-driven approach (about 50% from $M = 15$ onward), and the Shi's KKT-driven approach (about 55% from $M = 20$ onward). In fact, the utilization achieved by the proposed Combined-KKT-GA approach is close to that achieved by the GA-based approach. For the Combined-KKT-GA algorithm, the resource utilization drops from about 95% to about 65% when the number of links increases from $M = 5$ to 20, and the utilization stays around 60% from $M = 20$ onward. On the other hand, the resource utilization achieved by the GA-based

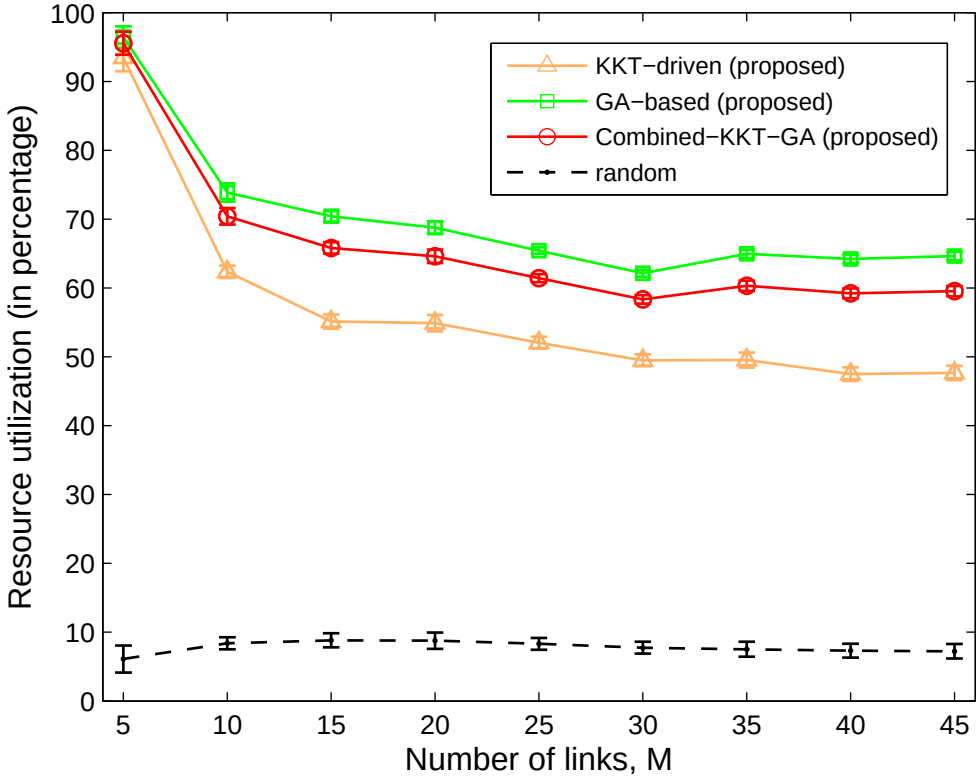


Figure 4.6: Comparison of the resource utilization of the three proposed approaches and the random scheme vs. the number of links for background data traffic.

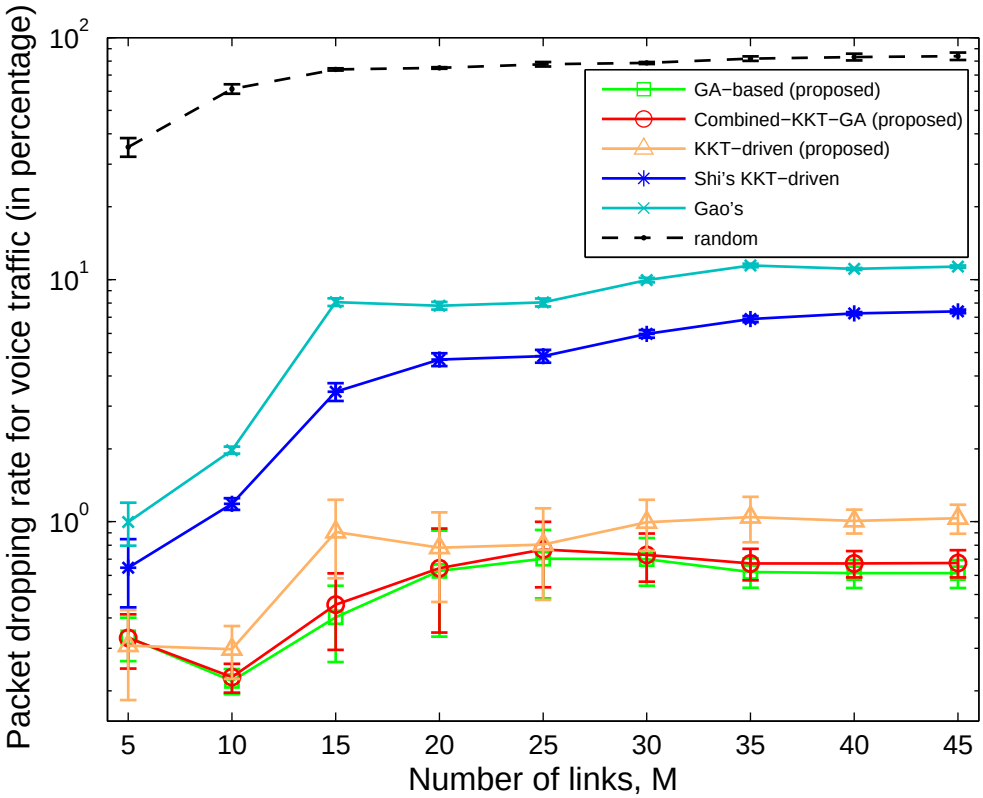


Figure 4.7: Comparison of the voice packet dropping rates of the proposed approaches (GA-based, KKT-driven, and Combined-KKT-GA), the Shi's KKT-driven scheme, the Gao's scheme, and the random scheme vs. the number of links for background data traffic.

approach drops from 97% to 71% when the number of links increases from $M = 5$ to 15, and then roughly stabilizes at about 65% from $M = 15$ onward. The resource utilization curves are depicted in Figure 4.6. For presentation clarity, only the curves for the proposed KKT-driven approach, the proposed GA-based approach, the proposed hybrid approach, and the random approach are shown. Notice that the reduction in the utilization is mainly ascribed to the bandwidth reservation for the real-time traffic. The simulation results confirm the fact that QoS provisioning and resource utilization maximization are conflicting with each other [22]. In Figure 4.7, the voice packet dropping rates are depicted. The packet dropping

rates for voice traffic are capped in our Combined-KKT-GA (about 0.7%), our KKT-driven (about 1%), and our GA-based (about 0.6%) approaches. From $M = 15$ onward, the packet dropping rates are about 6% and 11% in the Shi's KKT-driven approach and the Gao's approach, respectively. Suboptimal subcarrier allocation in the Gao's approach leads to the incapability of provisioning QoS effectively. On the other hand, in the Shi's KKT-driven approach, the subcarrier reallocation step aims at the system-wise throughput improvement at the cost of link-wise (or call-wise) QoS satisfaction, resulting in poor QoS provisioning performance. In the random scheme, the voice packet dropping rates increase as M increases. After all, the Shi's KKT-driven, the Gao's, and the random approaches are not designed for effective packet-level QoS provisioning and, therefore, they cannot provide the same level of QoS assurance offered by our proposed approaches. We observe a similar trend for the packet dropping rates for video traffic.

Case 2: with Bursty Data Traffic

Here, we study the case where data packet arrivals follow a Poisson process. Figure 4.8 shows the system throughput obtained in serving three traffic types versus the number of links. The curve of data capacity defined by the maximum achievable system throughput without QoS constraints is also plotted for reference. The trend of the system throughput performance in all approaches except the random one increases with the number of links. As anticipated, our GA-based scheme performs the best among all schemes. The performance of the proposed Combined-KKT-GA approach is almost the average of that of the proposed KKT-driven approach and that of the proposed GA-based approach. Due to the optimal power and subcarrier allocation, we observe that our KKT-driven approach performs better than the Gao's approach. Notice that our KKT-driven approach attains almost the same system throughput as the Shi's KKT-driven one. This phenomenon stems from the fact that the data traffic load is low and hence the effect of the throughput-sensitive subcarrier

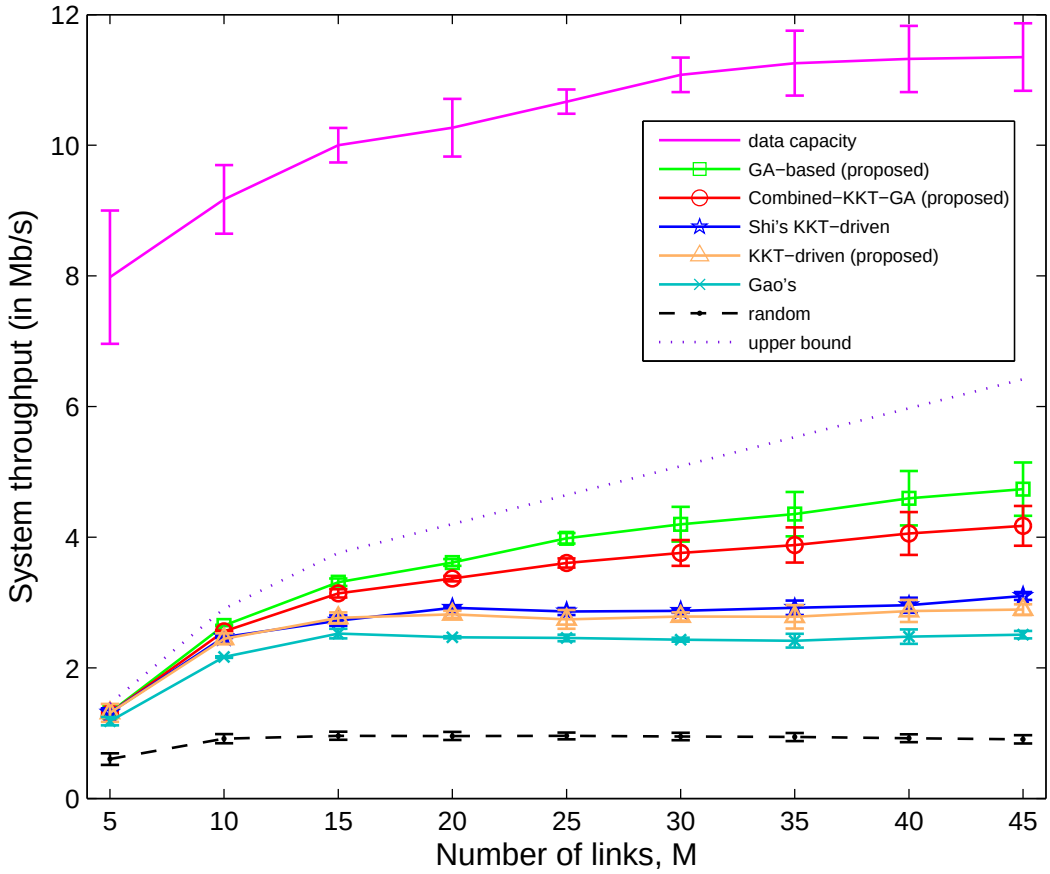


Figure 4.8: Comparison of the system throughput of the proposed approaches (GA-based, KKT-driven, and Combined-KKT-GA), the Shi's KKT-driven scheme, the Gao's scheme, the random scheme, and the data capacity vs. the number of links for bursty data traffic.

reallocation step in the Shi's approach is not substantial. In fact, there are two key factors determining the trend of the curves, namely traffic load and resource reservation for the real-time traffic. Since the network is not saturated, the higher the traffic load, the higher the throughput. The curves go up when M increases from 5 to 15, as there is a higher traffic load. From $M = 15$ onward, the influence of the resource reservation for QoS provisioning becomes more significant. Thus, the increment of system throughputs is smaller. On the other hand, the data capacity increases with the number of links because of the multi-user diversity. It is noteworthy that, in Figure 4.8, there is an obvious performance gap between the system throughputs obtained from all the approaches and the data capacity. This gap is ascribed to the low traffic load and resource reservation. We observe that both the system throughput and resource utilization for the proposed algorithms can be improved when the traffic load increases. The upper bound for the system throughput obtained by analysis is also plotted for comparison. In Figure 4.9, the resource utilization curves for our KKT-driven approach, our GA-based approach, our Combined-KKT-GA approach, and the random approach are depicted. For the dropping rates of the real-time traffic, we observe that the results are similar to Figure 4.7. Thus, it can be concluded that the data traffic model has little influence on the performance of the real-time traffic, for the data packets are assigned the lowest priority.

Simulation results show that our three proposed approaches can effectively provision packet-level QoS for real-time traffic and achieve satisfactory system throughput performance under both bursty and background data traffic models. In particular, the novel hybrid resource allocation approach, Combined-KKT-GA, is of low computational complexity, leading to a viable and preferred candidate for practical implementation. The characteristics of the aforementioned resource allocation approaches are summarized in Table 4.3.

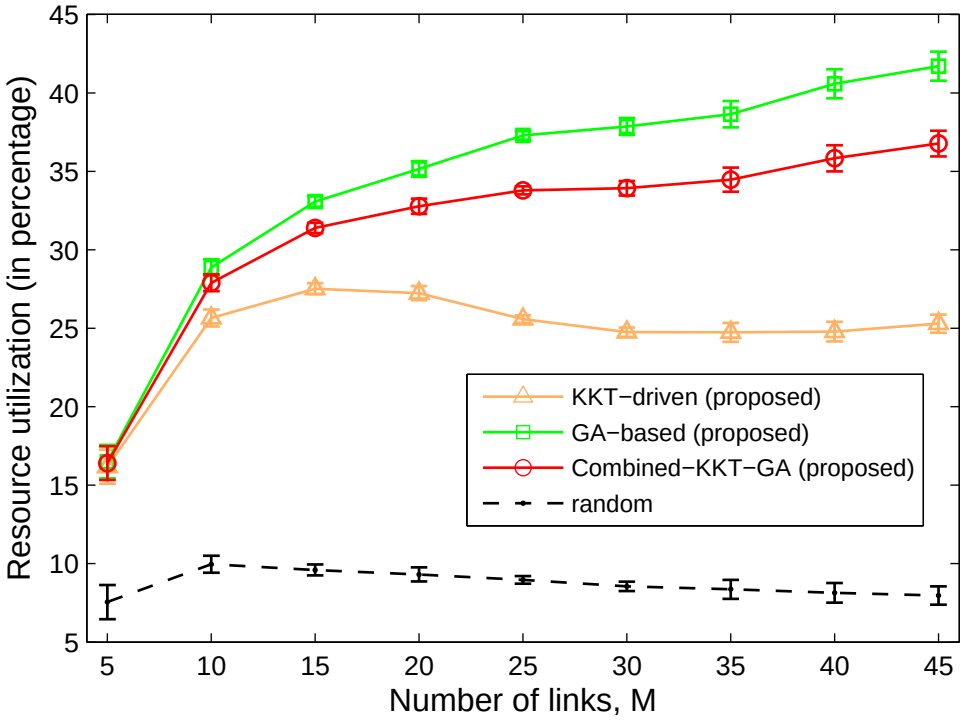


Figure 4.9: Comparison of the resource utilization of the three proposed approaches and the random scheme vs. the number of links for bursty data traffic.

Table 4.3: The characteristics of different resource allocation approaches.

Resource Allocation Approach	Complexity	Throughput Performance	QoS Provisioning	Practicality
Exhaustive Search	$O(M^{NL})$	Optimal	Yes	No
GA-based (proposed)	$O(TSLMN)$	Close to optimal*	Yes	Yes*
Combined-KKT-GA (proposed)	$O(CLMN)$**	Good	Yes	Yes
KKT-driven (proposed)	$O(LMN)$	Average	Yes	Yes
Shi's KKT-driven [98]	$O(LM^2N)$	Good	No	Yes
Gao's [33]	$O(LMN)$	Average	No	Yes
Random [21]	$O(N)$	Poor	No	Yes

Note *: The results are contingent on the number of iterations used in the GA-based resource allocation scheme.

Note **: The value of C is far less than that of TS.

4.6 Discussion

Our simulation results given in Section 4.5 show that resource utilization is poor when the traffic load is low (see Figure 4.9). However, recent studies show that the quantity of background data traffic has been skyrocketing, constituting a substantial portion of the total traffic in WMNs [3]. In future broadband WMNs with abundant background data traffic, resources are expected to be utilized more efficiently, whereby system throughput can be improved. On the other hand, with limited network resources, throughput maximization and QoS support are two conflicting goals, leading to a natural tradeoff. Thus, admitting more multimedia calls can, in fact, bring down the system throughput (see Figure 4.5). In WMNs supporting multimedia applications, radio resource allocation approaches should be strategically devised to effectively guarantee the QoS of heterogeneous traffic yet meliorate the system throughput. Thus, finding a desired performance balance between QoS support and throughput improvement is undoubtedly imperative. Similarly, acquiring a desired balance between system-wise performance optimization and user-wise performance optimization is also essential, as increasing system-wise performance would deteriorate user-wise perfor-

mance, or vice versa.

4.7 Chapter Summary

In this Chapter, three QoS-aware packet-level resource allocation approaches, namely KKT-driven, GA-based, and Combined-KKT-GA, are proposed for the joint power-subcarrier-time intra-cluster resource allocation problem in WMNs. Simulation results demonstrate that our GA-based approach performs the best in terms of throughput and packet dropping rate at the expense of high computational complexity. On the other hand, the proposed KKT-driven approach triumphs over its counterpart of the same time complexity. The proposed Combined-KKT-GA approach works out a desired performance tradeoff between computational complexity and system performance, giving rise to a preferred candidate for practical implementation. Besides, our results confirm that QoS provisioning and resource utilization are two conflicting performance metrics. Further, bounds for the throughputs obtained by realtime and non-real-time traffic are derived analytically, serving as viable performance benchmarks.

Chapter 5

Performance Enhancement I: Non-Altruistic Node Cooperation

5.1 Introduction

In WMNs, increasing throughput is the key to the success of providing an all-wireless environment. In previous Chapters, we have proposed novel and efficient resource allocation approaches to provide a high-speed mesh backbone with QoS assurance. To further enhance the system performance, the notion of node cooperation can be employed. With the unique characteristics of WMNs, however, a new node cooperative resource allocation strategy tailored for QoS-sensitive WMNs with heterogeneous traffic is imperative. Here, in lieu of designing a node cooperation protocol for WMNs, we assume that there is an efficient cooperation protocol in place, and we study the impact of node cooperation and resource allocation on the system performance. In specific, we focus on MAC-layer node cooperative resource allocation in WMNs with QoS assurance and service differentiation. We consider regenerative mesh nodes and non-altruistic node cooperation, meaning that there is no pure relay in the system and all the nodes have their own data to transmit. Further, the node

cooperation of interest is non-reciprocal or asymmetric, meaning that one node assists its neighbor but does not necessarily receive help from that neighbor. We also consider prioritized node cooperation so as to realize the notion of service differentiation. The contributions and significance of this research work are three-fold [20].

- First, we study the problem of non-altruistic node cooperative resource allocation for WMNs with QoS support, taking subcarrier allocation, power allocation, partner selection/allocation, service differentiation, and packet scheduling into account. Thanks to the KKT interpretations, we propose two resource allocation approaches tailored for the WMN of interest, namely a four-phase centralized approach and a two-phase distributed approach. Both approaches are shown to be effective in providing QoS assurance and service differentiation.

- Second, our proposed resource allocation solutions are Pareto optimal, making efficient use of scarce network resources. Modeled by a round-robin game, the proposed approaches attain Nash Equilibria, facilitating network stability. Further, our approaches are of low complexity, leading to viable candidates for practical implementation.

- Third, simulation results show that both proposed approaches are effective in packet-level QoS provisioning and system performance enhancement over their non-cooperative counterparts in terms of throughput, resource utilization, node cooperation gain, and packet dropping rate. Our results demonstrate that the proposed non-altruistic node cooperative approaches are less vulnerable to the changes in the system parameters such as the accuracy of traffic load estimates. Further, our study reveals a critical principle that whether node cooperation is beneficial to WMNs depends upon the nature of node cooperation, (i.e., non-altruistic or altruistic), the mode of network operation, (i.e., centralized control or distributed control), and the traffic pattern (i.e., bursty traffic or background traffic).

Table 5.1: Summary of important symbols used in this Chapter.

Symbol	Definition
$U_m(R_m(\cdot)\,\vert\boldsymbol{\Theta})$	utility function the m^{th} node given $\boldsymbol{\Theta}$
$R_m(\cdot)$	achievable data rate of the m^{th} node
R_m^d	QoS demand of the m^{th} node
$\boldsymbol{\Theta}$	utility selector of the WMN of interest
$c_{m,n}^l$	indicator of allocating the n^{th} subcarrier to the m^{th} node on the l^{th} timeslot, i.e., $c_{m,n}^l \in \{0,1\}$
$p_{m,n}^l$	transmit power over the n^{th} subcarrier to the m^{th} node on the l^{th} timeslot
z_{mu}	indicator of node cooperation offered to the m^{th} node by the u^{th} node, i.e., $z_{mu} \in \{0,1\}$
a_{mu}	power scaling factor of the transmit power of the u^{th} node allocated for the m^{th} node's transmissions
$G_{mu,n}^l$	channel gain from the u^{th} node to the receiver of the m^{th} node's transmissions over the n^{th} subcarrier on the l^{th} timeslot
$g_{mu,n}^l$	effective channel gain, i.e., $\varphi G_{mu,n}^l/\sigma_n^2$
σ_n^2	aggregate noise-plus-co-channel interference power on the n^{th} subcarrier
φ	BER measure
$P_m^{\max}$	maximum power constraint of the m^{th} node
M	number of mesh nodes
$\mathcal{M}_1$	set of RG nodes
$\mathcal{M}_2$	set of BE nodes
N	number of subcarriers available
L	number of timeslots (i.e., DATA slots) in a frame
$\mathcal{N}_m^l$	set of subcarriers allocated to the m^{th} node on the l^{th} timeslot

The remainder of this Chapter is organized as follows. Related work is given in Section 5.2. Section 5.3 introduces the idea of non-altruistic node cooperation with regenerative nodes. The problem of non-altruistic node cooperative resource allocation for WMNs with QoS support is described in Section 5.4. Two proposed QoS-driven resource allocation approaches with node cooperation are presented in Section 5.5. Efficiency evaluation by game theory is given in Section 5.6. A brief discussion is provided in Section 5.7. Performance evaluation is given in Section 5.8. Finally, a summary is drawn in Section 5.9. A summary of important symbols used in this Chapter is given in Table 5.1 for easy reference.

5.2 Related Work

In the literature, there exists a rich body of research work on node cooperation [8,12,14–16,28, 40,41,43,48,56,63,64,72,81,87,94,109,110,120–122,124]. Besides information-theoretic stud-

ies [28,40,63], most recent work on the topic of node cooperation aims at 1) distributed space-time coding design and 2) resource allocation with relay selection. The former group focuses on the design and performance evaluation of distributed space-time coding (e.g., [16,48,64]). Spectral efficiency can be further improved by means of dirty paper coding applied at the transmitters [56]. With appropriate detection techniques, an additional diversity order can be attained in time-varying wireless channels [15]. However, most node cooperation strategies based on distributed space-time coding conceive the existence of pure relay nodes, which is not always feasible in practical WMNs. In addition, most of the above research work focuses on the performance measures in the physical layer, while resource allocation with respect to the MAC layer is not addressed properly. The latter group aims at relay selection and resource allocation. Partner matching algorithms based on graph theory are proposed in [72,110], the objective of which is to optimize the energy efficiency; however, energy consumption is not a concern in WMNs. In fact, relay selection and relay allocation are expected to be relatively straightforward in altruistic node cooperative WMNs, for mesh routers are stationary and channel state information can be known. To achieve fairness in an OFDM-based system, a max-min cooperative subcarrier allocation approach is proposed in [94]. In order for each node to have a fair share of channel capacity, a relay node selection scheme with constant transmit power is suggested in [99]. On a different note, some work employs the notion of game theory to derive node cooperation strategies and cooperative resource allocation solutions (e.g., Nash bargaining approach [14,109,122]), where game-theoretic fairness is achieved. Auction-based cooperative resource allocation algorithms are also proposed to increase throughput and reduce outage probability [12,41,43,81]. The performance results of the aforementioned approaches can be appealing; however, powerful central controllers are mostly required to execute those preceding algorithms. In the case of austere suburban and rural environments, distributed control is preferred, yet directly applying those approaches to WMNs with decentralized control can be ineffective or inefficient.

In the presence of multiple relays, it has been shown that selecting the (single) best relay is sufficient to achieve the full diversity order and/or diversity-multiplexing tradeoff of a cooperative system [8,124]. The implication is that the design of space-time coding can be considerably simplified and the message overhead can be greatly reduced. However, most of the work in the literature focuses on altruistic node cooperation (i.e., pure relays exist) (e.g., [72,124]). Besides, many of the existing partner selection schemes consider only measured (instantaneous) SNRs in choosing a partner (e.g., [8]). In the context of WMNs for suburban/rural residential areas, all the mesh nodes mounted on the rooftops are expected to have their own data to transmit. Non-altruistic node cooperation is, therefore, more reasonable and realistic, where there is no pure relay in the system and all the nodes have their own data to transmit. Due to the necessity of the split of transmit power (i.e., part of the transmit power of a node is dedicated to transmitting its own data, while the rest is dedicated to assisting other nodes), the measured SNRs should not be the sole decision metric for partner selection and/or partner allocation. In fact, the availability of a potential partner and its QoS requirements should also be taken into account. In the case of non-altruistic node cooperation, the problem of node cooperative resource allocation becomes convoluted, as the problems of partner selection/allocation, power allocation, subcarrier allocation, service differentiation, and QoS provisioning are coupled. In [87,120], the problem of partner selection for non-altruistic node cooperation is studied. Three partner selection schemes with power control are proposed in [120] for balancing transmit power and system performance, whereas two partner selection schemes are proposed in [87] so as to minimize the average outage probability. However, MAC-layer service differentiation is not addressed, for only a single class of traffic is considered in the aforementioned work. In [121], Zhang *et al.* propose a simple two-step scheme for the system throughput maximization problem with physical-layer QoS assurance. Without considering beneficial node cooperation, however, system performance can be undermined. In addition, both packet scheduling and packet-

level QoS support are not taken into consideration, plausibly elevating the packet dropping rates for real-time traffic.

With no powerful central controllers, in this work, we propose two low-complexity QoS-driven node cooperative resource allocation approaches tailored for WMNs with heterogeneous traffic. Our proposed approaches are demonstrated efficient yet effective in providing QoS assurance and facilitating service differentiation for WMNs supporting multimedia applications.

5.3 Non-Altruistic Node Cooperation with Regenerative Nodes

In this research work, two classes of traffic are considered, namely 1) rate-guaranteed (RG) traffic (e.g., voice traffic) and 2) best-effort (BE) traffic (e.g., data traffic). In particular, the RG traffic has a minimum data rate requirement and a delay bound requirement, whereas the BE traffic has no QoS requirement. As regards aggregate traffic flows, a mesh node is treated as an RG node if any of its incoming traffic flows belongs to the RG traffic.

Node cooperation is considered and triggered as long as it is feasible and beneficial for system performance. In the presence of multiple available partners, we can prove that, with arbitrary power allocation, choosing the (single) best partner is sufficient to achieve the full diversity order (see Appendix-A.3.4). As such, a wireless node simply selects the one with which the achievable throughput is maximized. Some node cooperation scenarios in the WMN of interest are depicted in Figure 5.1. In our system model, Examples (a)-(c) are plausible node cooperation scenarios where strong source-relay links exist. Example (d), however, is less likely to happen because suburban/rural residential houses are sparsely distributed, and the transmit power levels for the transmissions between mesh routers and

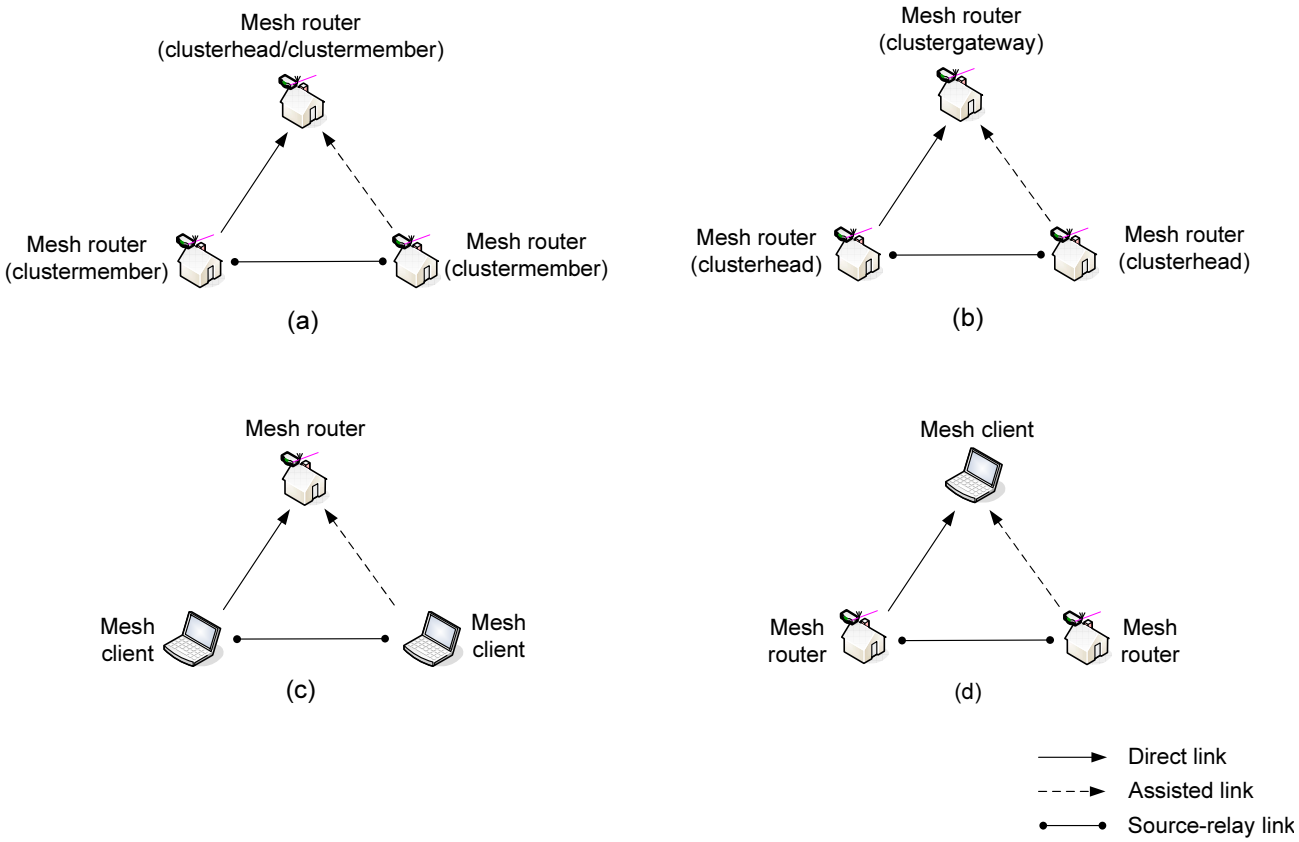


Figure 5.1: Examples of node cooperation in WMNs.

mesh clients are expected to be low. Therefore, the source-relay links in Example (d) are relatively weaker than the ones in Examples (a)-(c). Nonetheless, Example (d) can be justified when the mesh routers and mesh clients are closely placed (e.g., in the WMNs for metropolitan areas). It is noteworthy that Example (b) resembles downlink communications in a cellular network where a mobile terminal located at the borders of two neighboring cells is served by two corresponding base stations. Here, we focus on a node cooperation scenario where clustermembers cooperatively transmit their packets to a clusterhead or a clustermember in the same cluster (i.e., Figure 5.1(a)).

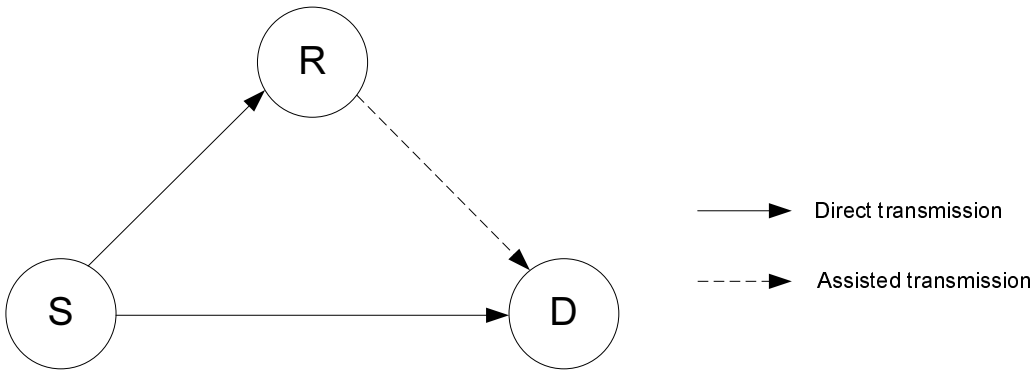


Figure 5.2: Schematic representation of relay-assisted transmissions.

Consider a transmission scenario involving three wireless nodes, namely Node S, Node R, and Node D. Node S is to transmit data to Node D, while Node R is viewed as a relay to help Node S forward the data to Node D. A schematic representation is illustrated in Figure 5.2. We employ the Cooperation Protocol I[1] suggested in [82] as our node cooperation strategy throughout this Chapter. In the first timeslot, Node S transmits a packet to both Node R and Node D. In the second timeslot, Node R forwards the packet received from Node S to Node D, while Node S transmits another packet to Node D. Notice that Node R can choose the decode-and-forward (DF) mode or amplify-and-forward (AF) mode of cooperation in the second timeslot. In this work, we consider the DF mode of cooperation (i.e., regenerative mesh nodes). The outage performances for the DF mode of cooperation (and the AF mode of cooperation) are presented in the Appendix-A.3. According to our analysis, if Node R can perfectly decode the symbols transmitted from Node S, the channel capacity C_{DF} achieved by the DF mode with perfect decoding is given by $C_{DF} = \frac{1}{2}\log_2\Big((1+\gamma_{SD})^2 + \gamma_{RD}\Big)$,

[1]Notice that, in this cooperation protocol, Node D receives two copies of the first packet and one copy of the second packet in two timeslots. In traditional cooperative networks with nodes employing full transmit power, the achievable throughput per timeslot due to cooperative transmissions can be higher than direct transmissions, thanks to both spatial multiplexing gain and power gain.

where $\gamma_{XY} = E_{XY}|h_{XY}|^2/\sigma^2$ with $E_{XY}(\geq 0)$ being the received average energy at Node Y from Node X, $X, Y \in \{S, R, D\}$, h_{XY} the Rayleigh fading coefficient for the $X \rightarrow Y$ link modeled as an independent zero-mean complex Gaussian random variable with unit variance, and σ^2 the noise-plus-cochannel interference power level. In the case of non-altruistic node cooperation, due to the necessity of splitting the transmit power, C_{DF} is given by $C_{DF} = \frac{1}{2}\log_2\Big((1 + a_S\gamma_{SD})^2 + a_R\gamma_{RD}\Big)$, where a_X is the scaling factor for the transmit power of Node X, i.e., $0 \leq a_X \leq 1$. As shown in Appendix-A.3.3, we can prove that arbitrarily positive power allocation has no impact on the diversity performance in the DF cooperation mode with perfect decoding. We can further prove that, with m potential relays, choosing the single best relay is sufficient to achieve the full diversity order in the DF cooperation mode with perfect decoding and arbitrarily positive power allocation. It is noteworthy that the case of $a_S = 1$ corresponds to the scenario of altruistic node cooperation. Compared to the channel capacity achieved by ordinary direct transmissions, denoted by $C_d = \log_2(1 + \gamma_{SD})$, we can envision that, in a 3-node model, a non-altruistic cooperative transmission may not always be advantageous over an ordinary direct transmission in terms of throughput. Therefore, for the sake of overall system performance, node cooperation among mesh nodes should be considered in a holistic manner.

Concerning packet-level node cooperation, after overhearing the packet transmissions from Node S, Node R can verify the MAC-layer error correction code(s) of the received packet (e.g., forward error correction (FEC) and cyclic redundancy check (CRC)) to determine whether that packet is corrupted or not. If a packet is received successfully, Node R can become one of Node S's potential partners, and node cooperative resource allocation can be carried out hereafter whenever feasible and favorable. On the other hand, unlike symbol-level cooperation where continuous symbol transmissions can be feasible, continuous packet transmissions are less realistic. Thus, in this work, we consider that a potential packet-level cooperation opportunity arises in every two timeslots.

5.4 QoS-Driven Non-Altruistic Node Cooperative Resource Allocation

There are various system constraints associated with the non-altruistic node cooperative resource allocation problem for WMNs. Let M, N, and L denote the number of mesh nodes, the number of subcarriers available, and the number of timeslots (i.e., DATA slots) in a frame, respectively. The sum of the (non-negative) transmit power of each mesh node on the (allocated) subcarriers is bounded by a maximum power level:

$$\sum_{n=1}^{N} p_{m,n}^{l} \leq P_{m}^{\max}, \forall m, l \tag{5.1}$$

$$p_{m,n}^{l} \geq 0, \forall m, n, l \tag{5.2}$$

where $p_{m,n}^{l}$ is the transmit power of the m^{th} node over the n^{th} subcarrier on the l^{th} timeslot and $P_{m}^{\max}$ is the maximum power constraint of the m^{th} node. With the help of node clustering, a number of subcarriers are allocated to each cluster. Here, we consider the case where each subcarrier can only be allocated to one transmission link without cooperation or at most two transmission links with node cooperation (i.e., a direct link and an assisted link) in a cluster. If node cooperation is employed, both a source node and its partner transmit data over the same subcarrier(s). In addition, as discussed in Section 5.3, since choosing the best partner is sufficient in terms of outage performance, a node can have at most one partner at a time on condition that the node cooperation is meritorious. The aforementioned constraints can

be formulated as follows:

$$\sum_{m=1}^{M} c_{m,n}^{l} \leq \begin{cases} 1, & \text{without cooperation} \\ 2, & \text{with cooperation} \end{cases}, \forall n, l \tag{5.3}$$

$$\sum_{m=1, m\neq u}^{M} z_{mu} \leq 1, \forall u \tag{5.4}$$

$$\sum_{u=1, u\neq m}^{M} z_{mu} \leq 1, \forall m \tag{5.5}$$

$$c_{u,n}^{l} = c_{m,n}^{l}, \text{ when } z_{mu} = 1, \forall m, u, n, l \tag{5.6}$$

$$c_{m,n}^{l} \in \{0, 1\}, \forall m, n, l \tag{5.7}$$

$$z_{mu} \in \{0, 1\}, \forall m, u \tag{5.8}$$

where $c_{m,n}^{l}$ is an indicator of allocating the n^{th} subcarrier to the m^{th} node on the l^{th} timeslot and z_{mu} is the indicator of node cooperation offered to the m^{th} node by the u^{th} node. Notice that, in general, $z_{yx} \neq z_{xy}, \forall x, y$, as the node cooperation considered in this work is non-reciprocal (i.e., asymmetric). For the sake of notational convenience, we set $z_{mm} = 1, \forall m$. Provided that node cooperation is in place, the transmit power of a node is split into two segments, one dedicated to its direct transmissions and the other to the assisted transmissions for its partner:

$$\sum_{m=1}^{M} z_{mu} a_{mu} = 1, \forall u \tag{5.9}$$

$$0 \leq a_{mu} \leq 1, \forall m, u \tag{5.10}$$

where a_{mu} is the normalized portion of the total transmit power of the u^{th} node for assisting the m^{th} node's transmissions. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be the set of RG nodes and that of BE nodes, respectively, i.e., $M = |\mathcal{M}_1| + |\mathcal{M}_2|$. In our problem formulation, we take the minimum rate requirements of the RG nodes in the current frame, if any, into account:

$$R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z}) \geq R_m^d, \forall m \tag{5.11}$$

where R_m^d is the (instantaneous) transmission rate demand of the m^{th} node in the current frame (i.e., $R_m^d > 0, \forall m \in \mathcal{M}_1$ and $R_m^d = 0, \forall m \in \mathcal{M}_2$) and $R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})$ is the achievable data rate of the m^{th} node over a frame, which can be computed by [64]

$$R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z}) = \sum_{n=1}^{N} \sum_{l=1}^{L} \frac{1}{2} c_{m,n}^l \log_2 \left(\left(1 + a_{mm} g_{mm,n}^l p_{m,n}^l\right)^2 + \sum_{u=1, u \neq m}^{M} z_{mu} a_{mu} g_{mu,n}^l p_{u,n}^l \right). \tag{5.12}$$

In (5.12), $\mathbf{c} = \left[c_{m,n}^l\right]_{M \times N \times L}$, $\mathbf{p} = \left[p_{m,n}^l\right]_{M \times N \times L}$, $\mathbf{a} = [a_{mu}]_{M \times M}$, $\mathbf{z} = [z_{mu}]_{M \times M}$, and $g_{mu,n}^l = \varphi G_{mu,n}^l / \sigma_n^2$, where φ is a BER measure, $G_{mu,n}^l$ the channel gain from the u^{th} node to the receiver of the m^{th} node's transmissions over the n^{th} subcarrier on the l^{th} timeslot, and σ_n^2 the aggregate noise-plus-cochannel interference power on the n^{th} subcarrier. Notice that constraint (5.6) is implicitly incorporated in (5.12). With the aforesaid system constraints, different objective functions can be considered. Here, we employ the well-known utility maximization framework to abstract our objective function. Let $U_m(R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z}) | \mathbf{\Theta})$ denote the utility function of the m^{th} node and $\mathbf{\Theta} \in \{1, 2\}$ the utility selector. The objective function is given by

$$\sum_{m=1}^{M} U_m(R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})) | \mathbf{\Theta}) = \begin{cases} \sum_{m=1}^{M} R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z}), & \text{when } \mathbf{\Theta} = 1 \\ \sum_{m=1}^{M} -\left[-\ln\left(\frac{R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})}{A}\right)\right]^{\kappa}, & \text{when } \mathbf{\Theta} = 2 \end{cases} \tag{5.13}$$

where A is a sufficiently large constant such that $0 < R_m/A < 1, \forall m$ [59]. Thus, the objective function is to maximize the system throughput when $\mathbf{\Theta} = 1$, to achieve proportional fairness when $\mathbf{\Theta} = 2$ and $\kappa = 1$, and to achieve max-min fairness when $\mathbf{\Theta} = 2$ and $\kappa \to \infty$, respectively. It can be shown that the utility function of a mesh node is an increasing function of its achievable data rate.

Lemma 1 *The utility function of the m^{th} node is an increasing function of its achievable data rate.*

In practice, the choice of $\mathbf{\Theta}$ is contingent on the purpose of the networking application and/or the prerogative of a system designer. In light of the fact that traffic demands vary over time, we need to update our resource allocation solution from time to time. As such, we consider that the node cooperation between a source node and its partner is committed merely for an active resource allocation interval.

Problem Formulation: Consider the following non-altruistic node cooperative resource allocation optimization problem (NCRAOP)

$$\max_{\mathbf{c},\mathbf{p},\mathbf{a},\mathbf{z}} \left\{ \sum_{m=1}^{M} U_m \left(R_m \left(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z} \right) | \mathbf{\Theta} \right) \right\} \tag{5.14}$$

$$\text{subject to} \quad R_m \left(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z} \right) \geq R_m^d, \forall_m \tag{5.15}$$

$$\sum_{n=1}^{N} p_{m,n}^l \leq P_m^{\max}, \forall_{m,l} \tag{5.16}$$

$$\sum_{m=1}^{M} c_{m,n}^l \leq 1, \forall_{n,l} \tag{5.17}$$

$$\sum_{m=1,m\neq u}^{M} z_{mu} \leq 1, \forall_u \tag{5.18}$$

$$\sum_{u=1,u\neq m}^{M} z_{mu} \leq 1, \forall_m \tag{5.19}$$

$$\sum_{m=1}^{M} z_{mu} a_{mu} = 1, \forall_u \tag{5.20}$$

$$p_{m,n}^l \geq 0, \forall_{m,n,l} \tag{5.21}$$

$$a_{mu} \geq 0, \forall_{m,u} \tag{5.22}$$

$$c_{m,n}^l \in \{0,1\}, \forall_{m,n,l} \tag{5.23}$$

$$z_{mu} \in \{0,1\}, \forall_{m,u} \tag{5.24}$$

$$z_{mm} = 1, \forall_m, \tag{5.25}$$

where $\mathbf{c}, \mathbf{p}, \mathbf{a}$, and $\mathbf{z}$ are the *optimization variables*. By reducing the well-known NP-complete *number partitioning problem* to the NCRAOP, it can be proved that the NCRAOP is an

NP-hard problem.

Proposition 5 *The NCRAOP is an NP-hard problem.*

5.5 Proposed Resource Allocation Approaches with Node Cooperation

5.5.1 KKT Interpretations

In general, solving the NP-hard NCRAOP requires exponential time complexity [9]. To design an efficient and effective resource allocation approach to solve the NCRAOP, we first investigate the relationships among different optimization variables by means of the KKT interpretations [9]. Relaxing the integer constraints (5.23) and (5.24) leads to

$$c_{m,n}^{l} \geq 0, \forall_{m,n,l} \tag{5.26}$$

and

$$z_{mu} \geq 0, \forall_{m,u} \tag{5.27}$$

respectively. Therefore, the Lagrangian of the NCRAOP is given by

$$\begin{aligned} L\left(\mathbf{c},\mathbf{p},\mathbf{a},\mathbf{z},\boldsymbol{\xi}\right) = & -\sum_{m=1}^{M} U_m\left(R_m\left(\mathbf{c},\mathbf{p},\mathbf{a},\mathbf{z}\right)|\boldsymbol{\Theta}\right) + \sum_{m=1}^{M} {\xi^{(1)}}_m \left(R_m^d - R_m\left(\mathbf{c},\mathbf{p},\mathbf{a},\mathbf{z}\right)\right) \\ & + \sum_{m=1}^{M}\sum_{l=1}^{L} {\xi^{(2)}}_m^l \left(\sum_{n=1}^{N} p_{m,n}^l - P_m^{\max}\right) + \sum_{n=1}^{N}\sum_{l=1}^{L} {\xi^{(3)}}_n^l \left(\sum_{m=1}^{M} c_{m,n}^l - 1\right) \\ & + \sum_{u=1}^{M} {\xi^{(4)}}_u \left(\sum_{m=1,m\neq u}^{M} z_{mu} - 1\right) + \sum_{m=1}^{M} {\xi^{(5)}}_m \left(\sum_{u=1,u\neq m}^{M} z_{mu} - 1\right) \\ & + \sum_{u=1}^{M} {\xi^{(6)}}_u \left(\sum_{m=1}^{M} z_{mu}a_{mu} - 1\right) - \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{l=1}^{L} {\xi^{(7)}}_{m,n}^l p_{m,n}^l - \sum_{m=1}^{M}\sum_{u=1}^{M} {\xi^{(8)}}_{mu} a_{mu} \\ & - \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{l=1}^{L} {\xi^{(9)}}_{m,n}^l c_{m,n}^l - \sum_{m=1}^{M}\sum_{u=1}^{M} {\xi^{(10)}}_{mu} z_{mu} + \sum_{m=1}^{M} {\xi^{(11)}}_m \left(z_{mm} - 1\right) \end{aligned} \tag{5.28}$$

where ${\xi^{(1)}}_m$, ${\xi^{(2)}}^l_m$, ${\xi^{(3)}}^l_n$, ${\xi^{(4)}}_u$, ${\xi^{(5)}}_m$, ${\xi^{(6)}}_u$, ${\xi^{(7)}}^l_{m,n}$, ${\xi^{(8)}}_{mu}$, ${\xi^{(9)}}^l_{m,n}$, ${\xi^{(10)}}_{mu}$, and ${\xi^{(11)}}_m$ are the Lagrange multipliers for the constraints (5.15), (5.16), (5.17), (5.18), (5.19), (5.20), (5.21), (5.22), (5.23), (5.24), and (5.25), respectively. To investigate the (necessary) conditions for optimality of the solution, given $\mathbf{\Theta}$, consider part of the KKT conditions of the relaxed problem with respect to $c^l_{m,n}$, z_{mu}, and $p^l_{m,n}$,

$$-\left(U'_m\left(R_m\left(\cdot\right)|\mathbf{\Theta}\right)+{\xi^{(1)}}_m\right)\frac{\partial R_m\left(\cdot\right)}{\partial c^l_{m,n}}+{\xi^{(3)}}^l_n={\xi^{(9)}}^l_{m,n},\forall m,n,l \tag{5.29}$$

$$-\left(U'_m\left(R_m\left(\cdot\right)|\mathbf{\Theta}\right)+{\xi^{(1)}}_m\right)\frac{\partial R_m\left(\cdot\right)}{\partial z_{mu}}+{\xi^{(4)}}_u+{\xi^{(5)}}_m+{\xi^{(6)}}_u a_{mu}={\xi^{(10)}}_{mu},\forall m,u,n,l \tag{5.30}$$

$$-\sum_{u=1}^{M}\left(U'_u\left(R_u(\cdot)|\mathbf{\Theta}\right)+{\xi^{(1)}}_u\right)\frac{\partial R_u(\cdot)}{\partial p^l_{m,n}}+{\xi^{(2)}}^l_m={\xi^{(7)}}^l_{m,n},\forall m,n,l \tag{5.31}$$

$${\xi^{(9)}}^l_{m,n}c^l_{m,n}=0,\forall_{m,n,l} \tag{5.32}$$

$${\xi^{(10)}}_{mu}z_{mu}=0,\forall_{m,u} \tag{5.33}$$

$${\xi^{(7)}}^l_{m,n}p^l_{m,n}=0,\forall_{m,n,l}. \tag{5.34}$$

Proposition 6 *(**Optimal Power Allocation**) Let*

$$f(p^l_{m,n})={\xi^{(2)}}^l_m-\sum_{u=1}^{M}\left(U'_u\left(R_u\left(\cdot\right)|\mathbf{\Theta}\right)+{\xi^{(1)}}_u\right)\frac{\partial R_u\left(\cdot\right)}{\partial p^l_{m,n}},\forall_{m,n,l}. \tag{5.35}$$

The optimal power allocation, denoted by $\mathbf{p}^*$*, for the NCRAOP is given by* ${p^*}^l_{m,n}=\max\{\tilde{p}^l_{m,n},0\},\forall_{m,n,l}$*, where* $\tilde{p}^l_{m,n}$ *is the solution of the function* $f(\tilde{p}^l_{m,n})=0$*.*

Proposition 7 *(**Subcarrier Allocation Criterion**) For the* n^{th} *subcarrier and the* l^{th} *timeslot, the necessary condition for* $c^l_{m,n}$ *being positive is*

$$m=\arg\max_{\tilde{m}}\left\{\left(U'_{\tilde{m}}\left(R_{\tilde{m}}\left(\cdot\right)|\mathbf{\Theta}\right)+{\xi^{(1)}}_{\tilde{m}}\right)\frac{\partial R_{\tilde{m}}\left(\cdot\right)}{\partial c^l_{\tilde{m},n}}\right\}. \tag{5.36}$$

Assuming $R_m\left(\mathbf{c},\mathbf{p},\mathbf{a},\mathbf{z}\right)\neq R^d_m$ (i.e., ${\xi^{(1)}}_m=0$), the subcarrier allocation criterion can be further simplified as follows. For the n^{th} subcarrier and the l^{th} timeslot, choose m^* such that

$$m^*=\arg\max_{m}\left\{U'_m\left(R_m\left(\cdot\right)|\mathbf{\Theta}\right)\frac{\partial R_m\left(\cdot\right)}{\partial c^l_{m,n}}\right\} \tag{5.37}$$

and set $c^l_{m^*,n}=1$.

Proposition 8 ***(Partner Selection Criterion)*** *The necessary condition for* z_{mu} *being positive is*

$$u = \arg\max_{\tilde{u}\neq m} \left\{ \frac{1}{a_{m\tilde{u}}} \left(\left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_m \right) \frac{\partial R_m \left(\cdot \right)}{\partial z_{m\tilde{u}}} - {\xi^{(4)}}_{\tilde{u}} - {\xi^{(5)}}_m \right) \right\}. \tag{5.38}$$

Assuming $R_m \left(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z} \right) \neq R_m^d$ (i.e., ${\xi^{(1)}}_m = 0$), and ignoring the constraints (5.18) and (5.19) (i.e., ${\xi^{(4)}}_u = {\xi^{(5)}}_m = 0$), the partner selection criterion can be deduced as follows. For the m^{th} node, choose u^* such that

$$u^* = \arg\max_{u\neq m} \left\{ \frac{U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right)}{a_{mu}} \frac{\partial R_m \left(\cdot \right)}{\partial z_{mu}} \right\} \tag{5.39}$$

and set $z_{mu^*} = 1$.

Corollary 2 ***(Partner Allocation Criterion)*** *The necessary condition for* z_{mu} *being positive is*

$$m = \arg\max_{\tilde{m}\neq u} \left\{ \frac{1}{a_{\tilde{m}u}} \left(\left(U'_{\tilde{m}} \left(R_{\tilde{m}} \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{\tilde{m}} \right) \frac{\partial R_{\tilde{m}} \left(\cdot \right)}{\partial z_{\tilde{m}u}} - {\xi^{(4)}}_u - {\xi^{(5)}}_{\tilde{m}} \right) \right\}. \tag{5.40}$$

Assuming $R_m \left(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z} \right) \neq R_m^d$ (i.e., ${\xi^{(1)}}_m = 0$), and ignoring the constraints (5.18) and (5.19) (i.e., ${\xi^{(4)}}_u = {\xi^{(5)}}_m = 0$), the partner allocation criterion can be deduced as follows. For the u^{th} node, choose m^* such that

$$m^* = \arg\max_{m\neq u} \left\{ \frac{U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right)}{a_{mu}} \frac{\partial R_m \left(\cdot \right)}{\partial z_{mu}} \right\} \tag{5.41}$$

and set $z_{m^*u} = 1$.

Notice that, in general, the results obtained from the partner selection criterion and that from the partner allocation criterion are not equal. The partner selection criterion refers to choosing the best relay for a particular source node, whereas the partner allocation criterion refers to allocating a relay node to the best source node. In a nutshell, the partner selection criterion is to maximize the node-wise utility (i.e., $U_m(R_m(\cdot))$), while the partner allocation criterion is to maximize the network-wise utility (i.e., $\sum_m U_m(R_m(\cdot))$).

Proposition 9 *(**Power Allocation Criterion for Assisted Transmissions**) For the u^{th} node, the sufficient condition for a_{mu} being positive is $a_{uj} > 0$ (and $z_{uj} = 1$), $\exists m, j$.*

Although a transmit power level for an advantageous cooperative transmission can be deduced by Proposition 9, in the case of non-altruistic node cooperation, with centralized control, the portion of transmit power used for assisted transmissions is usually governed by the system designer for the sake of network-wise system performance.

5.5.2 Examples

To better understand the conceptual implications of our subcarrier allocation criterion and partner selection criterion, we consider three different objective functions, namely 1) system throughput maximization (i.e., $\mathbf{\Theta} = 1$), 2) proportional fairness (i.e., $\mathbf{\Theta} = 2$ and $\kappa = 1$), and 3) max-min fairness (i.e., $\mathbf{\Theta} = 2$ and $\kappa \to \infty$).

System Throughput Maximization (i.e., $\mathbf{\Theta} = 1$)

$U_m\left(R_m\left(\cdot\right)|\mathbf{\Theta}\right) = R_m\left(\cdot\right)$ and hence $U'_m\left(R_m\left(\cdot\right)|\mathbf{\Theta}\right) = 1$; therefore, the subcarrier allocation criterion given in (5.37) and the partner selection criterion given in (5.39) become

$$m^* = \arg\max_m \left\{ \frac{\partial R_m\left(\cdot\right)}{\partial c^l_{m,n}} \right\} \tag{5.42}$$

$$= \arg\max_m \left\{ \log_2 \left(\left(1 + a_{mm} g^l_{mm,n} p^l_{m,n}\right)^2 + \sum_{u=1, u\neq m}^{M} z_{mu} a_{mu} g^l_{mu,n} p^l_{u,n} \right) \right\} \tag{5.43}$$

and

$$u^* = \arg\max_{u\neq m} \left\{ \frac{1}{a_{mu}} \frac{\partial R_m\left(\cdot\right)}{\partial z_{mu}} \right\} \tag{5.44}$$

$$= \arg\max_{u\neq m} \left\{ \sum_{n=1}^{N} \sum_{l=1}^{L} \frac{c^l_{m,n} g^l_{mu,n} p^l_{u,n}}{\left(1 + a_{mm} g^l_{mm,n} p^l_{m,n}\right)^2 + \sum_{u\neq m} z_{mu} a_{mu} g^l_{mu,n} p^l_{u,n}} \right\} \tag{5.45}$$

respectively. Thus, given the power allocation (i.e., $\mathbf{a}, \mathbf{p}$) and partner selection (i.e., $\mathbf{z}$), the condition (5.43) implies that the larger throughput the m^{th} node can contribute over the n^{th} subcarrier on the l^{th} timeslot, the better the n^{th} subcarrier on the l^{th} timeslot is assigned to the m^{th} node (i.e., $c^l_{m,n} = 1, \exists_{m,n,l}$). Likewise, given the power allocation and subcarrier allocation (i.e., $\mathbf{c}, \mathbf{p}, \mathbf{a}$), the condition (5.45) indicates that we should choose a partner who can contribute the largest marginal increase in the achievable data rate. Therefore, both criteria agree with respect to the notion of throughput maximization. Notice that, in the case of non-cooperation (i.e., $z_{mu} = 0$), the condition (5.43) reduces to the well-known subcarrier allocation criterion for a throughput maximization problem (e.g., [21]).

Proportional Fairness (i.e., $\mathbf{\Theta} = 2$ and $\kappa = 1$)

$U_m\left(R_m\left(\cdot\right)|\mathbf{\Theta}\right) = \ln\left(\frac{R_m(\cdot)}{A}\right)$ and hence $U'_m\left(R_m\left(\cdot\right)|\mathbf{\Theta}\right) = 1/R_m\left(\cdot\right)$; therefore, the conditions (5.37) and (5.39) become

$$m^* = \arg\max_m \left\{ \frac{1}{R_m\left(\cdot\right)} \frac{\partial R_m\left(\cdot\right)}{\partial c^l_{m,n}} \right\} \tag{5.46}$$

$$= \arg\max_m \left\{ \frac{1}{R_m\left(\cdot\right)} \log_2 \left(\left(1 + a_{mm} g^l_{mm,n} p^l_{m,n}\right)^2 + \sum_{u=1, u\neq m}^{M} z_{mu} a_{mu} g^l_{mu,n} p^l_{u,n} \right) \right\} \tag{5.47}$$

and

$$u^* = \arg\max_{u\neq m} \left\{ \frac{1}{a_{mu} R_m\left(\cdot\right)} \frac{\partial R_m\left(\cdot\right)}{\partial z_{mu}} \right\} \tag{5.48}$$

$$= \arg\max_{u\neq m} \left\{ \frac{1}{R_m\left(\cdot\right)} \sum_{n=1}^{N} \sum_{l=1}^{L} \frac{c^l_{m,n} g^l_{mu,n} p^l_{u,n}}{\left(1 + a_{mm} g^l_{mm,n} p^l_{m,n}\right)^2 + \sum_{u\neq m} z_{mu} a_{mu} g^l_{mu,n} p^l_{u,n}} \right\}, \tag{5.49}$$

respectively. As seen, it is more likely for a mesh node to get an extra subcarrier and/or a partner if its data rate obtained is small (i.e., small $R_m(\cdot)$), whereas it is less likely to assign an extra subcarrier or a partner to a node whose data rate obtained is already very high (i.e., large $R_m(\cdot)$). Thus, both criteria match with the notion of proportional fairness.

Max-min Fairness (i.e., $\boldsymbol{\Theta} = 2$ and $\kappa \to \infty$)

$U_m\left(R_m\left(\cdot\right)|\boldsymbol{\Theta}\right) = -\left[-\ln\left(\frac{R_m(\mathbf{c},\mathbf{p},\mathbf{a},\mathbf{z})}{A}\right)\right]^{\kappa}$, and hence $U'_m\left(R_m\left(\cdot\right)|\boldsymbol{\Theta}\right) = \kappa\left[-\ln\left(R_m(\cdot)/A\right)\right]^{\kappa-1}/R_m(\cdot)$. Since $0 < R_m(\cdot)/A < 1$, $\left[-\ln\left(R_m(\cdot)/A\right)\right]^{\kappa-1}$ becomes a dominant term as $\kappa \to \infty$. Hence, the conditions (5.37) and (5.39) become

$$m^* = \arg\min_{m} R_m(\cdot) \tag{5.50}$$

and

$$u^* = \arg\min_{u \neq m} R_m(\cdot) \tag{5.51}$$

respectively. From (5.50) and (5.51), we tend to assign subcarriers and partners to the nodes which have minimal achievable data rates, realizing the notion of max-min fairness.

In the following subsections, we propose a centralized resource allocation approach and a distributed resource allocation approach to solve the NCRAOP in Section 5.5.3 and Section 5.5.4, respectively.

5.5.3 Proposed Approach with Centralized Control

In order to obtain the global optimal solutions to the NCRAOP, subcarrier allocation, partner selection/allocation, and power allocation should be jointly considered, which brings about very high computational cost. To devise an efficient yet effective node cooperative resource allocation strategy with centralized control, we propose a four-phase resource allocation approach with QoS assurance and service differentiation: In Phase 1, we fix the power allocation and then solve the NCRAOP without considering node cooperation; In Phase 2, we perform water filling for power allocation to improve the system performance; In Phase 3, we allow node cooperation, if feasible and favorable, so as to add an additional performance gain to the system; In Phase 4, we perform water filling for power allocation again providing

the solutions of subcarrier allocation and partner allocation, thereby further improving the system performance. Note that, in this centralized approach, we consider that there is a clusterhead which collects the traffic demands from its clustermembers and executes our proposed centralized resource allocation.

Phase-1

With uniform power allocation and no node cooperation (i.e., $z_{mu} = 0$ and $a_{mu} = 0, \forall m \neq u$), the NCRAOP can be reduced to the following resource allocation optimization problem

$$\max_{\mathbf{c}} \left\{ \sum_{m=1}^{M} U_m \left(R_m \left(\mathbf{c} \right) | \mathbf{\Theta} \right) \right\} \tag{5.52}$$

$$\text{subject to} \quad R_m \left(\mathbf{c} \right) \geq R_m^d, \forall_m \tag{5.53}$$

$$\sum_{m=1}^{M} c_{m,n}^l \leq 1, \forall_{n,l} \tag{5.54}$$

$$c_{m,n}^l \in \{0, 1\}, \forall_{m,n,l} \tag{5.55}$$

and the subcarrier allocation criterion given in (5.36) is as follows. For the n^{th} subcarrier and the l^{th} timeslot, choose m^* such that $m^* = \arg\max_m \left\{ \left(U_m' \left(R_m \left(\mathbf{c} \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_m \right) \log_2 \left(1 + g_{mm,n}^l p_{m,n}^l \right) \right\}$. The variable ${\xi^{(1)}}_m$ is updated iteratively by ${\xi^{(1)}}_m = \max \left\{ 0, {\xi^{(1)}}_m - s_m^{(\mathrm{k})} d_m \right\}$, with $s_m^{(\mathrm{k})}$ being the step size at the k^{th} iteration for the m^{th} node and $d_m = R_m(\mathbf{c}) - R_m^d$. With CAC in place, the approach terminates when all the subcarriers are allocated and all the rate requirements are met (i.e., ${\xi^{(1)}}_m = 0, \forall m$), and the subcarrier allocation solution $\mathbf{c}^*$ is obtained such that $\sum_{m=1}^{M} {c^*}_{m,n}^l = 1, \forall n, l$.

Phase-2

It is well known that, for fixed subcarrier allocation, system performance can be improved by optimal power allocation, which can be carried out by means of utility-based water-filling [101]. With the subcarrier allocation obtained in Phase 1, the NCRAOP without

considering node cooperation becomes

$$\max_{\mathbf{p}} \left\{ \sum_{m=1}^{M} U_m \left(R_m \left(\mathbf{p} \right) | \mathbf{\Theta} \right) \right\} \tag{5.56}$$

$$\text{subject to} \quad R_m \left(\mathbf{p} \right) \geq R_m^d, \forall_m \tag{5.57}$$

$$\sum_{n=1}^{N} p_{m,n}^l \leq P_m^{\max}, \forall_{m,l} \tag{5.58}$$

$$p_{m,n}^l \geq 0, \forall_{m,n,l}. \tag{5.59}$$

Proposition 10 *With the subcarrier allocation solution, denoted by* $\tilde{\mathbf{c}}$*, the optimal power allocation solution, denoted by* $\mathbf{p}^*$*, for the optimization problem given in (5.56)-(5.59) is given*

$$p^{*l}_{m,n} = \tilde{c}^l_{m,n} \left[\frac{U'_m(R_m(\mathbf{p})|\mathbf{\Theta}) + \xi^{(1)}{}_m}{\xi^{(2)l}_m} - \frac{1}{g^l_{mm,n}} \right]^+, \forall_{m,n,l} \tag{5.60}$$

where $[x]^+ = \max\{0, x\}$.

Let $R_m(\hat{\mathbf{p}})$ and $R_m(\mathbf{p}^*)$ be the achievable data rate obtained of the m^{th} node with uniform power allocation (i.e., $\hat{\mathbf{p}}$) and that with optimal power allocation (i.e., $\mathbf{p}^*$), respectively. Since $R_m(\mathbf{p}^*) \geq R_m(\hat{\mathbf{p}})$ [101], the QoS demands of the mesh nodes can still be met after transmit power is allocated according to water filling. In a nutshell, Phase-2 resource allocation further improves the Phase-1 resource allocation solutions by employing optimal power allocation, conducing to improved system performance.

Phase-3

In this phase, we investigate the performance gain due to feasible and favorable node cooperation. Notice that partner *allocation* is feasible in our centralized approach, for a clusterhead can have complete knowledge on which mesh nodes can be the potential partners for a particular mesh node. On the contrary, applying the notion of partner *selection* to the WMNs

with distributed control is more pragmatic because of the characteristics of asymmetric non-altruistic node cooperation, to be discussed in Section 5.5.4. Nevertheless, with the subcarrier allocation solution obtained in Phase 1 and the power allocation solution obtained in Phase 2, the NCRAOP becomes

$$\max_{\mathbf{a},\mathbf{z}} \left\{ \sum_{m=1}^{M} U_m \left(R_m \left(\mathbf{a}, \mathbf{z} \right) | \mathbf{\Theta} \right) \right\} \tag{5.61}$$

$$\text{subject to} \quad R_m \left(\mathbf{a}, \mathbf{z} \right) \geq R_m^d, \forall_m \tag{5.62}$$

$$\sum_{m=1, m \neq u}^{M} z_{mu} \leq 1, \forall_u \tag{5.63}$$

$$\sum_{u=1, u \neq m}^{M} z_{mu} \leq 1, \forall_m \tag{5.64}$$

$$\sum_{m=1}^{M} z_{mu} a_{mu} = 1, \forall_u \tag{5.65}$$

$$a_{mu} \geq 0, \forall_{m,u} \tag{5.66}$$

$$z_{mu} \in \{0, 1\}, \forall_{m,u} \tag{5.67}$$

$$z_{mm} = 1, \forall_m. \tag{5.68}$$

Denote ρ (≥ 0) as a tunable system parameter to balance node cooperation and node non-cooperation, i.e., $a_{mu} = \rho a_{uu}, \exists m \neq u$. In fact, ρ indicates the *willingness* of a node to assist another mesh node's transmissions, i.e., the smaller the value of ρ, the less eager is a node to assist the m^{th} node for some m. For some u, if $\sum_{m \neq u} z_{mu} = 1$, then $a_{mu} = \rho a_{uu} (> 0), \exists m \neq u$; otherwise, $a_{uu} = 1$ and $a_{mu} = 0, \forall m \neq u$. On the other hand, in the case of non-altruistic node cooperation, it is conceivable that the value of ρ should be upper bounded (e.g., $\rho \leq 0.5$), as all the mesh nodes have their own data to transmit. Suppose the u^{th} node is to assist the m^{th} node, i.e., $z_{mu} = 1$. Then, the transmit power of the u^{th} node left for cooperation, denoted by P_u^{left}, is given by $P_u^{\text{left}} = P_u^{\max} - a_{uu} \sum_{n=1}^{N} p_{u,n}^l$, where $p_{u,n}^l$ is obtained from Phase-2. Let $\mathcal{N}_m^l$ be the set of subcarriers allocated to the m^{th} node on the l^{th} timeslot. If the transmit power of the u^{th} node dedicated to node cooperation is uniformly distributed over the subcarriers

allocated to the m^{th} node on the l^{th} timeslot, then $\hat{p}^l_{u,n} = P_u^{\max}/\sum_{n\in\mathcal{N}_m^l} c^l_{m,n}$. Thus, the portion of the transmit power of the u^{th} node in assisting the m^{th} node's transmissions is $a_{mu}\sum_{n\in\mathcal{N}_m^l}\hat{p}^l_{u,n}$.

In the presence of different traffic classes, service differentiation is indispensable for effective QoS provisioning, where RG nodes are assigned higher priority over BE nodes. As a consequence, node cooperation should also be prioritized. In particular, we consider that only RG nodes can receive assistance from either BE nodes or other RG nodes. Here, we further divide our Phase-3 resource allocation into two steps, namely 1) BE-assisting-RG and 2) RG-assisting-RG.

Step 1: In the case of BE-assisting-RG, allow BE nodes to assist RG nodes whenever favorable. Let $\mathcal{A} = \left\{m \mid m \in \mathcal{M}_1, \sum_{j\neq m} z_{mj} = 0\right\}$ and $\mathcal{B} = \left\{u \mid u \in \mathcal{M}_2, \sum_{i\neq u} z_{iu} = 0\right\}$. Set $\mathcal{A}$ consists of the RG nodes which do not receive any assistance from any other nodes, whereas set $\mathcal{B}$ consists of the BE nodes which do not offer any assistance to any nodes. The partner allocation criterion is as follows. For the u^{th} node, $u \in \mathcal{B}$, choose m^* such that $m^* = \arg\max_{m\in\mathcal{A}}\left\{\frac{U'_m(R_m(\mathbf{a},\mathbf{z})|\mathbf{\Theta})}{a_{mu}}\frac{\partial R_m(\mathbf{a},\mathbf{z})}{\partial z_{mu}}\right\}$. Then, check if this partner allocation process can help enhance the total utilities. Let $R_m(\mathbf{a},\mathbf{z})$ and $R_m(\tilde{\mathbf{a}},\tilde{\mathbf{z}})$ be the achievable data rate obtained of the m^{th} node without node cooperation and that with node cooperation, respectively. We have

$$R_{m^*}(\mathbf{a},\mathbf{z}) = \sum_{n=1}^{N}\sum_{l=1}^{L} c^l_{m^*,n}\log_2\left(1 + g^l_{m^*m^*,n}p^l_{m^*,n}\right) \tag{5.69}$$

$$R_{m^*}(\tilde{\mathbf{a}},\tilde{\mathbf{z}}) = \sum_{n=1}^{N}\sum_{l=1}^{L}\frac{1}{2}c^l_{m^*,n}\log_2\left(\left(1 + g^l_{m^*m^*,n}p^l_{m^*,n}\right)^2 + a_{m^*u}g^l_{m^*u,n}\hat{p}^l_{u,n}\right) > R_{m^*}(\mathbf{a},\mathbf{z}) \tag{5.70}$$

$$R_u(\mathbf{a},\mathbf{z}) = \sum_{n=1}^{N}\sum_{l=1}^{L} c^l_{u,n}\log_2\left(1 + g^l_{uu,n}p^l_{u,n}\right) \tag{5.71}$$

$$R_u(\tilde{\mathbf{a}},\tilde{\mathbf{z}}) = \sum_{n=1}^{N}\sum_{l=1}^{L} c^l_{u,n}\log_2\left(1 + a_{uu}g^l_{uu,n}p^l_{u,n}\right) < R_u(\mathbf{a},\mathbf{z}). \tag{5.72}$$

Since $R_u^d = 0, \forall u \in \mathcal{M}_2$, the solutions obtained by considering node cooperation are also feasible to the NCRAOP. Set $z_{m^*u} = 1$ and remove the m^{*th} node from $\mathcal{A}$ (i.e., $\mathcal{A} \leftarrow \mathcal{A} - \{m^*\}$) if the following condition is valid:

$$U_{m^*}\left(R_{m^*}(\tilde{\mathbf{a}}, \tilde{\mathbf{z}})|\mathbf{\Theta}\right) + U_u\left(R_u(\tilde{\mathbf{a}}, \tilde{\mathbf{z}})|\mathbf{\Theta}\right) > U_{m^*}\left(R_{m^*}(\mathbf{a}, \mathbf{z})|\mathbf{\Theta}\right) + U_u\left(R_u(\mathbf{a}, \mathbf{z})|\mathbf{\Theta}\right). \quad (5.73)$$

If (5.73) is satisfied, it means that assigning the u^{th} node to the m^{*th} node as its partner can increase the total utilities, thereby improving the network-wise system performance. The u^{th} node is removed from $\mathcal{B}$ (i.e., $\mathcal{B} \leftarrow \mathcal{B} - \{u\}$), and the process repeats until $\mathcal{A} = \{\phi\}$ or $\mathcal{B} = \{\phi\}$.

Step 2: In the case of RG-assisting-RG, allow an RG node to assist other RG nodes whenever favorable. Partner allocation in this step becomes convoluted, for an RG node decreases its achievable data rate and plausibly voids its own rate constraint (given in (5.11)) when assisting other nodes. Let $\mathcal{C} = \left\{m \mid m \in \mathcal{M}_1, \sum_{j \neq m} z_{mj} = 0\right\}$ and $\mathcal{D} = \left\{j \mid j \in \mathcal{M}_1, \sum_{i \neq j} z_{ij} = 0\right\}$. Set $\mathcal{C}$ consists of the RG nodes which do not receive any assistance from any other nodes, whereas set $\mathcal{D}$ consists of the RG nodes which do not offer any assistance to any RG nodes. For $j \in \mathcal{D}$, choose m^* such that $m^* = \arg\max_{m \in \mathcal{C} \setminus \{j\}} \left\{ \frac{U'_m(R_m(\mathbf{a},\mathbf{z})|\mathbf{\Theta})}{a_{mj}} \frac{\partial R_m(\mathbf{a},\mathbf{z})}{\partial z_{mj}} \right\}$. Then, check if the new resource allocation solutions are still feasible for the NCRAOP and can increase the total utilities. Let $(\mathbf{a}, \mathbf{z})$ and $(\tilde{\mathbf{a}}, \tilde{\mathbf{z}})$ denote the current node cooperative resource allocation solution and the new node cooperative resource allocation solution, respectively. We have

$$R_{m^*}(\mathbf{a}, \mathbf{z}) = \sum_{n=1}^{N} \sum_{l=1}^{L} c^l_{m^*,n} \log_2 \left(1 + a_{m^*m^*} g^l_{m^*m^*,n} p^l_{m^*,n}\right) \quad (5.74)$$

$$R_{m^*}(\tilde{\mathbf{a}}, \tilde{\mathbf{z}}) = \sum_{n=1}^{N} \sum_{l=1}^{L} \frac{1}{2} c^l_{m^*,n} \log_2 \left(\left(1 + a_{m^*m^*} g^l_{m^*m^*,n} p^l_{m^*,n}\right)^2 + a_{m^*j} g^l_{m^*j,n} \hat{p}^l_{j,n}\right) > R_{m^*}(\mathbf{a}, \mathbf{z}) \quad (5.75)$$

$$R_j(\mathbf{a},\mathbf{z}) = \sum_{n=1}^{N}\sum_{l=1}^{L}\frac{1}{2}c_{j,n}^{l}\log_2\left(\left(1+g_{jj,n}^{l}p_{j,n}^{l}\right)^2+\sum_{i\neq j}z_{ji}a_{ji}g_{ji,n}^{l}\hat{p}_{i,n}^{l}\right) \tag{5.76}$$

$$R_j(\tilde{\mathbf{a}},\tilde{\mathbf{z}}) = \sum_{n=1}^{N}\sum_{l=1}^{L}\frac{1}{2}c_{j,n}^{l}\log_2\left(\left(1+a_{jj}g_{jj,n}^{l}p_{j,n}^{l}\right)^2+\sum_{i\neq j}z_{ji}a_{ji}g_{ji,n}^{l}\hat{p}_{i,n}^{l}\right) < R_j(\mathbf{a},\mathbf{z}). \tag{5.77}$$

Set $z_{m^*j}=1$ and remove the m^{*th} node from $\mathcal{C}$ (i.e., $\mathcal{C}\leftarrow\mathcal{C}-\{m^*\}$) if the following two conditions are valid:

$$U_{m^*}\left(R_{m^*}(\tilde{\mathbf{a}},\tilde{\mathbf{z}})|\mathbf{\Theta}\right)+U_j\left(R_j(\tilde{\mathbf{a}},\tilde{\mathbf{z}})|\mathbf{\Theta}\right) > U_{m^*}\left(R_{m^*}(\mathbf{a},\mathbf{z})|\mathbf{\Theta}\right)+U_j\left(R_j(\mathbf{a},\mathbf{z})|\mathbf{\Theta}\right) \tag{5.78}$$

$$R_j(\tilde{\mathbf{a}},\tilde{\mathbf{z}}) \geq R_j^d. \tag{5.79}$$

Since $R_j^d > 0, \forall j \in \mathcal{M}_1$, we have to ensure that the rate constraints for all the RG nodes are not violated due to the aforesaid partner allocation (i.e., condition (5.79)). Condition (5.78) refers to the case where allocating the j^{th} node as a partner to the m^{*th} node can increase the total utilities. The j^{th} node is removed from $\mathcal{D}$ (i.e., $\mathcal{D}\leftarrow\mathcal{D}-\{j\}$), and the process repeats until $\mathcal{C}=\{\phi\}$ or $\mathcal{D}=\{\phi\}$.

With effective node cooperation, Phase-3 resource allocation improves the Phase-2 resource allocation solutions, thereby giving rise to higher total utilities.

Phase-4

The introduction of partner allocation (i.e., $(\mathbf{a},\mathbf{z})$) in Phase 3 changes the achievable rate function. Thus, the power allocation solution obtained in Phase 2 is no longer optimal. With the known solutions for subcarrier allocation and partner allocation, we carry out water filling for power allocation again so as to further improve the system performance. Suppose the u^{th} node is to assist the m^{th} node, i.e., $z_{mu}=1$. For the sake of optimality, water filling for the u^{th} node should be performed over both $\mathcal{N}_u^l$ and $\mathcal{N}_m^l$; however, procuring the optimal power allocation solutions requires a considerable number of recursive computations. Here,

to balance the computational complexity and the system performance, we make use of water filling for power allocation for the m^{th} node over $\mathcal{N}_m^l$ only, $\forall m, l$.

Lemma 2 *With the subcarrier allocation solution and the partner allocation solution, denoted by* $\tilde{\mathbf{c}}$ *and* $(\tilde{\mathbf{a}}, \tilde{\mathbf{z}})$*, respectively, the power allocation solution obtained from water filling, denoted by* $\mathbf{p}^*$*, is given by*

$$p^{*l}_{m,n} = \tilde{c}^l_{m,n} \left[\frac{-\left(2\xi^{(2)^l}_m - \Upsilon \tilde{a}^l_{mm} g^l_{mm,n}\right) + \sqrt{\Upsilon^2 (\tilde{a}^l_{mm})^2 (g^l_{mm,n})^2 - 4(\xi^{(2)^l}_m)^2 \Gamma_m}}{2\xi^{(2)^l}_m \tilde{a}^l_{mm} g^l_{mm,n}} \right]^+, \ \forall m, n, l \tag{5.80}$$

where $\Upsilon = U'_m(R_m(\mathbf{p})|\mathbf{\Theta}) + \xi^{(1)}_m$ *and* $\Gamma_m = \sum_{u \neq m} \tilde{z}_{mu} \tilde{a}_{mu} g^l_{mu,n} \hat{p}^l_{u,n}$.

With the aforementioned optimal power allocation, Phase-4 resource allocation further enhances the Phase-3 resource allocation solution, leading to higher total utilities.

The flowchart of our proposed centralized approach is depicted in Figure 5.3.

5.5.4 Proposed Approach with Distributed Control

When clusterheads are not available, distributed node cooperative resource allocation is essential. We consider that each mesh node can communicate with all other mesh nodes in the same cluster. Here, we propose a two-phase distributed node cooperative resource allocation approach with QoS support and service differentiation. We consider that an active resource allocation interval consists of two phases, namely contention phase and transmission phase. In the contention phase, the approach of Black-Burst jamming can be used for subcarrier contention [18]. In the transmission phase, mesh nodes transmit their data with the consideration of node cooperation, if feasible and favorable.

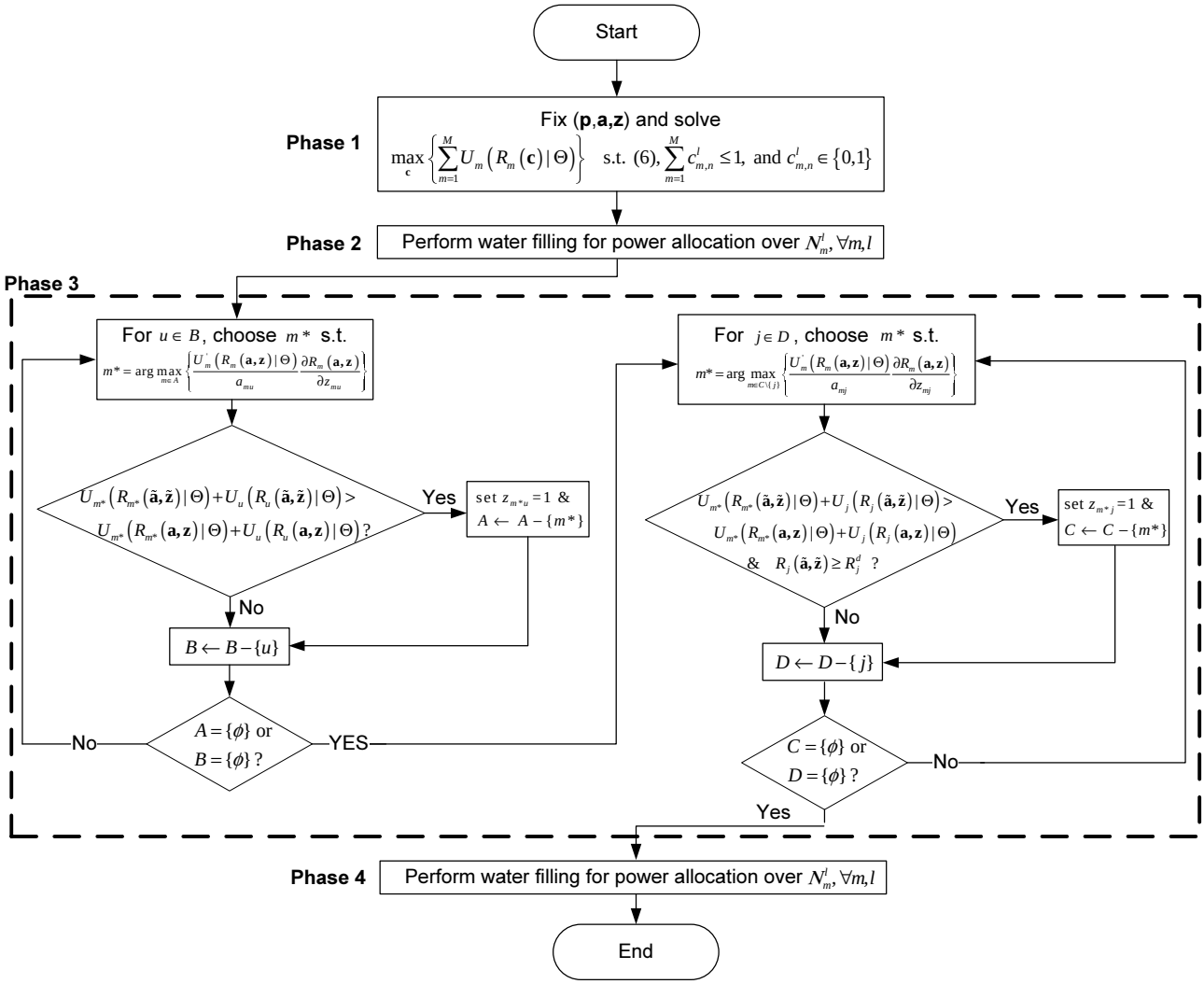
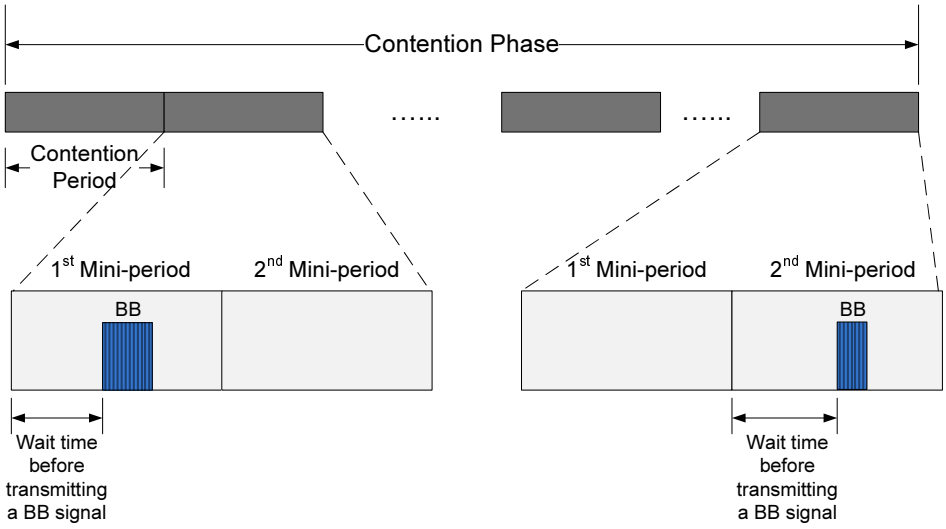


Figure 5.3: Flowchart of the proposed four-phase centralized node cooperative resource allocation approach.

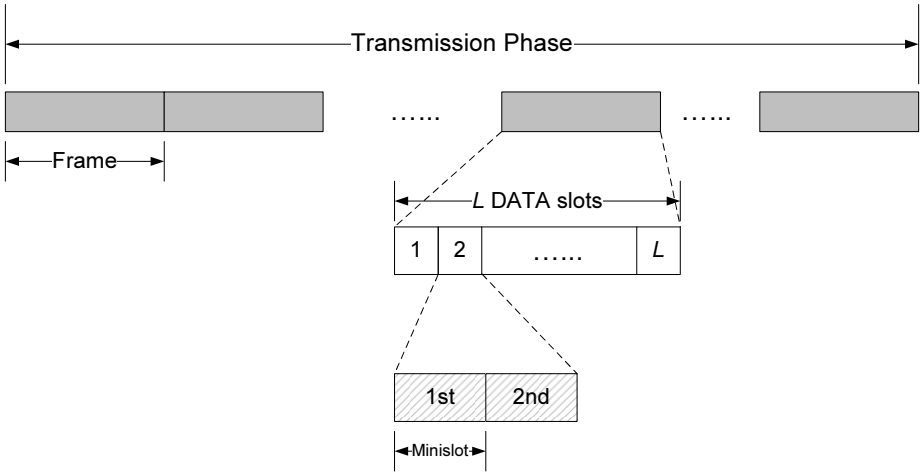
Contention Phase

A contention phase consists of a number of contention periods. For each contention period, the Black-Burst (BB) methodology is used for the subcarrier contention to achieve service differentiation [18]. Let $\mathcal{E} = \{m \mid m \in \mathcal{M}_1, R_m(\cdot) < R_m^d\}$ and $\mathcal{F} = \{m \mid m \in \mathcal{M}_1, R_m(\cdot) \geq R_m^d\} \bigcup \{m \mid m \in \mathcal{M}_2\}$. Set $\mathcal{E}$ consists of the RG nodes whose QoS demands are not satisfied, whereas set $\mathcal{F}$ consists of both the BE nodes and the RG nodes whose QoS demands are met. Each contention period is further divided into two mini-periods, where the first mini-period is dedicated to the contention among the nodes in $\mathcal{E}$, while the second mini-period is dedicated to the contention among the nodes in $\mathcal{F}$. Figure 5.4(a) depicts the dynamics of a contention phase.

At the beginning of each contention period, every node is in a listening mode and waits for a period of time before transmitting its BB signal. In the first mini-period, the waiting time of an RG node in $\mathcal{E}$ is inversely proportional to its minimum required rate. Therefore, the node with the highest minimum required rate sends its BB signal earlier than the other RG nodes in $\mathcal{E}$ so as to win the contention. Other nodes which detect the BB signal remain in the listening mode. An RG node becomes a winner for this contention if it senses an idle channel after transmitting a BB signal. To ensure that each contention period results in only one winner, we assume that the length of a BB signal sent from a node is proportional to its network ID (e.g., MAC address). The winner of this contention period then selects the best subcarriers according to the subcarrier allocation criterion given in (5.36) so as to meet its QoS demand. Notice that uniform power allocation is employed when the subcarriers are selected. After the subcarrier selection is finished, the winner transmits a BB signal over the allocated subcarriers so that the other nodes in a listening mode can detect and record which subcarriers have been chosen. Then, all the other RG nodes wait for the next contention and this process repeats until all the RG nodes in $\mathcal{E}$ have selected their subcarriers. Since we assume that CAC is in place, all the QoS requirements of the RG nodes in $\mathcal{E}$ can be met



(a) An illustration of the subcarrier contention phase among RG nodes and BE nodes.



(b) The structures of a frame and a DATA slot employed in the transmission phase.

Figure 5.4: The phase structures of the proposed two-phase distributed node cooperative resource allocation approach.

at the end of this phase.

Subcarrier contention among the nodes in $\mathcal{F}$ (i.e., satisfied RG nodes and BE nodes) occurs only if there is some unallocated subcarrier(s). If there is no BB detected in the first mini-period, meaning that the subcarrier contention among the RG nodes in $\mathcal{E}$ is complete, every node in $\mathcal{F}$ waits for a period of time in the second mini-period before sending out its BB signal, where the waiting time is inversely proportional to its marginal utility increase when choosing the best available subcarrier (i.e., $\frac{\partial U_m(R_m(\cdot))}{\partial c^l_{m,n}}$). Thus, the larger the marginal utility increases, the more likely that the node will be the winner. The winner then selects the best subcarrier according to the subcarrier allocation criterion given in (5.36). The process repeats until all the subcarriers are used. In fact, this contention phase can be further optimized. For instance, unnecessary first mini-periods should be eliminated after the contention among the unsatisfied RG nodes is finished; addressing this issue, however, is beyond the scope of this research work. Similar to the centralized resource allocation approach, after the subcarrier allocation is determined, each mesh node performs water filling for power allocation independently to further increase both its utility and the system performance.

Transmission Phase

A transmission phase consists of a number of frames, where each frame consists of L DATA slots. Each DATA slot is further divided into two (identical) minislots. Figure 5.4(b) depicts the frame structure used for a transmission phase. Figure 5.5 illustrates a typical (non-altruistic) node cooperation scenario in the transmission phase. In Figure 5.5(a), the m^{th} node is transmitting data to its destination node (i.e., the d^{th} node) in the first minislot. The m^{th} node's transmission can be overheard by the u^{th} node in the first minislot if the u^{th} node is in the listening mode. The u^{th} node then checks if it can decode the m^{th} node's transmissions successfully. If the u^{th} node fails to decode the m^{th} node's transmissions, the

u^{th} node will not help relay the m^{th} node's data (see Figure 5.5(b)). On the other hand, if the u^{th} node can decode the m^{th} node's transmissions reliably, the u^{th} node then becomes the partner of the m^{th} node if condition (5.73) is valid with the u^{th} node being a BE node or conditions (5.78) and (5.79) are valid with the u^{th} node being an RG node (see Figure 5.5(c)). As a matter of fact, the m^{th} node can have more than one partner at a time (e.g., the i^{th} node and u^{th} node in Figure 5.5(c)). In the presence of multiple potential partners, we employ the partner selection scheme proposed in [96] to choose the best partner for the m^{th} node with respect to our partner selection criterion given in (5.38). We also assume that a signal capture mechanism is in place so that a potential partner can only overhear the strongest neighboring node's transmission [113]. This node cooperation between a source node and its partner sustains until the next active interval.

The flowchart of our proposed distributed approach is depicted in Figure 5.6.

5.5.5 Complexity Analysis

Here, we analyze the time complexities of our proposed node cooperative resource allocation approaches. For the proposed centralized approach, the complexities of the Phase-1 resource allocation, the Phase-2 resource allocation, the Phase-3 resource allocation, and the Phase-4 resource allocation are on the order of $O(b'MNL)$, $O(M \max_m\{|\mathcal{N}_m^l|\}L)$, $O(|\mathcal{M}_1||\mathcal{M}_2|+|\mathcal{M}_1|^2)$, and $O(M \max_m\{|\mathcal{N}_m^l|\}L)$, respectively, where b' is a constant. As a consequence, the time complexity of the proposed four-phase centralized approach is $O(bMNL + |\mathcal{M}_1|M)$, where b is a constant. For the proposed two-phase distributed approach, since each mesh node behaves independently, the complexity of resource allocation in the contention phase and that of the transmission phase are on the order of $O(kNL)$ and $O(M)$, respectively, where k is a constant. Thus, the time complexity of the proposed two-phase distributed approach is $O(kNL + M)$. With the help of effective data structure, (e.g., binary tree implementation [112]), the time complexities of our centralize approach and our distributed approach

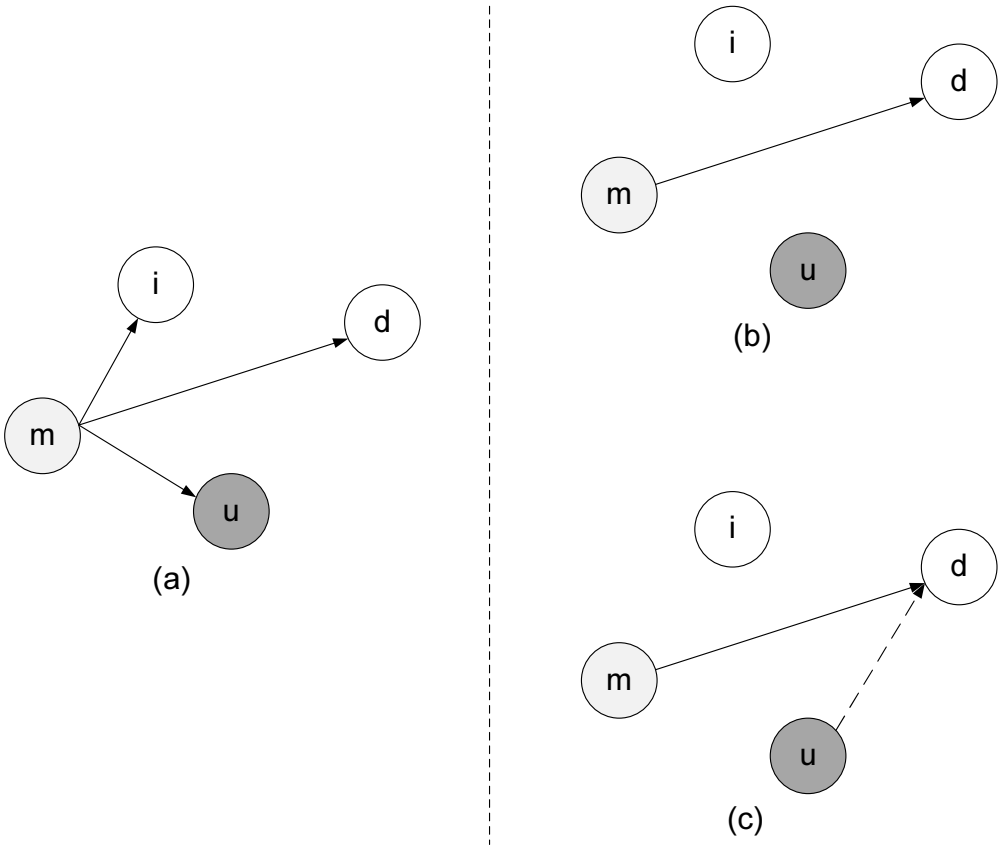


Figure 5.5: Example of a typical node cooperation scenario in the transmission phase.

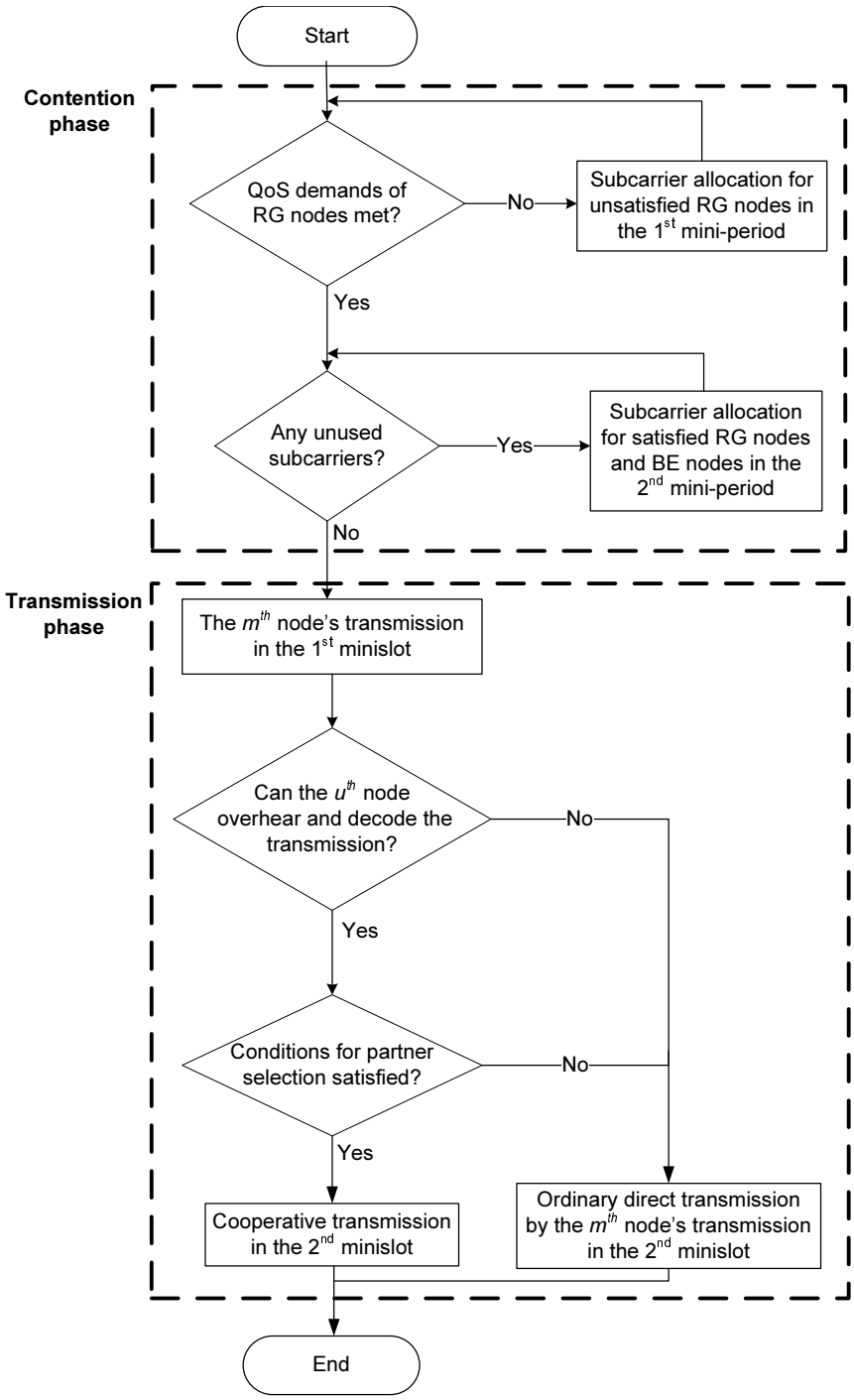


Figure 5.6: Flowchart of the proposed distributed approach.

can be reduced to the order of $O(bML\log_2 N + |\mathcal{M}_1|M)$ and $O(kL\log_2 N + M)$, respectively. It is noteworthy that the difference in time complexity between the two approaches stems from different modes of network operation (i.e., centralized control or distributed control) and different methodologies of resource allocation (i.e., partner allocation or partner selection). Consider a well-known Hungarian approach. Our subcarrier allocation problem can be viewed as a transportation problem [9]. By using the Hungarian approach, the complexity of subcarrier allocation is on the order of $O(MN^2L^2)$. By the same token, the complexity of partner allocation is on the order of $O(|\mathcal{M}_1|M^2)$. Therefore, the time complexity of a Hungarian approach is $O(|\mathcal{M}_1|M^2 + MN^2L^2)$. Despite a plausibly improved system performance, applying a Hungarian approach can be inefficient in the WMNs without powerful centralized controllers. In contrast, our proposed approaches are of low complexity and more suitable for the WMNs with affordable off-the-shelf mesh nodes.

5.6 Efficiency Evaluation by Game Theory

In this section, we show that the partner allocation/selection solutions in our proposed node cooperative resource allocation approaches achieve efficient use of network resources. In game theory, efficient resource utilization is determined by the concept of Pareto optimality [88]. Modeled by a round-robin game, the partner selection/allocation solutions attain Nash equilibria (NE) [88].

5.6.1 Nash Equilibrium

Proposition 11 *Modeled by a round-robin game played by the RG nodes with fixed subcarrier allocation and power allocation, the proposed partner selection solution in the distributed resource allocation achieves an NE.*

Proposition 12 *Modeled by a round-robin game played by the potential partners with fixed subcarrier allocation and power allocation, the partner allocation solution from the proposed centralized approach achieves an NE.*

An NE in the WMNs with distributed control is vital in providing network stability. Signalling among nodes can be kept to a minimum, for there is no further message exchange after resource allocation is completed. It is noteworthy that, in the centralized approach, if the round-robin game is played by the source nodes instead of the potential partners, in general, an NE cannot be achieved. The rationale is that partners are *allocated* rather than *selected* in the centralized approach. Some RG node might have the tendency to change its partner allocation solution so as to increase its achieved data rate and hence its own utility. All in all, with centralized control, the round-robin game can be administered by a clusterhead and hence played by the potential partners, whereby an NE can be attained.

5.6.2 Pareto Optimality

Proposition 13 *With fixed subcarrier allocation and power allocation, the partner selection solution from the proposed distributed approach is Pareto optimal.*

Proposition 14 *With fixed subcarrier allocation and power allocation, the partner allocation solution from the proposed centralized resource allocation is Pareto optimal.*

By Propositions 13 and 14, the partner allocation/selection solutions obtained from the proposed approaches achieve the Pareto optimality, and hence the resources are efficiently utilized.

5.7 Discussion

Via the virtue of node cooperation, system performance can be enhanced. Nevertheless, applying the proposed QoS-driven node cooperative resource allocation approaches will plausibly lead to a conservative CAC mechanism. Since the performance gain due to node cooperation is not taken into account in the Phase-1 resource allocation in the centralized approach or the contention phase in the distributed approach, the CAC capacity region of our system is smaller than anticipated (i.e., the number of RG traffic flows that can be admitted in the system is smaller than it should be). However, as discussed in Section 5.3, non-altruistic node cooperation is not always beneficial to the system because the rate achieved by node cooperative communications can be smaller than direct transmissions or mesh nodes can be unavailable at times. Despite the shrinkage of the CAC capacity region, without knowing whether or not node cooperation is helpful *a priori*, a conservative approach is a safe yet effective way to guarantee the QoS demands of the admitted traffic flows. Thus, the proposed resource allocation approaches with node cooperation can be viable candidates to be applied to QoS-aware WMNs with service differentiation.

It is worth mentioning that we perform optimal power allocation via water-filling on the allocated subcarriers for each node twice in the centralized approach (i.e., in Phase 2 and in Phase 4). In light of decoupling the NCRAOP, the Phase-2 resource allocation solutions have a direct impact on the feasible region of the Phase-3 resource allocation problem. Therefore, the rationale of performing optimal power allocation in Phase 2 is to procure the best possible feasible region for the partner allocation problem in Phase 3. Providing the solutions for subcarrier allocation and partner allocation, we perform the second optimal power allocation in Phase 4 to further optimize the system performance. Notice that optimal power allocation is not necessary in the case where the transmit power of a mesh node is very large, as the channel variations across the (allocated) subcarriers become insignificant. In other words, the transmit power on each allocated subcarrier is more or less tantamount to the *water level*

of interest. For the WMNs consisting of mesh nodes with very large transmit power, the proposed QoS-driven centralized resource allocation approach with node cooperation can be reduced to Phase 1 and Phase 3 only. Nevertheless, the proposed four-phase centralized resource allocation approach is imperative for our system model where mesh nodes only have limited transmit power. On a different note, our proposed centralized approach allocates subcarriers, transmit power, and partners in succession, leading to a compromised system performance. In order to achieve (near-)optimal solutions, some common approaches such as exhaustive search and GAs are usually required. In spite of jointly considering all the resources simultaneously, these approaches do not scale well due to their high computational costs (see Chapter 4). Besides, performing online computations in a timely fashion is hardly feasible even in small-scale WMNs without any powerful central controller. Therefore, applying such computationally expensive approaches directly to our specific system model can be ineffective in enhancing system performance or providing QoS assurance. On the other hand, our proposed centralized approach is of low complexity and a good fit for our system model under consideration in this research.

Concerning our proposed distributed node cooperative resource allocation approach, potential partners engage in node cooperation independently so long as they can further improve the system performance. In essence, a mesh node has no a priori knowledge of whether and when a potential partner will help relay its information. As a result, as opposed to the partner *allocation*, applying the notion of partner *selection* is more suitable to the WMNs with distributed control at the cost of performance degradation. In addition, since the node cooperation of interest is not symmetric, we refer to this type of node cooperation as *relay-initiated* node cooperation. With no central controller, the underlying assumption for our relay-initiated node cooperation is that every mesh node is cooperative rather than selfish. Even though the partner selection solutions obtained from our proposed distributed approach achieve an NE (see Proposition 11), selfish nodes might not participate in node cooperative

transmissions in practice, thereby capping the system performance. Therefore, without any proper coordination, the merit of node cooperation cannot be exploited. Devising an effective and efficient MAC mechanism with incentives and/or rebates for WMNs with selfish nodes is vital; however, addressing this issue is beyond the scope of this research.

To cope with varying traffic load demands in the WMNs, we not only need to update the resource allocation solutions from time to time, but we also need to adjust the number of admitted traffic flows governed by CAC at times. In our specific system model, we can envision that executing a CAC policy a couple of times a day would be sufficient. For example, in typical suburban/rural residential areas, traffic patterns are expected to be altered in the morning when people get to work, in the evening when people surf the Internet and watch online streaming videos, and in the midnight when people run peer-to-peer file sharing applications before going to bed [3,10,45]. In order for the WMNs to function efficiently, our proposed node cooperative resource allocation needs to be adaptive, catering to the variations of traffic loads. The impact of poor traffic estimates on the system performance is to be evaluated in Section 5.8.

5.8 Performance Evaluation

5.8.1 Simulation Environment

Consider a cluster with a number of wireless mesh nodes randomly located in a 1km x 1km coverage area. We adopt the path-loss model suggested in [44] (i.e., hilly/moderate-to-heavy tree density). Consider an OFDM-based wireless environment with N available subcarriers. The maximum transmission rate over each subcarrier is considered to be 200kb/s. We assume that the routing is predetermined so that the transmission source and destination pair of an incoming packet is known in advance. Other simulation parameters are chosen as follows:

$P_m^{\max} = 1\text{W}$, $\sigma_n^2 = 10^{-10}\text{W}$, $N = 100$, $L = 4$, $\rho = 1/3$, $\varphi = 1$, and $\mathbf{\Theta} = 1$. Since our resource allocation solutions sustain for an active resource allocation interval, the duration of the first frame (with 1 beacon slot, 1 control slot and L DATA slots) is $5(L + 2)$ ms and that of the subsequent frames (with L DATA slots) is $5L$ ms in the proposed centralized approach, with a 5ms DATA slot. We consider that the polling is done in every T ms, meaning that the node cooperative resource allocation solution is updated every T ms. Here, both the polling and the beacon packet transmissions are assumed to be error-free. For fair comparison, we consider that the duration of an active interval in the proposed distributed approach is T ms, where the first 10ms is dedicated to the contention phase while the remaining time period is dedicated to the transmission phase. We perform the simulations for 10,000 runs and average the results, where each simulation run sustains 5,000 frames.

Concerning the traffic models, the RG traffic is generated according to a two-state ON-OFF model. In the ON state, a fixed-size packet arrives in every 20ms with rate demand 384kb/s, whereas in the OFF state, no packet is generated. We consider that the duration of an ON period and that of an OFF period are independent, both follow an exponential distribution, where the mean ON period and the mean OFF period are 1s and 1.2s, respectively. The delay bound of RG traffic is assumed to be $5L$ ms. The required packet dropping rate is less than 1%. On the other hand, the BE traffic does not have any QoS requirements. BE packet arrivals follow a Poisson process with mean rate λ packets/second, where the packet size follows a Weibull distribution (i.e., Weibull(2,2)). To mimic the mixed traffic in a WMN, we assume that an RG node has one RG traffic flow and one BE flow, while a BE node has one BE flow. Regarding packet-level QoS provisioning, we conceive that the priority of RG traffic (packets) is related to the performance of their packet dropping rates. In the centralized approach, after gathering the transmission requests in the control slot, a clusterhead grants the requests of transmitting higher-priority packets first, facilitating QoS provisioning. In the distributed approach, we consider that the RG node with the highest

packet dropping rate transmits its BB signal earlier than the other nodes in the contention phase. If two or more RG nodes have the same packet dropping rate, the node with the highest minimum required rate will win the contention of interest. In other words, our QoS provisioning strategy for the RG traffic hinges upon the packet dropping rate. To further augment the effectiveness of QoS provisioning, the partners, timeslots, and subcarriers allocated to or chosen by a particular mesh node are reserved for packet transmissions until the next active resource allocation interval (e.g., next polling). The performance measurements used in our simulations are 1) throughout, 2) resource utilization, 3) packet dropping rate, and 4) node cooperation gain (NCG), where the NCG is defined the normalized throughput gain due to beneficial non-altruistic node cooperation.

5.8.2 Simulation Results

We evaluate the system performance of the proposed node cooperative resource allocation approaches versus M, T, $|\mathcal{M}_1|$, and λ in terms of throughput, resource utilization, packet dropping rate, and node cooperation gain. The standard deviations of the results are also plotted for reference. For performance comparison, we consider two baseline approaches and an approach suggested in [121]: 1) a centralized baseline approach which is the same as the proposed centralized approach without Phase-3 and Phase-4 resource allocation; 2) a distributed baseline approach which is the same as the proposed distributed approach without considering node cooperation in the transmission phase; and 3) the Zhang's approach proposed in [121] which first allocates the subcarriers with no QoS consideration to maximize throughput and then re-allocates the subcarriers trying to satisfy the QoS demands of the nodes without considering node cooperation. Note that the performance degradation due to signalling overhead is not taken into account in evaluating the system performance. An

upper bound[2] for average throughput performance is also plotted for reference.

Effect of the number of mesh nodes, M

For $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, $T = 150$ms, and $\lambda = 50$ packets/second, Figure 5.7 shows the throughput performance versus the number of mesh nodes. We can see that the throughputs of all considered approaches increase with M, since the system is not saturated. As expected, the centralized approaches (i.e., the proposed four-phase centralized approach and the centralized baseline approach) outperform their distributed counterparts (i.e., the proposed two-phase distributed approach and the distributed baseline approach) due to the merit of the existence of a clusterhead. However, the Zhang's approach achieves the highest throughput among all considered approaches, realizing the goal of throughput maximization. On a different note, our proposed four-phase centralized (two-phase distributed) approach outperforms the baseline centralized (distributed) approach, which stems from an additional performance gain due to beneficial node cooperation. Due to beneficial node cooperation, the proposed two-phase distributed approach achieves similar throughput performances as the centralized baseline approach. The node cooperation gain (NCG) is given in Table 5.2. As anticipated, the more the mesh nodes, the more the potential helpers and hence the higher the NCGs. In general, the NCG obtained in our proposed centralized approach is higher than that obtained in our proposed distributed approach. The rationale is mainly due to the partner *allocation* employed in our proposed centralized approach yet the partner *selection* in our proposed distributed approach. Another reason is that node cooperation in the proposed distributed approach can only be triggered when some mesh nodes are idle in the first minislot, thereby curbing some potential and favorable node cooperation opportunities. The NCG in our proposed centralized approach is roughly leveled off from $M = 30$ onward, which is ascribed

[2]We analytically obtain an upper bound for average throughput performance under the assumptions of no packet dropping for the RG traffic and perfect statistical traffic multiplexing.

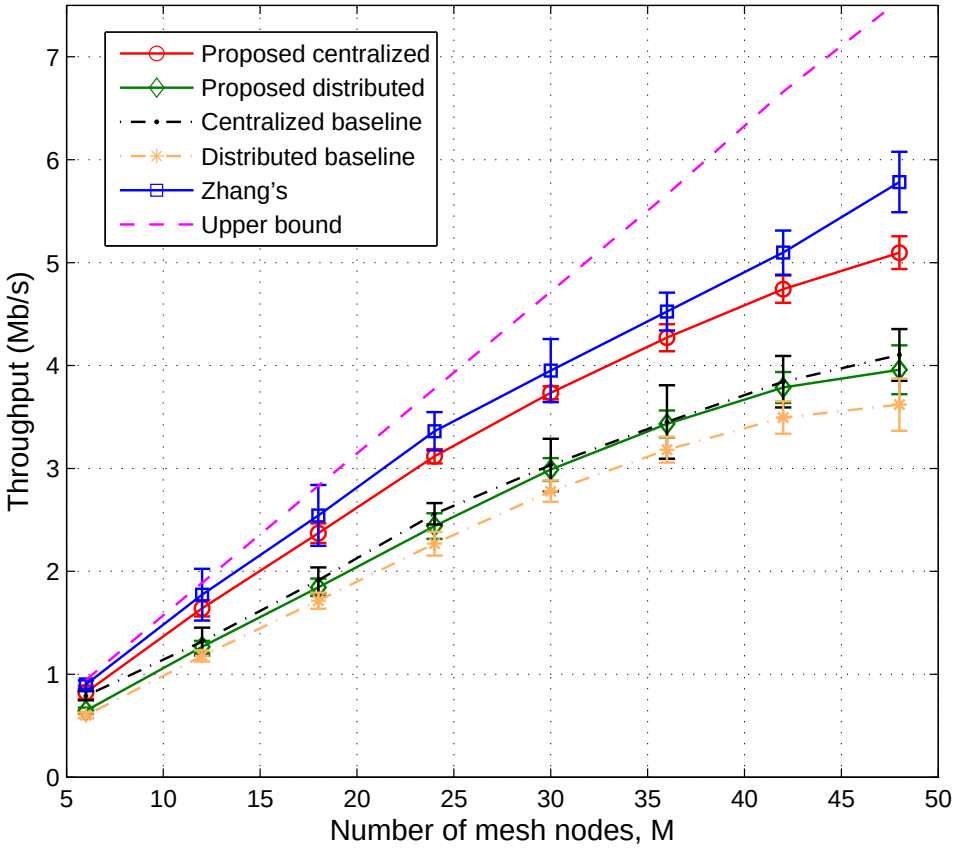


Figure 5.7: Comparison of the throughput performance of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the number of mesh nodes (with $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, $T = 150$ms, and $\lambda = 50$ packets/second).

Table 5.2: Relationship between the number of mesh nodes, M, and the node cooperation gain (NCG) (i.e., normalized throughput gain due to node cooperation) for the proposed approaches (with $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, $T = 150$ms, and $\lambda = 50$ packets/second)

M	6	12	18	24	30	36	42	48
NCG for the proposed centralized approach (in %)	4.29	10.11	16.92	21.81	23.14	23.74	23.44	24.21
NCG for the proposed distributed approach (in %)	7.47	7.48	7.50	7.54	7.59	7.89	8.36	9.35

to the limited available resources (i.e., subcarriers). We expect that the NCG can be higher with a larger value of N. We observe that, as the number of mesh nodes increases from 6 to 48, the resource utilization of the proposed four-phase centralized approach increases from 9% to 41%, that of the proposed two-phase distributed approach from 7% to 30%, that of the centralized baseline approach from 8% to 31%, that of the distributed baseline approach from 6% to 28%, and that of the Zhang's approach from 10% to 44%, respectively. The low resource utilization is due to low traffic load and resource reservation. We observe that the resource utilization (and throughput) for our proposed approaches can be improved when the traffic load increases. In Figure 5.8, the RG packet dropping rates are depicted. The packet dropping rates for RG traffic in our proposed approaches and two baseline approaches are well below 1% due to effective packet-level QoS provisioning. On the other hand, the RG packet dropping rate of the Zhang's approach increases and reaches 20% as M increases. Figure 5.8 shows that the Zhang's approach is ineffective in supporting the QoS requirements of RG traffic at the packet level. Nonetheless, the Zhang's approach aims at maximizing the (network-wise) throughput in lieu of focusing on (node-wise) QoS satisfaction. The results also assure the fact that provisioning QoS and increasing throughput are conflicting performance measures [22].

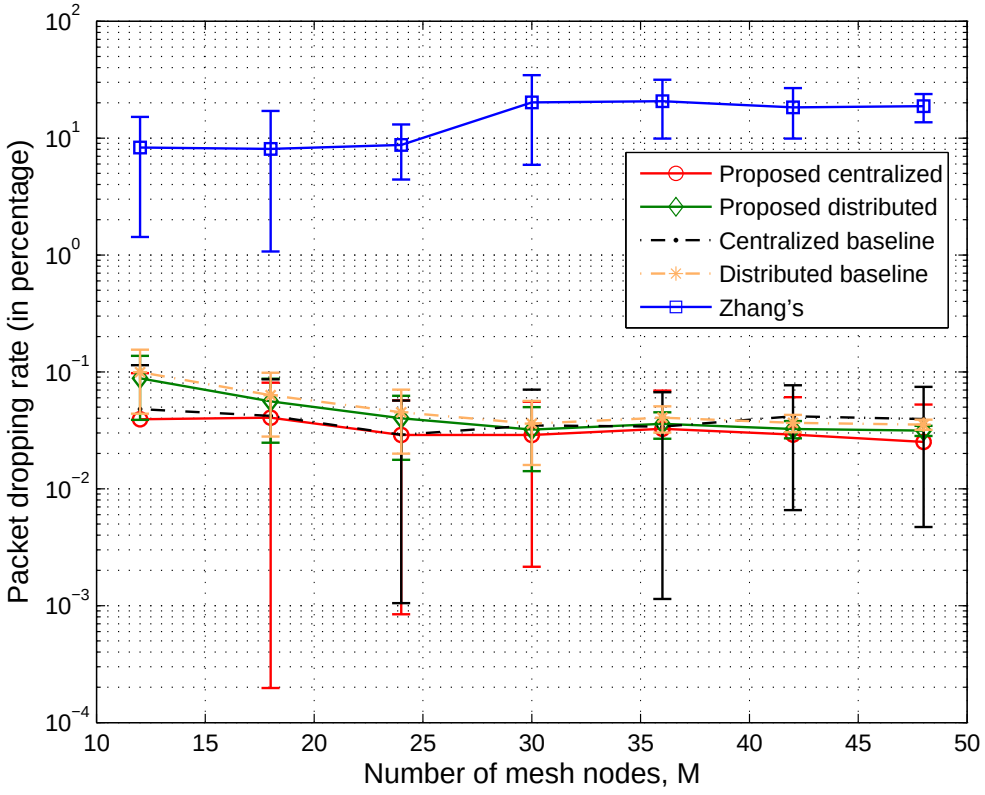


Figure 5.8: Comparison of the RG packet dropping rates of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the number of mesh nodes (with $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, $T = 150$ms, and $\lambda = 50$ packets/second).

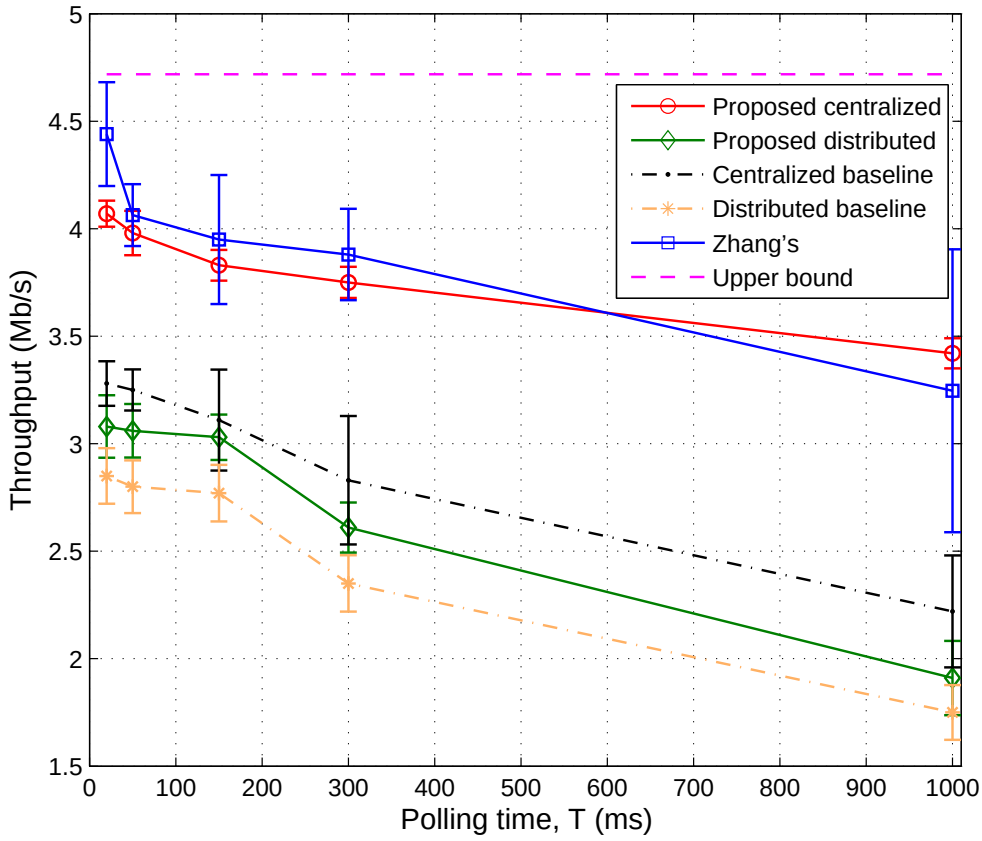


Figure 5.9: Comparison of the throughput performance of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the polling time (with $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $\lambda = 50$ packets/second).

Table 5.3: Relationship between the value of T and the node cooperation gain (NCG) (with $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $\lambda = 50$ packets/second)

T (in ms)	20	50	150	300	1000
NCG for the proposed centralized approach (in %)	21.13	22.46	23.15	32.50	54.05
NCG for the proposed distributed approach (in %)	5.48	6.99	7.58	11.06	13.71

Effect of the polling time, T

For $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $\lambda = 50$ packets/second, we study the impact of the polling time (i.e., the length of an active resource allocation interval) on the system performance measures. Figure 5.9 shows the throughput performance versus the value of T. The throughputs obtained in all considered approaches decrease with the value of T. The larger the value of T, the less accurate the traffic load estimate and hence the weaker the throughput performance. In particular, even though our proposed distributed approach maintains its NCG against T (see Table 5.3), its throughput drops significantly from 3.0Mb/s to 1.9Mb/s. On the contrary, with the help of a clusterhead, not only does our proposed centralized approach effectively sustain its throughput performance, but its NCG also ramps up considerably from 21% to 54% when T increases (see Table 5.3). As a result, the proposed four-phase centralized approach is shown to be less vulnerable to poor traffic load estimates. As to the Zhang's approach, however, its throughput obtained plummets sharply as T increases. The decline is due to the absence of effective packet-level QoS provisioning mechanism in place, and once the traffic load estimates are less accurate, the system performance deteriorates dramatically. This phenomenon can also be explained in Figure 5.10. Similar to Figure 5.8, the RG packet dropping rates of our proposed approaches and two baseline approaches are capped by 1%. In contrast, the packet dropping rate of the Zhang's approach increases from 10% to 43%, resulting in the worst RG packet dropping rate performance. In short, resource allocation approaches without considering packet-level QoS

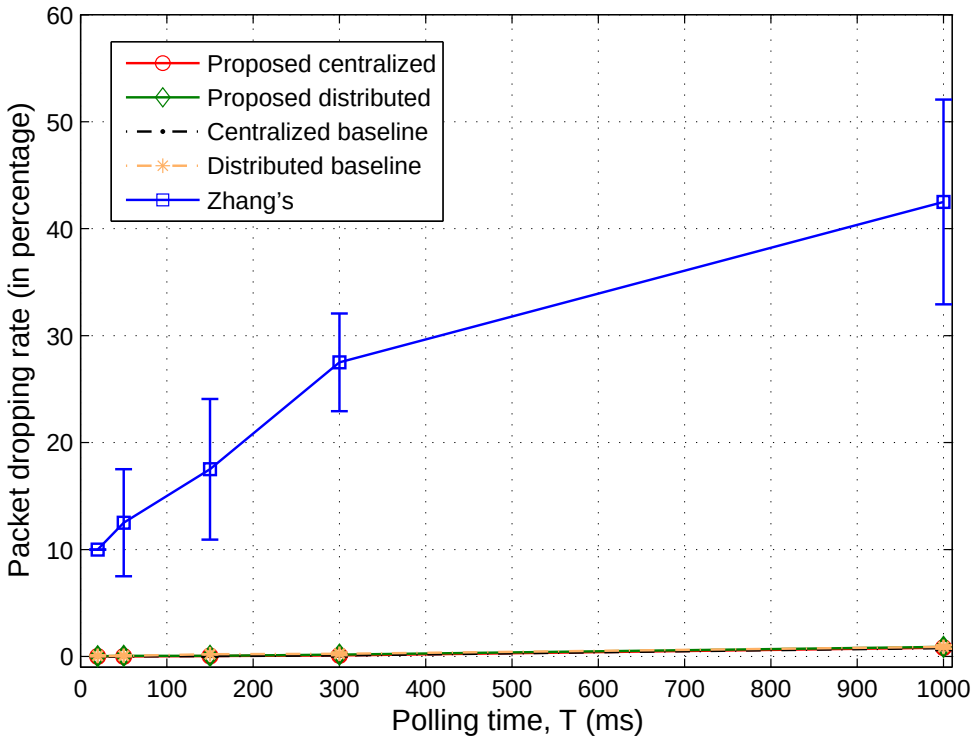


Figure 5.10: Comparison of the RG packet dropping rates of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the polling time (with $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $\lambda = 50$ packets/second).

provisioning are susceptible to the accuracy of traffic load estimates. For resource utilization, we observe that the trends of all curves are similar to those in Figure 5.9.

On a different note, we expect to obtain a poorer system performance if channel conditions change faster over time because the channel condition estimates become less accurate (e.g., Example (c) in Figure 5.1). To alleviate this problem, polling should be done more frequently in the proposed centralized approach, whereas a transmission phase should be shorter in the proposed distributed approach, whereby the estimates of the channel conditions (and the

traffic load) can be more accurate. One pitfall, however, is the additional MAC overhead introduced by extra beacon/control slots and contention phases in our centralized approach and distributed approach, respectively. Thus, the system performance of WMNs highly hinges on the value of T; how to acquire the optimum value of T (i.e., optimal MAC) is crucial but beyond the scope of this research study.

Effect of RG traffic, $|\mathcal{M}_1|$

For $M = 30$ mesh nodes, $T = 150$ms, and $\lambda = 50$ packets/second, we study the impact of RG traffic (i.e., the value of $|\mathcal{M}_1|$) on the system performance measures. In Figure 5.11, the throughput performance versus the value of $|\mathcal{M}_1|$ is depicted. Since the network is not saturated, in general, all the curves go up with the number of RG nodes, $|\mathcal{M}_1|$. From $|\mathcal{M}_1| = 20$ onward, the throughputs of our proposed distributed approach and the distributed baseline approach begin to level off, resulting from the effect of resource reservation for the RG traffic. By the same token, the rates of the throughput increment in our proposed centralized approach and the centralized baseline approach decrease from $|\mathcal{M}_1| = 25$ onward. The throughput performance of the Zhang's approach first rises from $|\mathcal{M}_1| = 10$ to 20 and then declines afterwards. Similar to the previous discussions, as the number of RG nodes increases, the Zhang's approach fails to effectively provision packet-level QoS, thereby increasing its RG packet dropping rate and decreasing its throughput. At $|\mathcal{M}_1| = 30$, the throughput obtained by the Zhang's approach is even lower than that by the proposed four-phase centralized approach. The resource utilizations attained by all the approaches have the trends similar to those in Figure 5.11. The NCGs of our two proposed approaches against $|\mathcal{M}_1|$ are given in Table 5.4. It can be noted that the values of the NCGs stay more or less the same against $|\mathcal{M}_1|$. It shows that the node cooperation opportunities of both proposed approaches are almost independent of the number of RG nodes (traffic flows) in the system. It is noteworthy that, in Figure 5.11, there is an obvious performance gap between the throughputs obtained

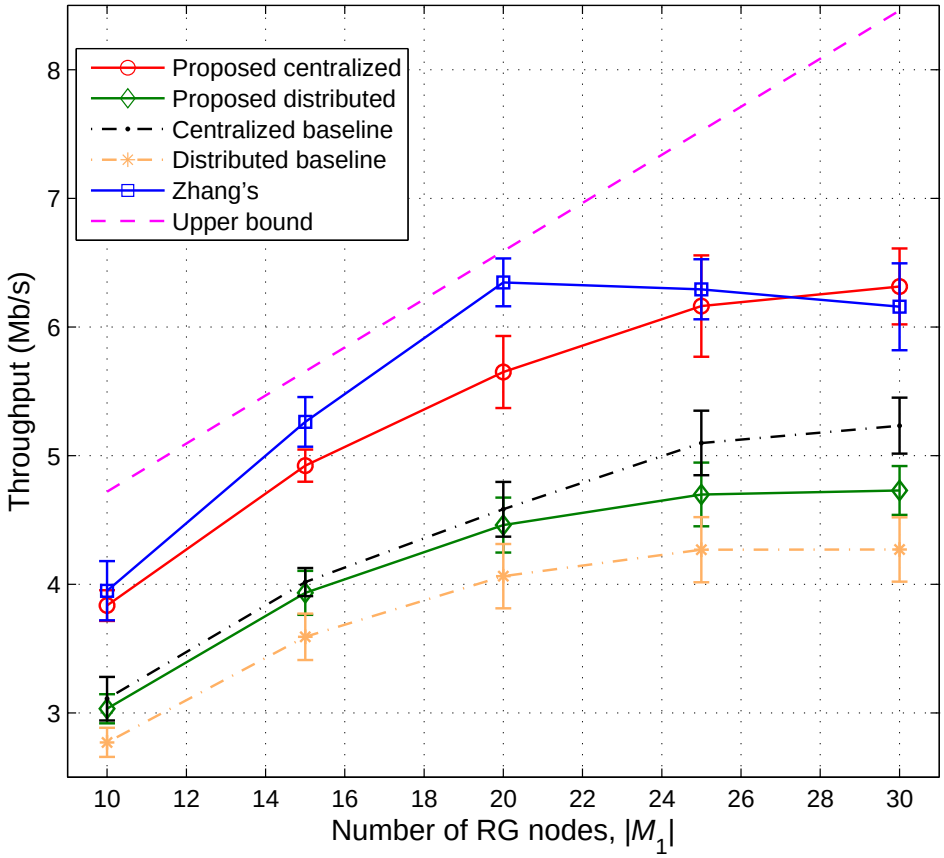


Figure 5.11: Comparison of the throughput performance of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the number of RG nodes (with $M = 30$ mesh nodes, $T = 150$ms, and $\lambda = 50$ packets/second).

Table 5.4: Relationship between the number of RG nodes, $|\mathcal{M}_1|$, and the node cooperation gain (NCG) (where $M = 30$ mesh nodes, $T = 150$ms, and $\lambda = 50$ packets/second)

$\lvert\mathcal{M}_1\rvert$	10	15	20	25	30
NCG for our proposed centralized approach (in %)	23.15	22.69	23.36	20.78	20.65
NCG for our proposed distributed approach (in %)	7.58	7.24	7.64	7.73	8.63

from the considered approaches and the upper bound at a large $|\mathcal{M}_1|$. This gap is mainly ascribed to the bandwidth reservation for RG traffic. Concerning the RG packet dropping rates, the trends of the considered approaches are similar to the ones shown in Figure 5.10. In a nutshell, with the virtue of node cooperation, our four-phase centralized approach can achieve better system performance and provide more effective QoS provisioning than its throughput-oriented counterpart when there is a large number of RG nodes in the network.

Effect of BE traffic, λ

For $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $T = 150$ms, we study the impact of BE traffic (i.e., the value of λ) on the system performance measures. Here, we consider two cases: 1) $\lambda = 50$ packets/second (i.e., bursty data traffic); and 2) $\lambda \to \infty$ packets/second (i.e., background data traffic). Figure 5.12 shows the throughput performance for $\lambda = 50$ and $\lambda \to \infty$. It is clear that the throughputs obtained in the presence of background data traffic (i.e., $\lambda \to \infty$) are higher than in the case of Poisson arrivals (i.e., $\lambda = 50$). The NCGs of the proposed four-phase centralized approach and the proposed two-phase distributed approach are given in Table 5.5. As seen, the NCGs obtained drop from 23.14% to 9.01% in the proposed centralized approach and from 7.59% to 0% in the proposed distributed approach, respectively, when the data traffic changes from bursty (i.e., $\lambda = 50$) to background (i.e., $\lambda \to \infty$). Concerning our proposed centralized approach, since the rate function is a concave function and the throughput obtained by Phase-1 and Phase-2 resource allocation

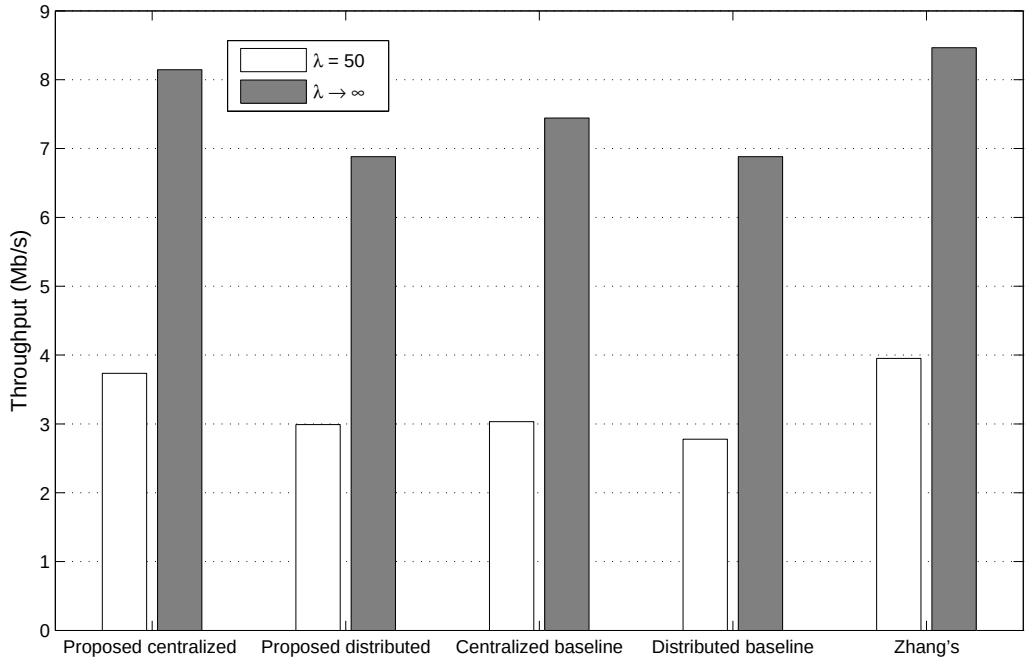


Figure 5.12: Comparison of the throughput performance of the proposed four-phase centralized approach, the proposed two-phase distributed approach, the centralized baseline approach, the distributed baseline approach, and the Zhang's approach vs. the value of λ (with $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $T = 150$ms).

Table 5.5: Relationship between the value of λ and the node cooperation gain (NCG) (where $M = 30$ mesh nodes, $|\mathcal{M}_2| = 2|\mathcal{M}_1|$, and $T = 150$ms)

λ	50	$\rightarrow \infty$
NCG for our proposed centralized approach (in %)	23.14	9.01
NCG for our proposed distributed approach (in %)	7.59	0

is already high in the case of $\lambda \rightarrow \infty$, the room for further throughput increment due to node cooperation is relatively smaller than in the case of $\lambda = 50$. As a result, the performance gain of the proposed four-phase centralized approach over the centralized baseline approach is less substantial. On the other hand, it is notable that the NCG of our proposed distributed approach vanishes in the case of $\lambda \rightarrow \infty$, the rationale for which is that all the mesh nodes are busy all the time and no mesh nodes are idle in the first minislot in the transmission phase. Thus, there is no partner available, wiping out the effectiveness of node cooperation. In other words, if a WMN with decentralized control is saturated (with background data traffic), the benefits of non-altruistic node cooperation cannot be exploited. In a nutshell, when designing and deploying an efficient and effective WMN in practice, we should take notice of the nature of node cooperation (i.e., non-altruistic or altruistic), the traffic pattern (i.e., bursty traffic or background traffic), and the mode of network operation (i.e., centralized control or distributed control). As regards the packet dropping rate of the RG traffic, we observe that the results are nearly the same as the ones shown in Figure 5.8. All in all, the BE packets are assigned the lower priority and hence the change in the value of λ has almost no influence on the packet dropping rate performance of RG traffic.

5.9 Chapter Summary

In this Chapter, we propose two non-altruistic node cooperative resource allocation approaches tailored for WMNs with QoS support and service differentiation. Both the pro-

posed four-phase centralized approach and the proposed two-phase distributed approach are shown to be effective in QoS provisioning for RG traffic and system performance melioration. Simulation results demonstrate that our proposed approaches achieve satisfactory system performance and are less susceptible to the inaccuracy of traffic load estimates. Our study reveals a crucial design principle that whether or not node cooperation is useful depends upon the nature of node cooperation, the traffic pattern, and the mode of network operation. Both two proposed node cooperative resource allocation approaches are Pareto optimal, utilizing meager network resources efficiently.

Chapter 6

Performance Enhancement II: Instinctive Channel Sensing for Cognitive Radio

6.1 Introduction

Since many frequency bands in the radio spectrum are underutilized most of the time, shown in recent studies [103], efficient resource utilization depends upon how effective secondary users can utilize the temporarily available spectrum bands. As such, effective and efficient channel exploitation is imperative for secondary users to seize available network resources as efficiently and effectively as possible in cognitive radio networks (CRNs). On the other hand, in typical multi-channel systems (e.g., OFDM-based networks), due to distinct fading characteristics, a (secondary) user can experience different fading channel gains across different frequency channels. In a multi-user environment, different users can experience different channel conditions over the same channel(s). This phenomenon gives rise to the notion of *multi-user diversity* [106]. With effective channel allocation, system performance can be

improved, and fine-grain QoS can be guaranteed [24]. In the context of cognitive radio, the difference in the channel gains and hence the maximum achievable rates of the channels can help devise a more effective channel sensing strategy for secondary users, thereby plausibly increasing resource utilization. In this Chapter, we study the problem of channel exploitation for CRNs by taking the intrinsic nature of fading channels into consideration. The contributions and significance of this research work are three-fold [19].

- First, we propose a simple channel sensing order for secondary users in CRNs, referred to as instinctive channel sensing. By sensing the channels according to the descending order of their achievable rates, we can prove the optimality of our proposed rule that a secondary user should stop at the first sensed free channel. Provided that all the channel availabilities are equal, we can further prove that the proposed instinctive channel sensing is optimal in terms of throughput.

- Second, in a multi-secondary user scenario where our instinctive channel sensing is employed, we analytically derive an expression for the probability of packet transmission collision, which is a function of primary-free probabilities, the number of channels, and the number of secondary users. Our results show that the probability of collision decreases as the number of channels increases, the number of secondary users decreases, and/or the values of primary-free probabilities decrease. Our results also demonstrate that, as the number of secondary users increases, both the total throughput and resource utilization rise due to increased transmission opportunities and multi-user diversity. In addition, it is observed that resource utilization can be further improved when the number of secondary users is close to the number of channels.

- Third, we compare our proposed channel exploitation approach with two other approaches, namely a random approach and the approach suggested in [51]. Simulation results show that our proposed instinctive sensing achieves the best reward (i.e.,

Table 6.1: Summary of important symbols used in this Chapter.

Symbol	Definition
s_k	the k^{th} channel to be sensed in a sensing sequence
c_k	effectiveness of data transmissions at the k^{th} sensed channel
θ_{s_k}	primary-free probability of the k^{th} sensed channel
R_{s_k}	achievable transmission rate of the k^{th} sensed channel
λ_k	instantaneous reward for a secondary user at the k^{th} sensed channel
Λ_{k+1}	expected reward for a secondary user if skipping the k^{th} sensed channel
τ	time spent on channel sensing
T	duration of a timeslot
N	number of available channels in the CRN
M	number of secondary users in the CRN
P_c	probability of collision

throughput) performance, outperforming its counterparts. We also show that our sensing approach can preserve its performance superiority over the other two approaches no matter what the level of resource utilization in a CRN is (i.e., good or poor resource utilization by primary users).

The remainder of this Chapter is organized as follows. Related work is given in Section 6.2. The cognitive radio network model is described in Section 6.3. The proposed channel sensing with optimal stopping are presented in Section 6.4. Performance evaluation is given in Section 6.5. A brief discussion is provided in Section 6.6. Finally, we summarize this Chapter in Section 6.7. A summary of important symbols used in this Chapter is given in Table 6.1 for easy reference.

6.2 Related Work

In the literature, the topic of channel exploitation has drawn a plethora of attention [50,51, 95,123]. In [51], a channel sensing order based on the channel availabilities (or primary-free

probabilities) is proposed. The proposed sensing approach performs well only if the channel availabilities are known to the secondary users. Without having channel availabilities in advance, the channel sensing scheme is not optimal in terms of throughput. An optimal stopping rule is suggested in [95] to achieve satisfactory gains in throughput in a conventional (non-cognitive) wireless network. In [123], a Markov-based model is proposed to characterize the channel availabilities of primary users. With complete knowledge of channel availabilities, secondary users can opportunistically select and exploit the channels. In a system with a large number of primary users, however, the availability of each channel is hardly predictable, especially when traffic arrivals are bursty and resource reservation is considered for primary users [23]. In [50], channel exploitation is studied from an MAC-layer perspective, where a decentralized cognitive MAC protocol is proposed. Despite optimal stopping in place, the intrinsic features of channel fading are not exploited, thereby leading to suboptimal system performance. In this work, we devise a simple channel sensing order, by exploiting the benefits of the intrinsic features of physical fading environments.

6.3 Cognitive Radio Network Model

Consider a synchronized CRN with M non-mobile secondary users and N available channels. We assume that a secondary user and its target stationary destination are known. Time is partitioned into slots of duration T. In each timeslot, each channel is either available (i.e., no primary activities) or busy (i.e., with primary activities). The availability of a channel in each timeslot is independent of the status of other channels, and the availability of a channel in a timeslot is independent of that in other timeslots. In each timeslot, a secondary user senses the channels according to its sensing sequence. Denote $(s_1, s_2, ..., s_N)$ as a sensing sequence. Note that a sensing sequence is a permutation of the set $(1, 2, ..., N)$. Denote τ as the time needed for sensing a channel, where $N\tau < T$. Here, we assume that the sensing is

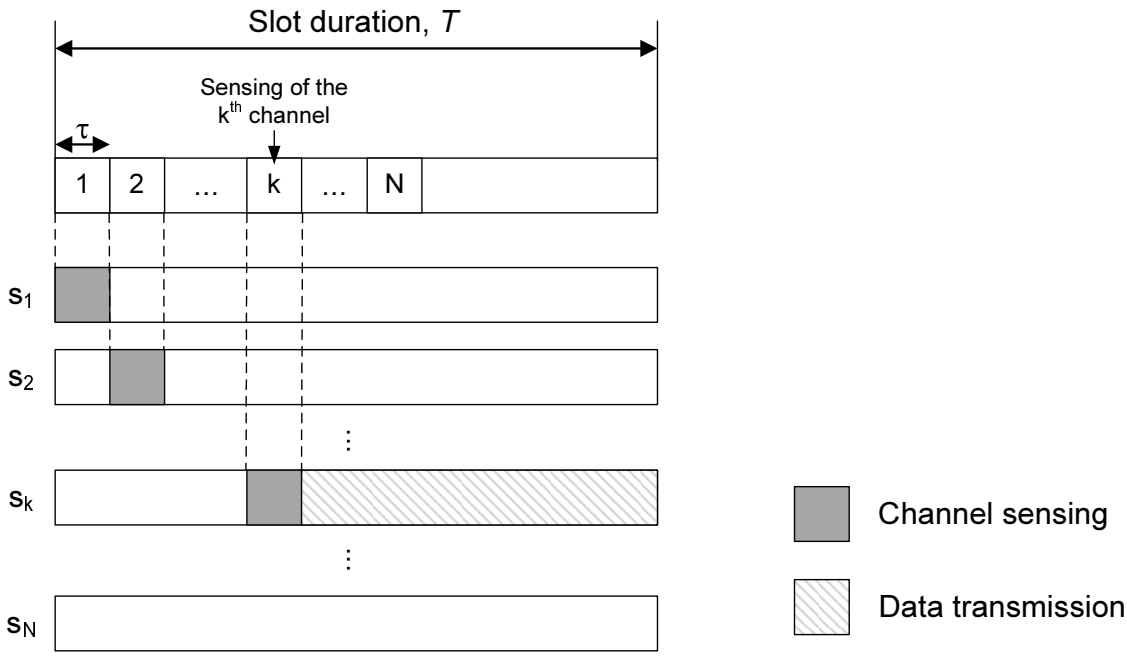


Figure 6.1: The slot structure and the channel sensing procedure for a secondary user.

perfect, and τ is sufficient to determine whether or not a channel is free of primary activities. Figure 6.1 depicts the slot structure and channel sensing under consideration, with negligible channel switching time. In Figure 6.1, a secondary user senses the first $(k-1)$ channels busy and stops at the k^{th} sensed channel, which is the first free channel. Since the optimal stopping criterion is satisfied at the k^{th} sensed channel, to be discussed in Section 6.4, this secondary user transmits its information over the k^{th} sensed channel in the reminder of that timeslot. The period of time spent on channel sensing and that on data transmissions are $k\tau$ and $T-k\tau$, respectively. Notice that we consider a constant slot time with actual data transmission time changing with k as in [95]. Denote c_k as the effectiveness of data transmissions if a secondary

user stops at the k^{th} sensed channel, which is given by

$$c_k = 1 - \frac{k\tau}{T}. \tag{6.1}$$

With non-mobile transmission pairs, the channel gains between source nodes and destination nodes can be estimated accurately. On the other hand, a secondary user can transmit only when there is no primary activity (i.e., no concurrent transmissions). As such, the achievable transmission rate of a secondary user over each channel can be determined in advance. We consider that each secondary user is equipped with one transceiver and, therefore, it can only sense channels one at a time. A secondary user needs only one channel to transmit its data.

6.4 Proposed Stopping Rule and Channel Exploitation

In CRNs, the gist of channel exploitation is to find and utilize a desired unoccupied channel for secondary users as effective and efficient as possible. In fact, the problem of channel exploitation can be viewed as a general stopping rule problem, the objective of which is to stop at some channel that maximizes the (expected) reward for a secondary user. In our work, since the number of available channels is finite, we can formulate our channel exploitation problem as a finite-horizon stopping problem [32]. In the following subsections, we first consider the stopping rule and channel exploitation for one secondary user. Then, we extend our investigation to a two-secondary user case and a multi-secondary user case in Section 6.4.2 and in Section 6.4.3, respectively.

6.4.1 Single-Secondary User Scenario ($M = 1$)

In this scenario, an active secondary user can transmit its own information as long as any of the available channels is sensed free. Since the transmission rates of all the channels

are known *a priori*, it is proposed that the secondary user senses the channels according to the descending order of their achievable rates, referred to as *instinctive sensing*. Therefore, for $i < j$, $R_{s_i} \geq R_{s_j}$, $\forall i, j$, where R_{s_k} is the achievable transmission rate of the k^{th} sensed channel. If the ${s_k}^{th}$ channel is sensed free, a secondary user transmits its data at the rate of R_{s_k}; otherwise, proceeds to sense the ${s_{k+1}}^{th}$ channel. To characterize the dynamics of our instinctive channel sensing, we employ the notion of rewards in the context of stopping rule. Denote λ_k as the instantaneous reward for a secondary user at the k^{th} sensed channel. The instantaneous reward at the k^{th} sensed channel for $1 \leq k < N$ can be written as

$$\lambda_k = \begin{cases} c_k R_{s_k}, & \text{sensed free with probability } \theta_{s_k} \\ \Lambda_{k+1}, & \text{sensed busy with probability } (1 - \theta_{s_k}) \end{cases} \tag{6.2}$$

and that for $k = N$ as

$$\lambda_N = \begin{cases} c_N R_{s_N}, & \text{sensed free with probability } \theta_{s_N} \\ 0, & \text{sensed busy with probability } (1 - \theta_{s_N}) \end{cases} \tag{6.3}$$

where θ_{s_k} ($\in [0, 1]$) is the *primary-free probability* of the k^{th} sensed channel and Λ_{k+1} is the expected reward if a secondary user skips the k^{th} sensed channel and uses one of the remaining $(N - k)$ channels. The expected reward Λ_k is given by

$$\Lambda_k = \mathbb{E}[\lambda_k] = \begin{cases} \theta_{s_k} c_k R_{s_k} + (1 - \theta_{s_k}) \Lambda_{k+1}, & 1 \leq k < N \\ \theta_{s_N} c_N R_{s_N}, & k = N \end{cases}. \tag{6.4}$$

Notice that $c_k R_{s_k}$ represents the effective throughput obtained if a secondary user transmits its data over the k^{th} sensed channel, and Λ_1 can be interpreted as the average throughput obtained by a secondary user, and can be re-written as

$$\Lambda_1 = \sum_{k=1}^{N} \left[\prod_{i=1}^{k} \left(1 - \theta_{s_{i-1}}\right) \right] \theta_{s_k} c_k R_{s_k} \tag{6.5}$$

where we define $\theta_{s_0} = 0$. Since our channel exploitation problem is a finite-horizon stopping problem, the expected rewards $\{\Lambda_1, \Lambda_2, ..., \Lambda_N\}$ acquired at all the channels can be obtained, and hence the stopping rule can be completely specified by $\{\Lambda_k\}_{k=1}^{N}$.

Proposed Stopping Rule – We propose that a secondary user should stop at the first sensed free channel. With the proposed instinctive sensing order, we can prove that a secondary user achieves the maximal (expected) reward by stopping at the first sensed free channel and hence, our proposed stopping rule is optimal.

Proposition 15 *(**Instinctive Sensing**) Provided that the channels are sensed according to the descending order of their achievable transmission rates, a secondary user can achieve the maximal reward by stopping at the first sensed free channel.*

According to Proposition 15, our proposed stopping rule is optimal. On the condition that all the θ_{s_k}'s are equal, we can further prove that our proposed instinctive channel sensing order is optimal in terms of throughput (i.e., expected reward Λ_1).

Proposition 16 *Provided that $\theta_{s_i} = \theta_{s_j}, \forall i, j$, the proposed instinctive sensing is throughput-optimal.*

Proposition 16 shows the optimality of our instinctive sensing order on the condition that the primary-free probabilities are all equal. In general, however, as $\{\theta_{s_k}\}_{k=1}^{N}$ may vary with k, our proposed sensing order is not always optimal in terms of the expected reward Λ_1 for a secondary user. Nevertheless, it is noteworthy that, in a system consisting of a large number of primary users, if there is any imbalance in the channel availabilities, more (less) primary users are likely to compete for the most (least) available channel(s), resembling the notions of max-min fairness and solidarity [22]. Therefore, the condition that the primary-free probabilities are equal can be realized in some cases, whereby the maximal throughput can be attained by the proposed instinctive sensing.

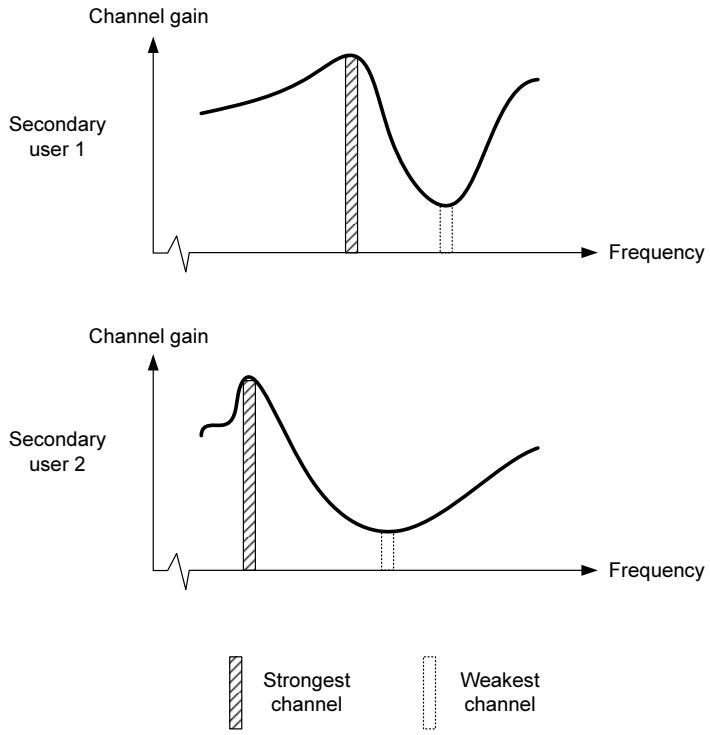


Figure 6.2: An illustration of the channel gains of two secondary users.

6.4.2 Two-Secondary User Scenario ($M = 2$)

Here, we extend our study to a two-secondary user scenario. We assume that the two (active) secondary users in a CRN are non-mobile and placed at different locations, and are 1-hop neighbors (i.e., no hidden terminal). Due to the random nature of multi-path propagation, the channel gains of the available channels for secondary user 1 are plausibly different from that for secondary user 2 (see Figure 6.2) [106]. As such, it is very unlikely for the two secondary users to have the same instinctive channel sensing sequence. In a CRN without any coordination between the secondary users, however, it is still possible that a particular channel is sensed free by these two secondary users simultaneously, thereby leading to packet

collisions. In our numerical analysis, for simplicity, we assume that the sensing sequences of the two secondary users are independent and equally likely, and the order of a channel is a uniform discrete random variable in $[1, N]$. Denote $\mathbf{s}^{(m)} = \left(s_1^{(m)}, s_2^{(m)}, ..., s_N^{(m)}\right)$ as the sensing sequence of the m^{th} secondary user, and $\mathcal{S}$ the set of all possible sensing sequences. Similar to the single-secondary user scenario discussed in Section 6.4.1, if the k^{th} channel in an instinctive sensing sequence is sensed free (i.e., no primary or secondary activities), the secondary user of interest transmits its information; otherwise, proceeds to sense the next channel. Therefore, a collision occurs if the two secondary users transmit over the same channel at the same time. Letting $P_{c|\mathbf{s}^{(1)}}$ denote the conditional probability of collision given $\mathbf{s}^{(1)}$, we have

$$
\begin{aligned}
&P_{c|\mathbf{s}^{(1)}} \\
&= P(s_1^{(1)} \text{ sensed free}, s_1^{(1)} = s_1^{(2)} | \mathbf{s}^{(1)}) + P(s_1^{(1)}, s_1^{(2)} \text{ sensed busy}, s_2^{(1)} = s_2^{(2)}, s_2^{(1)} \text{ sensed free} | \mathbf{s}^{(1)}) + \\
&\quad \ldots + P(s_1^{(1)}, s_1^{(2)}, ..., s_{N-1}^{(1)}, s_{N-1}^{(2)} \text{ sensed busy}, s_N^{(1)} = s_N^{(2)}, s_N^{(1)} \text{ sensed free} | \mathbf{s}^{(1)}) \qquad (6.6) \\
&= \sum_{k=1}^{N} \frac{1}{N-k+1} \left(\prod_{i=1}^{k} \left(1 - \theta_{s_{i-1}^{(1)}}\right) \right) \cdot \\
&\quad \left(\frac{1}{\binom{N}{k-1}} + \sum_{j=2}^{\min(k, N-k+1)} \frac{\binom{k-1}{k-j}}{\binom{N}{k-1}} \sum_{i_1 > k} \sum_{i_2 > i_1} \cdots \sum_{i_{j-1} > i_{j-2}} \left(1 - \theta_{s_{i_1}^{(1)}}\right) \left(1 - \theta_{s_{i_2}^{(1)}}\right) \cdots \left(1 - \theta_{s_{i_{j-1}}^{(1)}}\right) \right) \theta_{s_k^{(1)}} \qquad (6.7)
\end{aligned}
$$

where $\theta_{s_0^{(1)}} = 0$ by definition. Thus, the average probability of collision for a secondary user, denoted by P_c, is given by

$$
P_c = \sum_{\mathbf{z} \in \mathcal{S}} P_{c|\mathbf{z}} P(\mathbf{s}^{(1)} = \mathbf{z}) \qquad (6.8)
$$

where $P(\mathbf{s}^{(1)} = \mathbf{z})$ is the probability that the first secondary user employs the channel sensing sequence $\mathbf{z}$. In the case $\theta_{s_k^{(1)}} = 1, \forall k > 0$, (6.8) corresponds to the probability of collision in conventional multi-channel MAC with two users employing random channel selection [18], given by $P_c = \frac{1}{N}$.

In general, there is a positive correlation among neighboring channels. The rationale is that, if a channel is of good (bad) quality, its neighboring channels will possibly be of good (bad) quality as well. However, such a correlation diminishes as the number of channels increases. On the other hand, the probability of collision is expected to decrease with the number of available channels, for two secondary users will be less likely to sense the same channel free at the same time.

6.4.3 Multi-Secondary User Scenario ($M > 2$)

In a CRN with M (> 2) secondary users, a collision occurs when at least two secondary users transmit over the same channel at the same time. However, we envision that, at a large N, the probability of collision that three or more secondary users transmit over the same channel simultaneously is very low. Therefore, to simplify the derivation for P_c, we consider the collisions due to two simultaneous transmissions over the same channel only. Assuming that the sensing sequences are independent and equally likely, and the channels are randomly placed in a channel sensing sequence, we obtain the conditional probability of collision given $\mathbf{s}^{(1)}$, which can be approximated by $(M-1)P_{c|\mathbf{s}^{(1)}}$, where $P_{c|\mathbf{s}^{(1)}}$ is given in (6.7). Thus,

$$P_c \approx (M-1)\sum_{\mathbf{z}\in\mathcal{S}} P_{c|\mathbf{z}} P(\mathbf{s}^{(1)} = \mathbf{z}). \tag{6.9}$$

The probability of collision given in (6.9) should provide a good approximation when N is relatively larger than M.

6.5 Performance Evaluation

6.5.1 Simulation Environment

We consider secondary users randomly located in a 1km x 1km coverage area, and adopt the channel model suggested in [44] (i.e., hilly/moderate-to-heavy tree density). Packets are generated from a secondary user and sent to a destination located at the center of the coverage area. The maximum transmission rate of each channel is 200kb/s. We consider that the maximum power constraint of a secondary user is 1mW. The background noise is assumed to be white Gaussian with zero mean and standard deviation 10^{-12}W. Other system parameters are chosen as follows: $\tau = 1\mu$s and $T = 5$ms. In simulations, we use the Shannon capacity equation [97] to calculate the achievable transmission rates of the channels. We perform the simulations for 100,000 runs and average the results, where each simulation run sustains 50,000 timeslots. The performance measurements used in our simulations are 1) reward (i.e., throughput), 2) channel utilization, 3) resource utilization, and 4) collision probability.

6.5.2 Simulation and Numerical Results

We evaluate the system performance of the proposed instinctive channel sensing with optimal stopping versus the primary-free probability θ, the number of channels N, and the number of secondary users M. We first study the performance in a single-secondary user scenario against θ and N. To further demonstrate the merits of our channel exploitation approach, we consider two channel exploitation counterparts for comparison, namely a random approach and an approach suggested in [51]. In the random approach, the sensing sequence of a secondary user is a random permutation of the set $(1, 2, ..., N)$, whereas in the Jiang's approach proposed in [51], a secondary user senses the channels according to the descending

order of their primary-free probabilities. Next, we investigate the system performance of our proposed instinctive channel sensing in a two-secondary user scenario and a multi-secondary user scenario against N and M. In the simulations, we consider two cases for θ: 1) equal θ, where $\theta_{s_k} = \theta, \forall k$ and $\theta = H$ with $H \in \{0.001, 0.005, 0.01, 0.03, 0.1, 0.3, 0.6, 0.9, 1\}$; and 2) unequal θ, where θ is uniformly distributed on $[0, H]$.

Effect of primary-free probability, θ

For $N = 512$, we study the impact of the value of θ on the reward performance. Figure 6.3 depicts the simulation and numerical results for the reward performance versus the value of θ. As anticipated, our simulation and numerical results agree with each other. Also, it is clear that the rewards in both cases of equal θ and unequal θ increase with the value of primary-free probability. However, it is worth mentioning that the rates of the reward performance improvement in both equal θ and unequal θ cases decrease with θ. The rationale for such a performance saturation is that, when the value of θ is small, a marginal increase in θ can greatly increase the transmission opportunities of a secondary user. In addition, with the increased primary-free probabilities, there are more free channels for a secondary user to choose from, resembling the case of multi-user diversity. Therefore, the reward performance increases sharply. At a large θ, however, a marginal increase in the value of θ only has a minimal impact on the channel availabilities and the channel gain variations across those free channels. Thus, the reward performance almost levels off from $\theta = 0.6$ onward. The channel utilization versus θ is shown in Figure 6.4. The trend of channel utilization is the same as that of reward performance shown in Figure 6.3. As observed in Figure 6.4, our proposed sensing approach can achieve almost 100% channel utilization at a large θ, making efficient use of unoccupied channels. There is a performance gap between the curve with equal θ and that with unequal θ. The performance difference is mainly due to the randomness of primary-free probabilities in the case of unequal θ. In other words, on average, more (less)

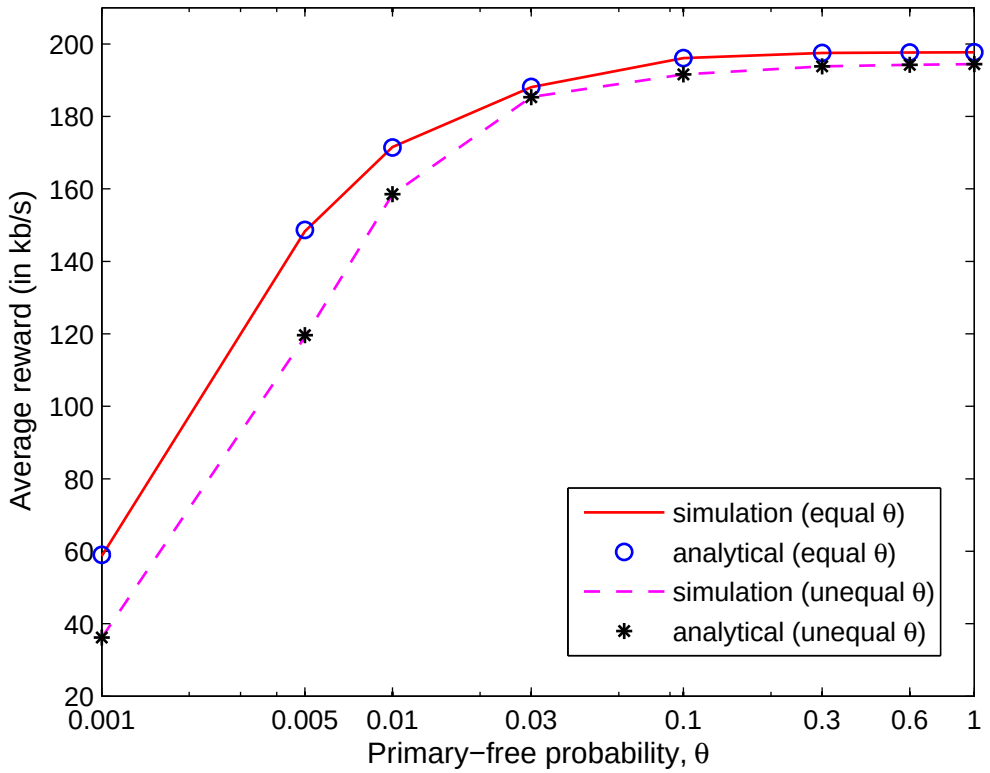


Figure 6.3: Reward performance of the proposed instinctive channel sensing in a single-secondary user scenario vs. the value of θ (where $N = 512$).

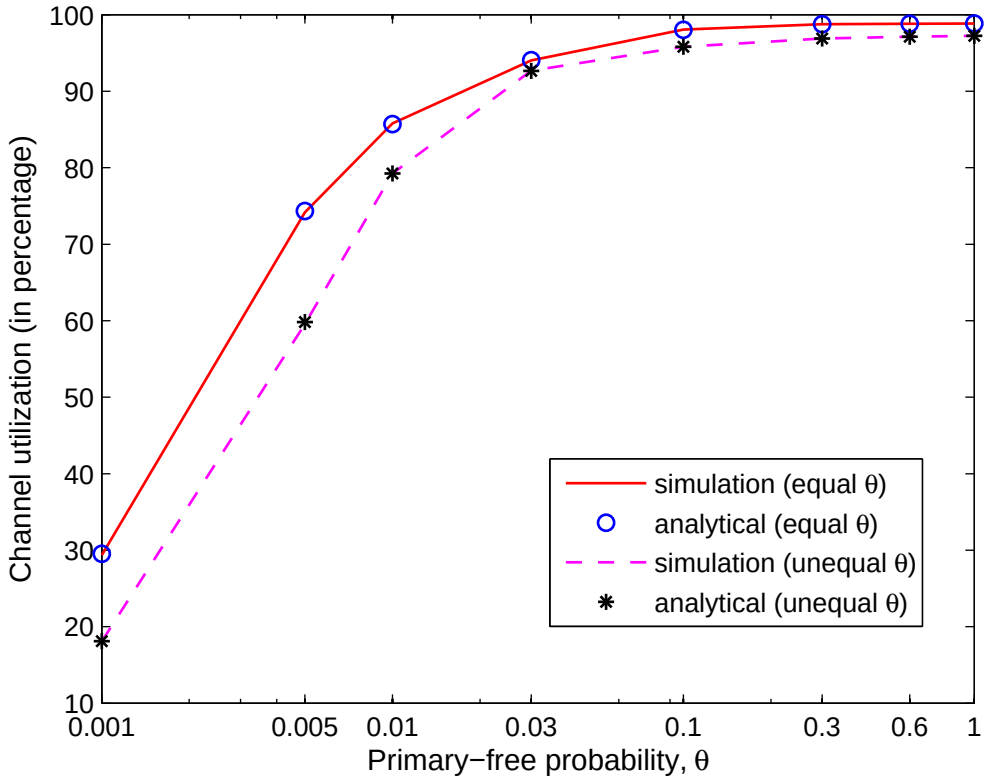


Figure 6.4: Channel utilization of the proposed instinctive channel sensing in a single-secondary user scenario vs. the value of θ (where $N = 512$).

channels are available in the case of equal (unequal) θ. On the other hand, for the same mean value of θ, we observe that the channel utilization for equal θ and that for unequal θ behave differently, depending on the mean value of θ. In Figure 6.4, at a small θ, the curve for unequal θ outperforms that for equal θ. For example, at the mean value of 0.005, the curve for the unequal θ attains 79%, whereas the curve for the equal θ attains only 75%. In contrast, at a large θ, the curve for equal θ achieves a higher channel utilization than that for unequal θ. For example, at the mean value of 0.3, the curves for the equal θ and unequal θ attain 98% and 96%, respectively. These observations can be explained by our previous discussion on the impact of increased transmission opportunities on the reward performance.

Effect of the number of channels, N

For $\theta = 0.3$ and $\theta \in [0, 0.3]$, we study the impact of the number of available channels on the reward performance shown in Figure 6.5. As seen, the curves for the reward performance in both cases of equal θ and unequal θ rise with the value of N. Similar to the discussion in Section 6.5.2, the reward increase is due to the fact that the more the available channels in a CRN, the more the variations in the channel gains across the (free) channels, realizing multi-user diversity. In Figure 6.6, the channel utilization is also plotted for reference. It is observed that the channel utilization can be further improved at a larger N. We observe that, in Figures 6.5 and 6.6, our simulation results and numerical results closely match, validating our performance analysis.

Performance comparison

For $N = 512$, we compare our proposed instinctive channel sensing order with the random sensing order and the Jiang's sensing order. In the performance comparison, we further consider two different settings for the CRN, namely good resource utilization by primary users (e.g., $\theta = 0.005$ and $\theta \in [0, 0.005]$) and poor resource utilization by primary users

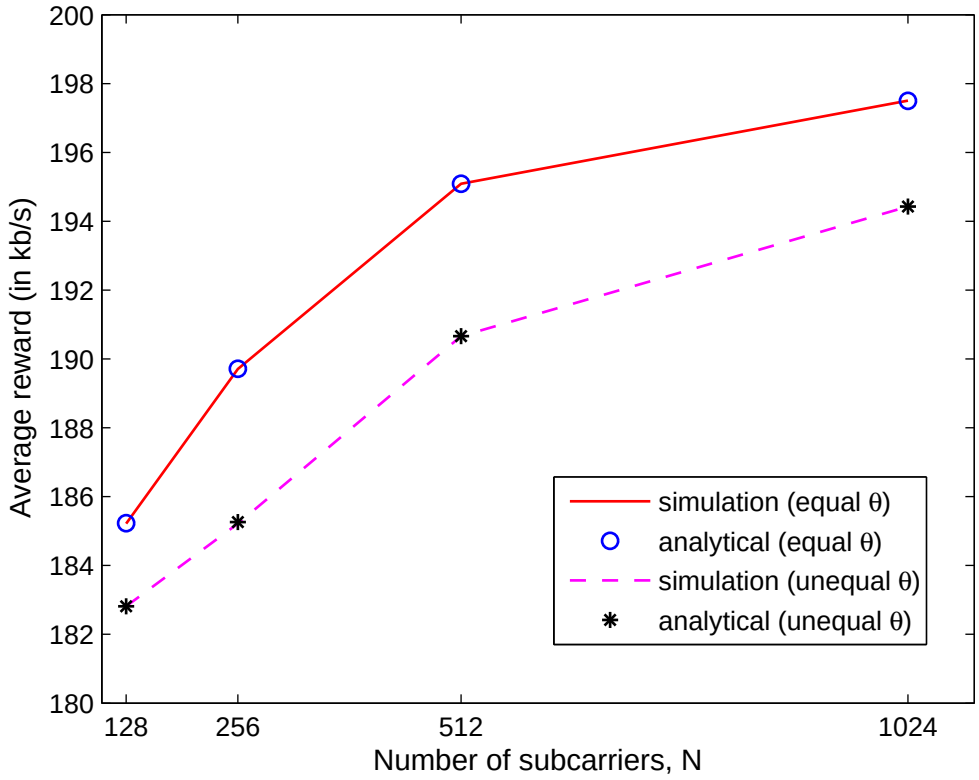


Figure 6.5: Reward performance of the proposed instinctive channel sensing in a single-secondary user scenario vs. the value of N (where $\theta = 0.3$).

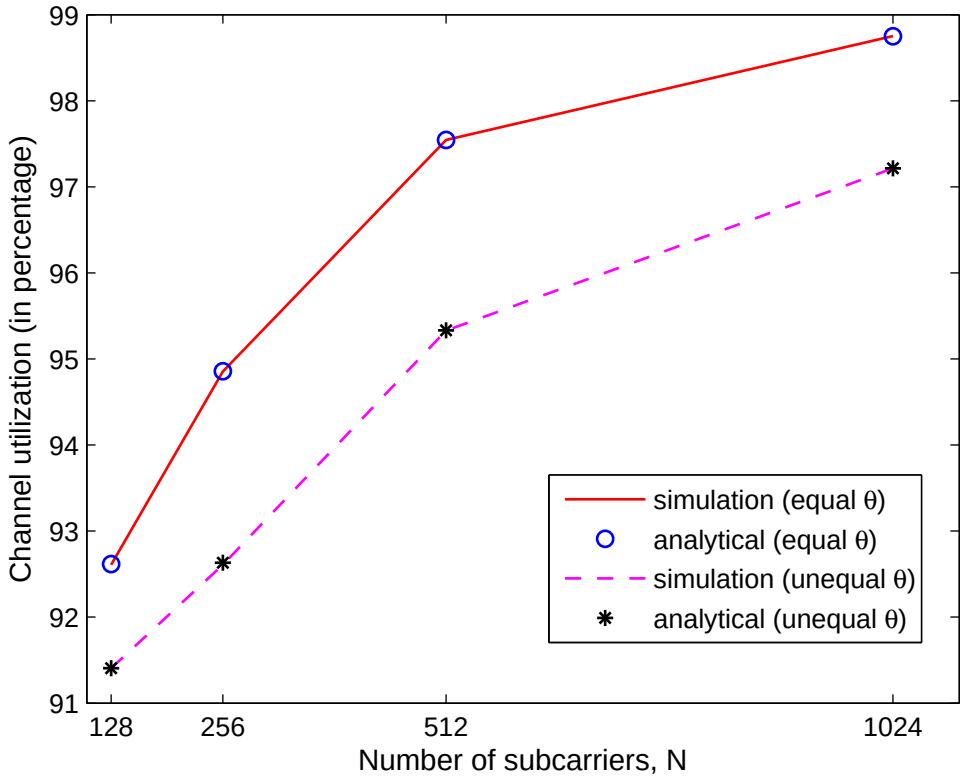


Figure 6.6: Channel utilization of the proposed instinctive channel sensing in a single-secondary user scenario vs. the value of N (where $\theta = 0.3$).

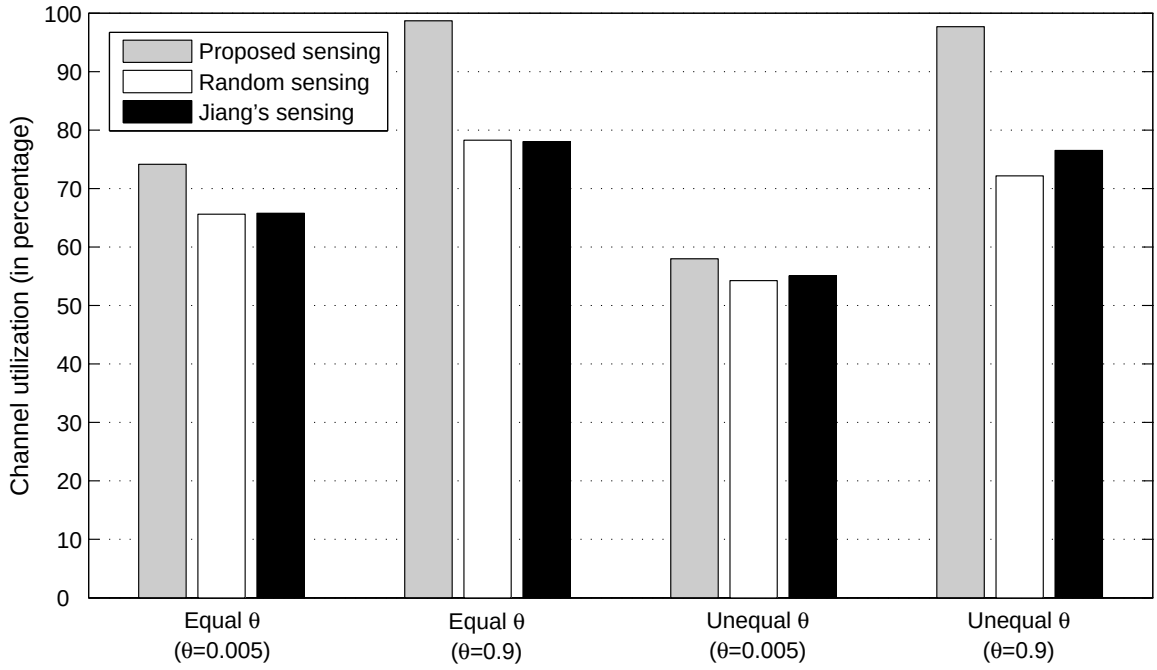


Figure 6.7: Comparison of the channel utilization of our proposed instinctive sensing approach, a random sensing approach, and the Jiang's sensing approach vs. the value of θ (where $N = 512$).

(e.g., $\theta = 0.9$ and $\theta \in [0, 0.9]$). Four observations can be made from the channel utilization achieved by the sensing approaches shown in Figure 6.7: First, the channel utilization for equal θ is generally better than that for unequal θ, as expected; Second, since the Jiang's approach is sensitive to the order of primary-free probabilities, it performs better than the random approach in the cases of unequal θ. On the other hand, the random approach and the Jiang's approach attain almost the same channel utilization in the equal θ case. Note that a secondary user employing the Jiang's sensing order is to sense the channels according to their primary-free probabilities. As such, if all the primary-free probabilities are equal, the sensing sequence of a secondary user essentially becomes $(s_1, s_2, ..., s_N) = (1, 2, ..., N)$. Since the channel gains in our simulations are generated randomly, sensing the channels

in a sequential manner is more or less the same as in a random manner. Nevertheless, our proposed instinctive channel sensing achieves the highest channel utilization in all the cases. The performance gains are due to the fact that our instinctive sensing order takes into account the unique characteristics of the channels. By sensing the channels according to the descending order of their achievable rates, a secondary user can take advantage of the channel variations, whereby an improved performance can be acquired; Third, when the primary-free probabilities are low (i.e., $\theta = 0.005$ and $\theta \in [0, 0.005]$), all three approaches achieve similar channel utilization. The justification is that, since the number of free channels is very small and the transmission opportunities for a secondary user are low, the channels sensed free in all of the three approaches are more or less the same. Our proposed approach performs slightly better than the others, as the benefits of channel variations can be exploited to a certain extent; Fourth, since more channels are available for a secondary user in the CRN with poor resource utilization by primary users (i.e., $\theta = 0.9$ and $\theta \in [0, 0.9]$), the channel utilization achieved by a secondary user is higher than in the CRN with good resource utilization (i.e., $\theta = 0.005$ and $\theta \in [0, 0.005]$). With more free channels and hence more channel variations, our proposed sensing order outperforms its two counterparts by at least 18%, leading to the best candidate for channel exploitation in CRNs. Concerning the reward performance, we observe that the trend for the three channel exploitation approaches is the same as that shown in Figure 6.7.

Collision probability P_c in a two-secondary user case

Here, we evaluate the probability of collision (P_c) in a two-secondary user scenario, where the two secondary users under consideration employ the same proposed instinctive sensing. In the case of equal θ, we study the impacts of the values of θ and N on the probability of collision. Both simulation and analytical results are given in Table 6.2. There are two main trends: First, for the same value of θ, P_c generally deceases with N; Second, for the same

Table 6.2: Simulation and analytical results for the probability of collision, P_c, in a two-secondary user scenario vs. the number of channels, N, for different values of θ.

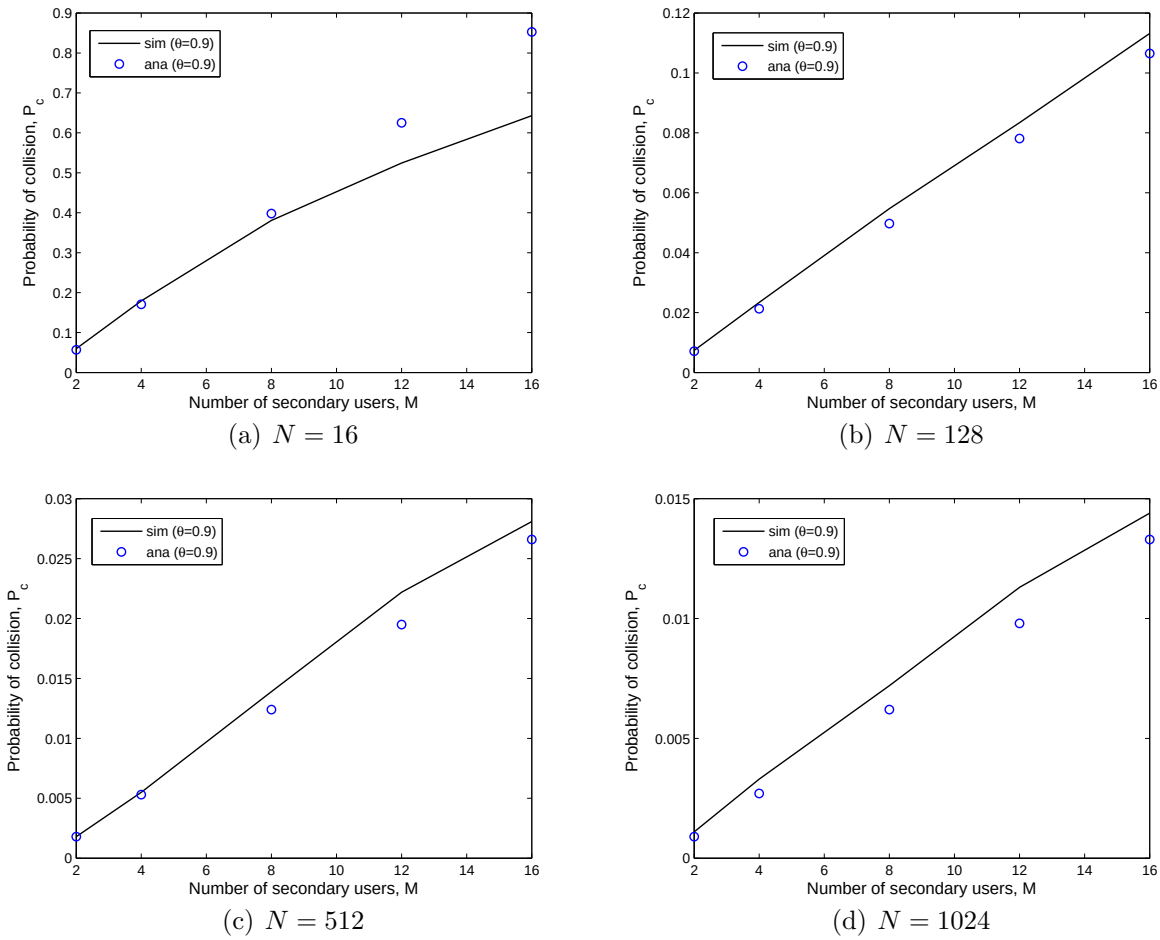
N		2	4	16	32	128	256	512	1024
$\theta = 0.005$	simulation	0.0055	0.0058	0.0046	0.0045	0.0034	0.0030	0.0016	0.0005
	analytical	0.0050	0.0050	0.0048	0.0045	0.0033	0.0023	0.0013	0.0006
$\theta = 0.01$	simulation	0.0103	0.0110	0.0100	0.0070	0.0045	0.0029	0.0012	0.0005
	analytical	0.0100	0.0098	0.0091	0.0082	0.0046	0.0025	0.0011	0.0005
$\theta = 0.1$	simulation	0.1092	0.1033	0.0439	0.0205	0.0049	0.0023	0.0012	0.0005
	analytical	0.0950	0.0831	0.0408	0.0202	0.0043	0.0021	0.0010	0.0005
$\theta = 0.3$	simulation	0.2603	0.1902	0.0404	0.0199	0.0059	0.0032	0.0014	0.0006
	analytical	0.2550	0.1721	0.0410	0.0192	0.0046	0.0023	0.0012	0.0006
$\theta = 0.6$	simulation	0.4212	0.2223	0.0480	0.0247	0.0069	0.0033	0.0015	0.0007
	analytical	0.4200	0.2100	0.0459	0.0226	0.0056	0.0028	0.0014	0.0007
$\theta = 0.9$	simulation	0.5033	0.2652	0.0588	0.0306	0.0082	0.0037	0.0020	0.0009
	analytical	0.4950	0.2351	0.0572	0.0285	0.0071	0.0036	0.0018	0.0009
$\theta = 1$	simulation	0.5061	0.2621	0.0672	0.0312	0.0076	0.0042	0.0021	0.0009
	analytical	0.5000	0.2500	0.0625	0.0313	0.0078	0.0039	0.0020	0.0010

value of N, P_c generally increases with θ. The first observation can be explained as follows. As the number of channels increases, the chances for the two secondary users to sense the same channel free simultaneously decrease, thereby leading to a smaller P_c. Concerning the second observation, on one hand, the transmission opportunities for both secondary users increase with the value of θ. On the other hand, since the number of channels is fixed, the increased transmission opportunities cause a higher probability of collision, for the two secondary users are more likely to sense the same channel free at the same time. On a different note, we observe that there is some discrepancy between the simulation results and the analytical results, especially for a small N. As mentioned in Section 6.4.2, there is a (positive) correlation among neighboring channels, and choosing which channel to be sensed next generally depends on the previously sensed channels. This phenomenon exacerbates as

the number of channels decreases. Therefore, the assumption that the order of a channel is uniformly distributed in our numerical analysis might be void. Nonetheless, both simulation and analytical results agree with each other at a large N. In practice, the number of available channels is mostly large (e.g., 512 or more channels) and, therefore, our numerical analysis is valid in typical OFDM-based networks. Further, we observe that the total reward obtained in a two-secondary user scenario is higher than that in a single-secondary user scenario given the same value of θ. Since the system is not saturated, a higher reward can be attained with more secondary users in the system.

Collision probability and resource utilization in a multi-secondary user case

For the case of equal θ with $\theta = 0.9$, we study the impact of the number of secondary users M (> 2) on the probability of collision and resource utilization. In specific, we gauge the system performance with different values of N and M. Numerical and simulation results for the collision probability are depicted in Figure 6.8. As expected, the probability of collision increases with M. We also notice that P_c rises more drastically with M when N is small, compared to the case when N is large. This phenomenon is ascribed to the elevated chances of simultaneous transmissions over the same sensed free channel(s) when the number of channels is limited. Similar to the previous discussions, the probability of collision increases as N decreases. We observe that there is some performance discrepancy between our simulation and numerical results at a small N. In typical OFDM-based systems with a large number of channels, however, our analytical results generally provide a good approximation of the collision probability in a multi-secondary user scenario, validated by the simulation results. We also notice that the trend of P_c versus M for a small θ is similar to that shown in Figure 6.8. Due to less transmission opportunities, however, it is observed that a lower P_c is achieved in the case of a smaller θ. Figure 6.9 depicts the total reward (i.e, throughput) versus M. Clearly, the total reward increases with M. The rationale is



(a) $N = 16$ (b) $N = 128$

(c) $N = 512$ (d) $N = 1024$

Figure 6.8: Simulation and analytical results for collision probability, P_c, in a multi-secondary user scenario vs. the number of secondary users, M, for different values of N (where $\theta = 0.9$).

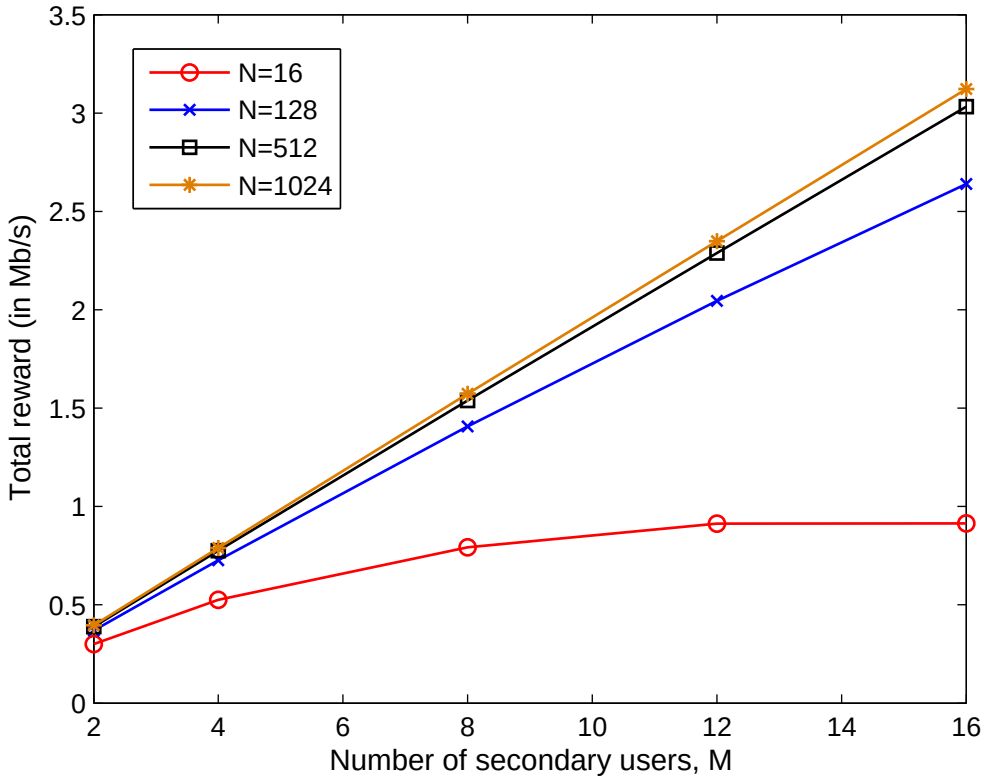


Figure 6.9: Total reward performance of the proposed instinctive channel sensing vs. the value of M (where $\theta = 0.9$).

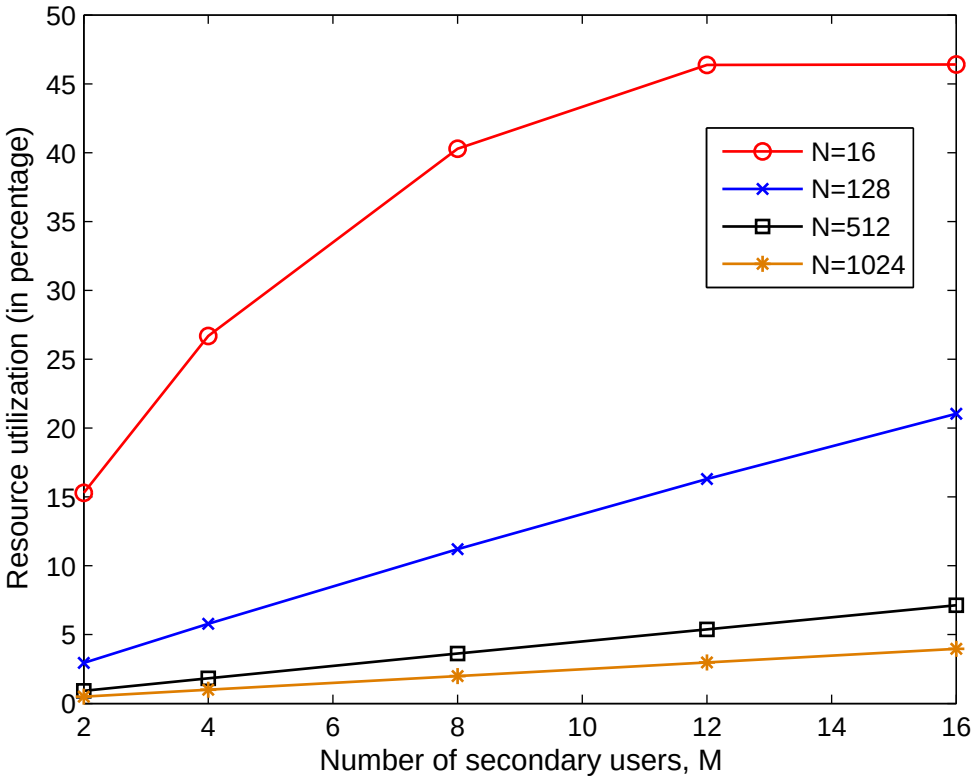


Figure 6.10: Resource utilization of the proposed instinctive channel sensing vs. the value of M (where $\theta = 0.9$).

that, since the system is not saturated, the more the secondary users, the higher the total reward. On the other hand, the rate of increment in the total reward increases with N due to a lower P_c and increased transmission opportunities (and hence multi-user diversity). On the contrary, the impact of N has the opposite effect on the resource utilization. Figure 6.10 shows the resource utilization versus M. As expected, the resource utilization increases with M. However, given the value of M, the smaller the value of N, the higher the resource utilization. In particular, when $N = 16$, the resource utilization increases from 15% to about 47% as M increases from 2 to 16, whereas when $N = 1024$, the resource utilization is less than 5%. The difference in the resource utilization is due to the fact that a secondary

user can transmit its data over a single channel only. Therefore, in a system with a large number of channels (e.g., 512 channels), a small number of secondary users can utilize only a small portion of temporarily unoccupied channels, even if our throughput-optimal instinctive sensing order is employed. In contrast, when the number of channels is small (e.g., 16 channels), the resources can be better utilized by the same number of secondary users. The implication is that resource utilization can be improved when M approaches N at the cost of a high P_c. Therefore, given the number of available channels, how to procure a desired number of secondary users such that the resource utilization and collision probability are optimized is certainly of great importance; however, addressing this issue is left for further work. On the other hand, if *channel bundling* is permitted (see Section 6.6), we observe that both the total reward and resource utilization can be further improved.

6.6 Discussion

In practice, channel bundling can be employed so as to reduce signalling overhead and facilitate efficient resource allocation. The idea is that, instead of allocating the channels one by one, we group a set of channels, referred to as a channel bundle, and allocate them at once, thereby reducing computational cost. In the context of CRNs, since all the channels in the same bundle are either available or busy simultaneously, a secondary user can sense each bundle of channels instead of each channel individually. For example, every 8 channels can be grouped as a bundle, and the 4^{th} channel can be viewed as its representative. As such, with respect to our proposed instinctive sensing, a secondary user essentially senses the channel bundles according to the descending order of the achievable rates of their 4^{th} channels. Therefore, if the 4^{th} channel of a bundle is sensed free, a secondary user transmits its information over the channels in that bundle, or proceeds to sense the 4^{th} channel of the next bundle otherwise. With the channel bundling in place, the system throughput can

be increased, for more free channels can be utilized by each secondary user per timeslot. In general, however, the best 4^{th} channel is not tantamount to the best channel bundle. For instance, some channel bundle can achieve the maximum total achievable rate among the channel bundles, but the rate of its representative (e.g., 4^{th} channel) may not be the maximum among all the individual channels. As a result, in the presence of channel bundling, the optimality of our proposed stopping rule and instinctive sensing can be void. All in all, optimal channel bundling is strongly desired for the sake of system performance enhancement and computational complexity reduction; addressing this issue, however, is left for further work. On the other hand, node clustering can also facilitate channel sensing, since each cluster is allocated a subset of the available channels (see Chapter 3). Instead of sensing all the available channels, secondary users in clustered WMNs only need to sense the channels allocated to their corresponding clusters. As such, since the number of channels to be sensed by secondary users is smaller, the overhead due to channel sensing can be reduced, and the system performance can be improved. In conventional CRNs, secondary users can only transmit when there is free of primary activity; yet, such a traditional methodology can, in fact, hinder network resources from being efficiently utilized. According to our research findings in Chapter 5, the benefits of node cooperation can be exploited so as to further improve the system performance of WMNs provided that the QoS demands of performance-guaranteed nodes can be satisfied. By the same token, as long as the transmissions of licensed primary users assisted by unlicensed secondary users are beneficial in terms of QoS support and throughput increase, secondary users should be allowed to transmit their packets together with primary users' packet transmissions, thereby making better use of scarce radio resources.

6.7 Chapter Summary

In this Chapter, a simple instinctive channel sensing order with optimal stopping is proposed for CRNs. Provided that the primary-free probability is the same for all the channels, we prove that our proposed channel sensing order is throughput-optimal. In a single-secondary user scenario, our proposed channel sensing approach is demonstrated promising, outperforming a random sensing approach and the approach suggested in [51] by at least 18% in terms of channel utilization when the network resources are poorly utilized by primary users. In the case of two secondary users in a CRN, our numerical and simulation results show that, by employing the proposed instinctive sensing, the probability of collision generally decreases as the number of channels increases and/or the primary-free probabilities decrease. In a CRN with multiple secondary users, our results show that the more the secondary users, the higher the collision probability. On the other hand, as the number of secondary users increases, the total reward increases. Our study also indicates that, when the number of secondary users approaches the number of channels, resource utilization can be further improved at the expense of a higher collision probability.

Chapter 7

Balancing Throughput and Fairness with QoS Assurance

7.1 Introduction

With limited network resources, throughput maximization, QoS provisioning, and fairness assurance are conflicting goals [71], conducing to a natural tradeoff among these three performance measures. For WMNs, in particular, balancing system throughput and fairness with QoS assurance and high resource efficiency is of great interest. As such, a unified (optimization) framework to effectively attain different degrees of performance tradeoff between throughput and fairness with QoS support is strongly desired. In this research, from the perspective of call-level resource allocation, we address the problem of balancing throughput and fairness for WMNs with QoS assurance.

The contributions and significance of this research work are two-fold [22].

- First, we study the problem of balancing throughput and (weighted max-min) fairness for interference-limited WMNs with QoS assurance. By introducing the notion of

bargaining floor, we derive a unified optimization framework to bridge the problem of throughput maximization and the problem of weighted max-min fairness.

- Second, with our proposed QoS-aware framework, an optimal relationship curve of system throughput and fairness is obtained. In specific, different degrees of performance tradeoff between system throughput and fairness can be attained, simply by adjusting the value of the proposed bargaining floor. Our study shows that the tradeoff curve of system throughput against fairness is concave in shape, meaning that a unit decrease in system throughput leads to a larger marginal improvement in fairness performance. Further, the resource allocation solutions from our proposed framework achieve Pareto optimality, making efficient use of network resources.

The remainder of this Chapter is organized as follows. Related work is discussed in Section 7.2. The optimization problem formulation is presented in Section 7.3. Efficiency of the proposed resource allocation is addressed in Section 7.4. Numerical results are given in Section 7.5. Practical implementation issues are discussed in Section 7.6. Finally, a summary is drawn in Section 7.7. A summary of important symbols used in this Chapter is given in Table 7.1 for easy reference.

7.2 Related Work

In the literature, utility optimization is a commonly used mathematical tool to measure system performance subject to certain constraints (e.g., QoS requirements) [26,54,60,69], where a utility function is described as a measure of user satisfaction. Proportional fairness can be obtained by choosing suitable utility functions [60], and other performance measures such as throughput can also be incorporated into this optimization formulation [54] (see (5.13)). With different problem formulations (or utility functions), different optimal solutions can be

Table 7.1: Summary of important symbols used in this Chapter.

Symbol	Definition
R_m^d	effective bandwidth of a call on the m^{th} link
$R_m(\mathbf{a})$	actual transmission rate of the m^{th} link
$Q_m(\mathbf{a})$	utility function (i.e., extra throughput obtained) of the m^{th} link, i.e., $Q_m(\mathbf{a}) = R_m(\mathbf{a}) - R_m^d$
w_m	weighting factor of the m^{th} link, i.e., $w_m > 0$
P_m	maximum transmit power level of the m^{th} link's transmitter
G_{mk}	channel gain from the k^{th} link's transmitter to the m^{th} link's receiver
a_m	power scaling factor of the m^{th} link's transmitter, i.e., $a_m \in [0, 1]$
γ_m	signal-to-interference-plus-noise ratio (SINR) of the m^{th} link
σ	cross-correlation factor between any two signals, i.e., $\sigma \in (0, 1]$
η	background noise power
M	total number of active links in the network
J	bargaining floor, i.e., $J \in [0, J^*]$, where J^* is the maximum value of J
D	deviation of $\min_m \{w_m Q_m(\mathbf{a})\}$ from J^*, i.e., $D = \lvert J^* - \min_m \{w_m Q_m(\mathbf{a})\}\rvert$
U	measure of system throughput, i.e., $U \in [0, 1]$
V	measure of fairness, i.e., $V \in [0, 1]$
W	shaped Jain's fairness index, i.e., $W \in [0, 1]$

obtained. Pricing schemes [60] can be employed to achieve a tradeoff between throughput and fairness, to a certain extent. However, the utility functions used in the preceding work carry no or little physical meaning. How to find a meaningful utility function with an appropriate pricing scheme can be problematic. In addition, most of the current work assumes that the utility functions can be separable in the dual problem [26,54,60,69], which is not always feasible, especially for interference-limited systems (e.g., CDMA and ultra-wideband (UWB) systems), meaning that directly applying existing approaches (e.g., [26]) generally results in suboptimal solutions. In the interference-limited systems, we need to consider all the (user) utility functions for the sake of optimality.

Ideal (weighted max-min) fairness can be obtained by generalized processor sharing (GPS) [89,90] or its variants [70], where all nodes in a network share the total resources.

With GPS, the weight of a node determines the amount of allocated resources. As such, each node can have a fair share of network resources. However, the notion of a weight is an abstract concept, and the question of how to relate QoS requirements to the weight effectively remains an open research problem. In GPS, even though all the weights or QoS requirements are already satisfied, only (weighted max-min) fairness is considered, and its throughput performance needs further investigation.

Although some research aims at the tradeoff between throughput and fairness in telecommunications networks, only heuristic schemes are proposed without any optimality consideration [29,77,108]. Thus far, only limited relationship of system throughput and fairness is addressed. Joint consideration of both performance metrics and hence a unified framework attaining different degrees of performance tradeoff between them with QoS support are strongly desired. Further, with scarce radio resources, efficient resource utilization is vital, which can be verified by game theory [88]. All these aspects are addressed in this Chapter.

7.3 Balancing Throughput and Fairness

7.3.1 Interference-Limited Network Model

We consider a generic system model which is an interference-limited WMN. Let R_m^d denote the *effective bandwidth* of a call on the m^{th} link, which is the minimum rate required to satisfy the QoS requirements and depends on source traffic characteristics [58,115]. Let $R_m(\mathbf{a})$ denote the actual transmission rate of the m^{th} link where $\mathbf{a} = (a_1, a_2, ..., a_m, ..., a_M)$ and a_m is the power scaling factor of the m^{th} link's transmitter, i.e., $a_m \in [0, 1]$, and M the total number of active links in the network. For simplicity, the actual transmission rate of the m^{th} link is given by [97]

$$R_m(\mathbf{a}) = \log_2 (1 + \gamma_m) \tag{7.1}$$

where γ_m is the SINR of the m^{th} link. In (7.1),

$$\gamma_m = \frac{G_{mm} P_m a_m}{\sigma \sum_{k \neq m} G_{mk} P_k a_k + \eta} \tag{7.2}$$

where P_m is a maximum transmit power level of the m^{th} link's transmitter, G_{mk} the channel gain from the k^{th} link's transmitter to the m^{th} link's receiver, σ the cross-correlation factor between any two signals, i.e., $\sigma \in (0, 1]$, and η the background noise power.

In order to effectively balance system throughput and fairness, CAC is indispensable, which can be contingent on our optimization framework (to be discussed in Section 7.3.2). The call admission routine[1] is invoked whenever a new call arrives. Each incoming new call is admitted as long as there exists a feasible solution of the optimization problem, meaning that the QoS requirements (i.e., in terms of effective bandwidth) of this new call and all other calls in service can be satisfied. If there is no feasible solution of the optimization problem, the new call is rejected due to insufficient resources available to meet its QoS requirements. In other words, the criterion of call admission is tantamount to the feasibility of the solution of our optimization framework. With the CAC in place, the QoS requirements of all admitted calls can be guaranteed, and the operation of balancing throughput and fairness can be carried out effectively.

Given that all the QoS requirements are met, the network performance can be further improved for increasing system throughput and/or maintaining (weighted max-min) fairness. In this research, the notion of weighted max-min fairness is employed as the fairness performance of interest, which corresponds to the ideal fairness achieved by the GPS.

[1]Notice that routing is involved in the CAC. Particularly, the tasks of *QoS routing* include 1) route discovery, 2) call admission control over each link, and 3) route repair. Therefore, which link a call is to traverse along can be determined in advance [1].

7.3.2 Optimization Problem Formulation

We first consider two optimization problem formulations, namely system throughput optimization and weighted max-min fairness optimization. Next, we propose a generic optimization problem formulation with the consideration of system throughput and fairness. For the system throughput optimization, the optimization problem is given by

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} R_m(\mathbf{a}) \right\} \tag{7.3}$$

$$\text{subject to } R_m(\mathbf{a}) \geq R_m^d, 0 \leq a_m \leq 1, \forall m \tag{7.4}$$

where $\mathbf{a} = (a_1, a_2, ..., a_m, ..., a_M)$ is the *optimization variable*. In fact, we can rewrite (7.3) as

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} \left(R_m(\mathbf{a}) - R_m^d \right) \right\} \tag{7.5}$$

as the optimality still maintains for a linear-shifted objective function [107]. The system throughput optimization problem (STOP) can be rewritten as

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \tag{7.6}$$

$$\text{subject to } Q_m(\mathbf{a}) \geq 0, 0 \leq a_m \leq 1, \forall m \tag{7.7}$$

where $Q_m(\mathbf{a}) = R_m(\mathbf{a}) - R_m^d$. The physical meaning of $Q_m(\mathbf{a})$ is the amount of extra resources allocated to (i.e., excess throughput obtained by) the m^{th} link. In fact, $Q_m(\mathbf{a})$ can be viewed as the utility function of user m in the conventional utility maximization [54,60,69], as $Q_m(\mathbf{a})$ of user m is: 1) increasing; 2) strictly concave; and 3) twice differentiable (i.e., continuous) on a_m. The gist of (7.6) is to maximize the sum of all the utility functions. In the STOP, however, the utility functions are not separable in the dual problem as $Q_m(\mathbf{a})$ only increases with a_m but not over $\mathbf{a}$. Consider the following (partial) dual problem [69]

$$\min_{\boldsymbol{\lambda} \succeq \mathbf{0}} D(\boldsymbol{\lambda}) \tag{7.8}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_m, ..., \lambda_M)$ is a set of Lagrange multipliers and

$$D(\boldsymbol{\lambda}) = \max_{\mathbf{a}} \left\{ \sum_m Q_m(\mathbf{a}) + \sum_m \lambda_m Q_m(\mathbf{a}) \right\}. \tag{7.9}$$

Suppose that the optimality still maintains when each (user) utility function, $Q_m(\mathbf{a})$, is optimized separately. Intuitively, each user m will simply choose the maximum power scaling factor, i.e., $a_m = 1, \forall m$. In most cases, due to the cross-interference (i.e., $\sigma \neq 0$), the overall result will not contribute to the maximum value of (7.6). Optimizing user utility functions separately does not always conduce to the maximal solution of the STOP, for, in general, from (7.9), we have

$$D(\boldsymbol{\lambda}) = \max_{\mathbf{a}} \left\{ \sum_m Q_m(\mathbf{a}) + \sum_m \lambda_m Q_m(\mathbf{a}) \right\} \tag{7.10}$$

$$= \max_{\mathbf{a}} \left\{ \sum_m (1 + \lambda_m) Q_m(\mathbf{a}) \right\} \tag{7.11}$$

$$\neq \sum_m \max_{a_m} \{(1 + \lambda_m) Q_m(\mathbf{a})\}. \tag{7.12}$$

Thus, the solution space (i.e., resource allocation) is not necessarily the same as those proposed in the literature (e.g., [69]). Obtaining the optimal solution requires a joint consideration of all user utility functions, meaning that the existing solutions suggested in the literature cannot be directly applied.

For the weighted max-min fairness optimization problem (WMMFOP), the corresponding formulation is given by [93]

$$\max_{\mathbf{a}} \left\{ \min_m \{w_m Q_m(\mathbf{a})\} \right\} \tag{7.13}$$

$$\text{subject to } Q_m(\mathbf{a}) \geq 0, 0 \leq a_m \leq 1, \forall m \tag{7.14}$$

where w_m is a weighting factor of the m^{th} link, i.e., $w_m > 0$, which indicates the *unwillingness* of obtaining extra allocated resources[2], i.e., the smaller the value of w_m, the more eager is

[2]Notice that the meaning of the weighting factor w_m of our interest is different from that in GPS.

the m^{th} link to obtain more extra resources. The use of w_m is necessary for effective and efficient resource allocation in WMNs with heterogeneous traffic (e.g., voice, video, and data traffic). For example, after its effective bandwidth requirement is satisfied, allocating extra resources to voice traffic may be wasteful as the quality of signal reception is already good, and hence this traffic should be assigned a large weighting factor. On the other hand, a smaller weighting factor should be assigned to data traffic, as it demands more throughput. With different weighting factors, different traffic classes can be differentiated, facilitating per-class packet scheduling and hence QoS provisioning.

Proposition 17 *The set of feasible weighted utilities (i.e., $w_m Q_m(\mathbf{a}), \forall m$) in the WMMFOP has the solidarity property [93].*

Here, we transform (7.13) into [107]

$$\max_{\mathbf{a}} L \tag{7.15}$$

$$\text{subject to } w_m Q_m(\mathbf{a}) \geq L, \forall m. \tag{7.16}$$

Therefore, the WMMFOP can be rewritten as follows.

$$\max_{\mathbf{a}} L \tag{7.17}$$

$$\text{subject to } w_m Q_m(\mathbf{a}) \geq L, 0 \leq a_m \leq 1, \forall m. \tag{7.18}$$

As we prove in Proposition 17, the WMMFOP possesses the solidarity property. As such, at the optimal point, all $w_m Q_m(\mathbf{a})$, $\forall m$, are to be equal, and the maximum value of L for the WMMFOP is unique [93]. Let J^* and $\hat{\mathbf{a}}$ denote the optimal solutions (i.e., maximal L and optimal $\mathbf{a}$) of the WMMFOP. If the constraints $w_m Q_m(\mathbf{a}) \geq J^*, \forall m$, are added to the STOP, the *modified* STOP becomes

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \tag{7.19}$$

$$\text{subject to } w_m Q_m(\mathbf{a}) \geq J^*, 0 \leq a_m \leq 1, \forall m. \tag{7.20}$$

Proposition 18 *The optimal solution* $\hat{\mathbf{a}}$ *obtained from the WMMFOP is also the optimal solution for the modified STOP.*

With the new constraint set, the solution $\mathbf{a}$ obtained from the modified STOP not only achieves weighted max-min fairness, but also optimizes the system throughput. Therefore, to bridge the system throughput and fairness performance measures together, we introduce a parameter called *bargaining floor*, denoted by J, where $J \in [0, J^*]$ with J^* being the optimal solution (i.e., the maximal value) of the WMMFOP. Motivated by Proposition 18, we propose the following generic optimization problem (GOP)

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \tag{7.21}$$

$$\text{subject to } w_m Q_m(\mathbf{a}) \geq J, 0 \leq a_m \leq 1, \forall m. \tag{7.22}$$

Clearly, the solutions of the GOP for maximal system throughput can be obtained when $J = 0$ while that for maximal (weighted max-min) fairness when $J = J^*$, where J^* can be obtained from the WMMFOP. In this research, our focus is not to solve the GOP. Instead, we treat the GOP as a unified framework for deducing the optimal relationship between system throughput and fairness with QoS support. Let $\mathbf{a}^*$ be the optimal solution obtained from the GOP.

Proposition 19 *The system throughput (i.e.,* $\sum_{m=1}^{M} Q_m(\mathbf{a}^*)$*) is a non-increasing function of bargaining floor* J*.*

Corollary 3 *The minimum value of* $w_m Q_m(\mathbf{a}^*)$ *(i.e.,* $\min_m \{w_m Q_m(\mathbf{a}^*)\}$*) is a non-decreasing function of* J*.*

Theorem 1 *The system throughput (i.e.,* $\sum_{m=1}^{M} Q_m(\mathbf{a}^*)$*) does not increase with* J*, but the minimum value of* $w_m Q_m(\mathbf{a}^*)$ *(i.e.,* $\min_m \{w_m Q_m(\mathbf{a}^*)\}$*) does not decrease with* J*.*

Corollary 4 *A relationship between the system throughput and weighted max-min fairness performance can be achieved by solving the GOP with different values of J.*

From the perspective of network design, since the network resources are limited, improving fairness performance will reduce the system throughput, which matches with the perspective of Corollary 4. With different values of J, a tradeoff curve of system throughput and fairness can be obtained. Thus, the GOP should be solved with different values of J iteratively. The iterative procedure to obtain the tradeoff curve is described below:

Step 1: Find J^* by solving the WMMFOP;

Step 2: Set $J = 0$ and solve the GOP, whereby the obtained solution $\mathbf{a}$ corresponds to the maximal throughput performance;

Step 3: Increase J by δJ and solve the GOP;

Step 4: Repeat Step 3 until $J = J^*$, which corresponds to the maximal fairness.

Through the above procedure, different sets of the optimal solution $\mathbf{a}$, the corresponding relationships of system throughput and fairness, and hence a desired tradeoff curve can be attained by suitably selecting the value of J. Notice that, after the tradeoff curve is procured, the choice of the J value is usually contingent upon the purpose of the application of interest and/or the prerogative of the system designer. With the fixed value of J, the existence of the optimal solution of the GOP (i.e., the desired tradeoff) is assured due to the CAC in place, discussed in Section 7.3.1. Should the value of J be allowed to change on the fly when the system is in use, a conservative approach can be employed: whenever a new call arrives, we only check the feasibility condition on the solution of the WMMFOP (i.e., the GOP with $J = J^*$). The new call is admitted if there exists a solution, or rejected otherwise. Since the feasible region does not shrink and hence the solution obtained initially will not

become infeasible when the value of J decreases, this approach prevents the ongoing calls in service from being dropped. In spite of performance degradation, this conservative yet safe approach is imperative for providing QoS assurance to new calls and ongoing calls already admitted in the system.

7.4 Efficiency Evaluation by Game Theory

In this section, we show that our solutions obtained from the GOP given by (7.21)–(7.22) achieve efficient use of resources. In game theory, efficient resource utilization is determined by the concept of Pareto optimality [88].

Proposition 20 *The optimal solution* $\mathbf{a}$ *of the GOP is Pareto optimal.*

From the perspective of game theory, the resource allocation solutions procured from the GOP are efficiently utilized. For the relationship between system throughput and fairness, every point (i.e., resource allocation) on a tradeoff curve (to be discussed in Section 7.5) is Pareto optimal, thereby making efficient use of scarce network resources.

7.5 Numerical Results

This section presents numerical results on: 1) system throughput and fairness performance versus the value of J in the GOP; and 2) a relationship/tradeoff curve of system throughput and fairness. In our numerical analysis, we solve the GOP by an exhaustive search with an increment size of $\delta a = 0.01$. Suppose that there are I iterations in the iterative procedure for the throughput and fairness tradeoff curve. Let $\mathbf{a_i}$ be the optimal solution obtained from the GOP in the i^{th} iteration. The measure of system throughput in the i^{th} iteration (i.e.,

$i \in I$) is given by

$$U = \frac{\left(\sum_{m=1}^{M} Q_m(\mathbf{a_i})\right) - \min_{i \in I}\left\{\sum_{m=1}^{M} Q_m(\mathbf{a_i})\right\}}{\max_{i \in I}\left\{\sum_{m=1}^{M} Q_m(\mathbf{a_i})\right\} - \min_{i \in I}\left\{\sum_{m=1}^{M} Q_m(\mathbf{a_i})\right\}} \tag{7.23}$$

where $U \in [0, 1]$, i.e., the larger the value of the system throughput, the larger the value of U. Let $D_i = |J^* - \min_m \{w_m Q_m(\mathbf{a_i})\}|$, where J^* is the solution of the WMMFOP and D_i represents the deviation of the minimum value of $w_m Q_m(\mathbf{a_i})$ among all M links from J^*, i.e., the larger the value of D_i, the poorer the weighted max-min fairness performance. The measure of weighted max-min fairness in the i^{th} iteration is given by

$$V = \frac{\max_{i \in I}\{D_i\} - D_i}{\max_{i \in I}\{D_i\} - \min_{i \in I}\{D_i\}} \tag{7.24}$$

where $V \in [0, 1]$, i.e., the larger the value of D_i, the smaller the value of V. V indicates the fairness performance of the worst link. In the literature, Jain's fairness index [47] is widely employed as a measure of network-wise fairness performance. Let JFI_i be the Jain's fairness index in the i^{th} iteration, where

$$JFI_i = \frac{\left(\sum_m w_m Q_m(\mathbf{a_i})\right)^2}{M \sum_m \left(w_m Q_m(\mathbf{a_i})\right)^2}. \tag{7.25}$$

The shaped Jain's fairness index in the i^{th} iteration is given by

$$W = \frac{JFI_i - \min_{i \in I}\{JFI_i\}}{\max_{i \in I}\{JFI_i\} - \min_{i \in I}\{JFI_i\}} \tag{7.26}$$

where $W \in [0, 1]$, i.e., the larger the value of Jain's fairness index, the larger the value of W.

In this numerical analysis, we consider four active links in the network, i.e., $M = 4$. The fading coefficient of a link is modeled as a complex Gaussian random variable with zero mean and unit variance. The channel gain matrix $\mathbf{G}$, i.e., $\mathbf{G} = [G_{mk}]_{M \times M}$, used for the numerical analysis is randomly generated and normalized, which is given below:

$$\mathbf{G} = \begin{bmatrix} 0.2818 & 0.3299 & 0.2739 & 0.0350 \\ 0.2418 & 0.1761 & 0.5019 & 1.0000 \\ 0.1823 & 0.9345 & 0.2802 & 0.0068 \\ 0.2016 & 0.4150 & 0.4480 & 0.0400 \end{bmatrix}. \tag{7.27}$$

Table 7.2: System parameters for the numerical analysis for the case of equal weighting factors

Parameter	Value
M	4
η	0.01
σ	0.1
P_m	1, $\forall m$
w_m	1, $\forall m$
$(R_1^d, ..., R_4^d)$	(2, 1, 0.5, 0.1)
J^*	0.3290

Under the condition that the solution exists in the WMMFOP, we consider two cases: 1) equal weighting factors; and 2) different weighting factors.

7.5.1 Equal Weighting Factors

The system parameters for the case of equal weighting factors are given in Table 7.2, where J^* is computed by solving the WMMFOP. We follow the iterative procedure described in Section 7.3.2 and obtain the numerical results given in Table 7.3. First, we study the impact of the value of J on the system throughput measure U and the fairness measure V. Figure 7.1 depicts system throughput performance versus fairness performance. As mentioned in Section 7.3.2, a tradeoff curve is obtained by solving the GOP iteratively, starting at the maximal system throughput and ending at the maximal fairness. As expected, in Figure 7.1, U decreases from the maximum value (i.e., $U = 1$) to the minimum value (i.e., $U = 0$) with J, while V increases from the minimum value (i.e., $V = 0$) to the maximum value (i.e., $V = 1$) with J, which shows that increasing system throughput and maintaining fairness are conflicting with each other. The shaped Jain's fairness index is also plotted for comparison. From the results shown in Table 7.3, the minimum weighted utility value, $\min_m \{w_m Q_m(\mathbf{a})\}$, increases with J, as expected. It is worth mentioning that the fairness performance measure

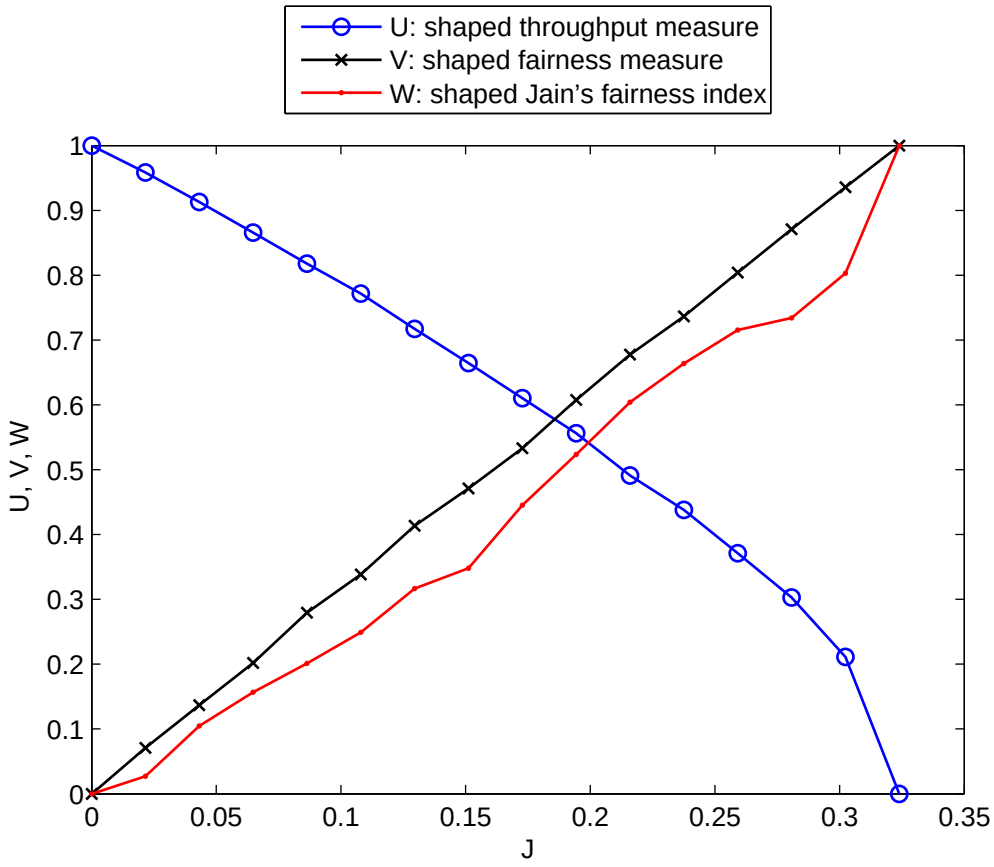


Figure 7.1: The system throughput measure and the fairness measures vs. the value of J with equal weighting factors.

Table 7.3: Numerical results for the case of equal weighting factors

J	$Q_1(\mathbf{a})$	$Q_2(\mathbf{a})$	$Q_3(\mathbf{a})$	$Q_4(\mathbf{a})$	$\sum Q_m(\mathbf{a})$	$\min_m \{w_m Q_m(\mathbf{a})\}$	U	V	W	$\mathbf{a} = [a_1, a_2, a_3, a_4]$
0	0.9115	0.0063	1.3134	0.0001	2.2313	0.0001	1.0000	0	0	[1.00, 0.46, 0.64, 0.14]
0.0219	0.8548	0.0230	1.2921	0.0236	2.1935	0.0230	0.9584	0.0696	0.0272	[1.00, 0.50, 0.66, 0.18]
0.0439	0.8716	0.0478	1.1874	0.0457	2.1525	0.0457	0.9131	0.1385	0.0892	[1.00, 0.52, 0.61, 0.21]
0.0658	0.8809	0.0674	1.0971	0.0672	2.1126	0.0672	0.8693	0.2041	0.1445	[1.00, 0.54, 0.57, 0.24]
0.0877	0.7930	0.0885	1.0959	0.0900	2.0674	0.0885	0.8194	0.2686	0.1792	[1.00, 0.60, 0.61, 0.29]
0.1097	0.8088	0.1097	0.9938	0.1117	2.0240	0.1097	0.7717	0.3331	0.2488	[1.00, 0.62, 0.56, 0.32]
0.1316	0.8997	0.1353	0.8077	0.1321	1.9748	0.1321	0.7175	0.4012	0.3198	[1.00, 0.61, 0.45, 0.33]
0.1535	0.8662	0.1553	0.7520	0.1541	1.9276	0.1541	0.6656	0.4682	0.3786	[1.00, 0.65, 0.44, 0.37]
0.1755	0.7390	0.1775	0.7835	0.1766	1.8766	0.1766	0.6093	0.5367	0.4544	[1.00, 0.74, 0.50, 0.44]
0.1974	0.8167	0.1986	0.6100	0.1983	1.8236	0.1983	0.5509	0.6025	0.5004	[1.00, 0.73, 0.40, 0.45]
0.2193	0.7079	0.2227	0.6189	0.2195	1.7690	0.2195	0.4908	0.6671	0.6038	[1.00, 0.82, 0.44, 0.52]
0.2413	0.6777	0.2417	0.5501	0.2425	1.7120	0.2417	0.4281	0.7345	0.6755	[1.00, 0.87, 0.42, 0.57]
0.2632	0.6693	0.2646	0.4535	0.2643	1.6517	0.2643	0.3616	0.8032	0.7262	[1.00, 0.91, 0.38, 0.61]
0.2851	0.6609	0.2864	0.3569	0.2858	1.5900	0.2858	0.2936	0.8685	0.7433	[1.00, 0.95, 0.34, 0.65]
0.3071	0.5515	0.3081	0.3187	0.3098	1.4881	0.3081	0.1815	0.9365	0.8596	[0.94, 1.00, 0.33, 0.70]
0.3290	0.3312	0.3292	0.3340	0.3290	1.3234	0.3290	0	1.0000	1.0000	[0.78, 1.00, 0.33, 0.71]

of V and that of W are different (i.e., V for the worst-link fairness performance while W for the network-wise fairness performance); however, in the case of equal weighting factors, the general trend of both curves agrees with each other. Thus, both fairness measures match with the max-min fairness performance, as both V and W increase with $\min_m \{w_m Q_m(\mathbf{a})\}$, in general. Consider the trend of each utility function. For every link, its utility value (i.e., $Q_m(\mathbf{a})$ of the m^{th} link) against the value of J is given in Table 7.3. For a small J, the links with smaller effective bandwidths usually obtain larger utility values (e.g., $Q_3(\mathbf{a})$). It is intuitive that those links with smaller effective bandwidths are more flexible to increase their throughputs than other links with larger effective bandwidths. As J increases, the utility values of those links with smaller effective bandwidths decrease for the sake of fairness.

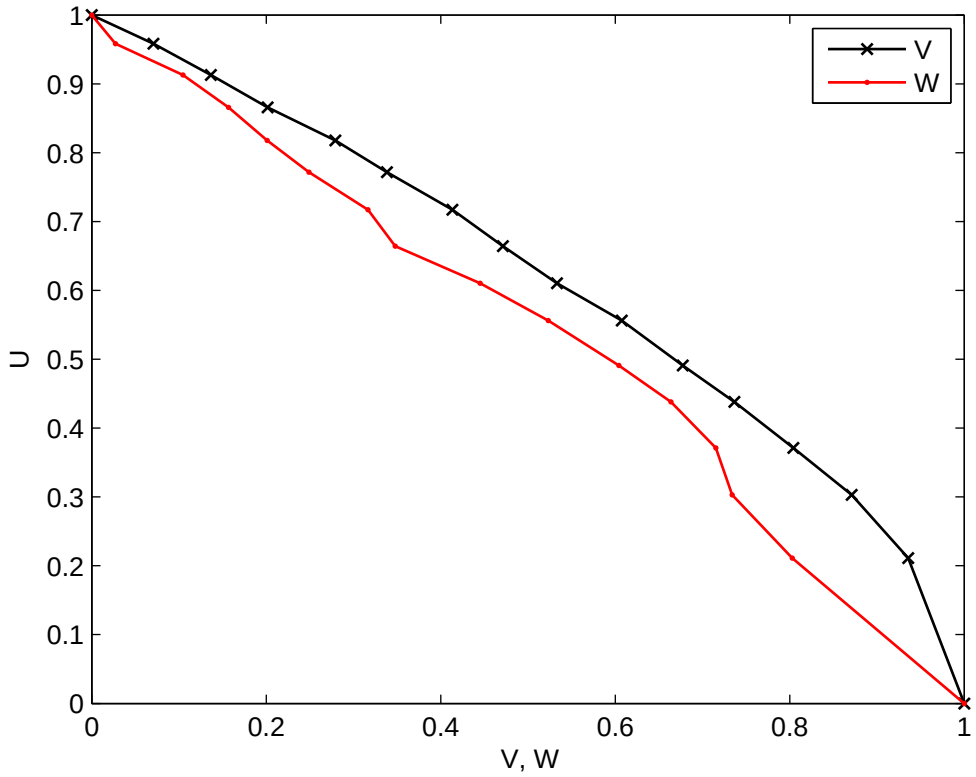


Figure 7.2: The relationship of system throughput and fairness with equal weighting factors.

However, with different channel gains, some link, say the m^{th} link, with a small effective bandwidth may be forced to use a small value of a_m so that only a small value of $Q_m(\mathbf{a})$ is achieved, for example, $Q_4(\mathbf{a})$ in our example. From (7.27), $G_{24} = 1.0$, meaning that the interference impact from the 4^{th} link to the 2^{nd} link is significant. In order to meet all the effective bandwidth requirements, the 4^{th} link can only use a small value of a_4, which results in a small value of $Q_4(\mathbf{a})$. Nonetheless, the utility values of all links converge to the same value when $J = J^*$ (i.e., maximal fairness). Notice that the performance discrepancies in Table 7.3 are merely due to the discrete exhaustive search used in the numerical analysis. The desired tradeoff curve of system throughput and fairness performances is shown in Figure 7.2. The curve is a bit concave in shape, meaning that in a nearly unfair situation

Table 7.4: System parameters for the numerical analysis for the case of different weighting factors

Parameter	Value
M	4
η	0.01
σ	0.1
P_m	1, $\forall m$
$(w_1, ..., w_4)$	(1, 2, 4, 8)
$(R_1^d, ..., R_4^d)$	(2, 1, 0.5, 0.1)
J^*	0.9249

(i.e., $V \approx 0$), a unit decrease in system throughput gives a larger marginal improvement in weighted max-min fairness performance. At a near-maximal fairness point (i.e., $V \approx 1$), a larger decrease of system throughput is required to further increase the fairness measure. From this curve, different degrees of performance tradeoff between system throughput and fairness can be acquired by suitably choosing the value of J. The shaped Jain's fairness index is also plotted for reference. This tradeoff curve is undoubtedly useful for effective and efficient resource allocation. With application-specific constraints (such as fairness or throughput requirements), a desired tradeoff point can be obtained from this relationship curve and hence the corresponding resource allocation $\mathbf{a}$ can be deduced.

7.5.2 Different Weighting Factors

For the case of different weighting factors, the system parameters used for the numerical analysis are given in Table 7.4, and the numerical results are given in Table 7.5. In Figure 7.3, the curve of the system throughput measure U and that of the fairness measure V with different values of J are plotted, and the tradeoff curve of system throughput and fairness performances is shown in Figure 7.4. For U and V, the trends of these two curves in Figure 7.3 and Figure 7.4 are more or less the same as those shown in Figure 7.1 and Figure 7.2,

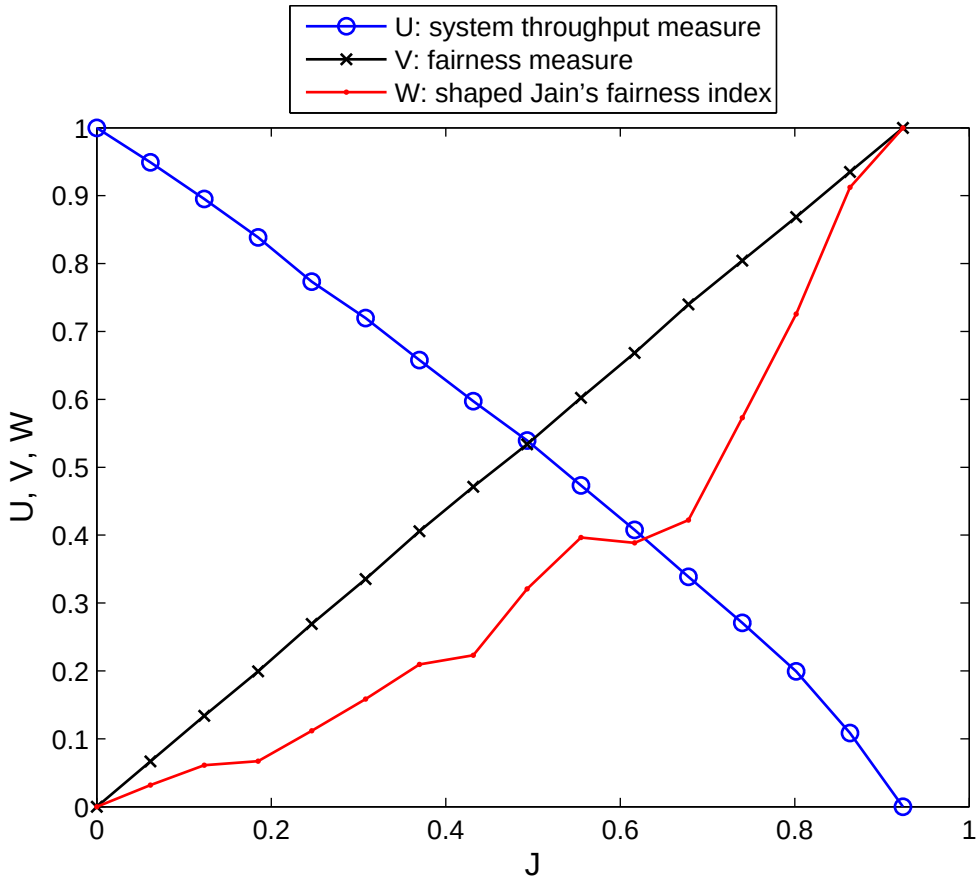


Figure 7.3: The system throughput measure and the fairness measures vs. the value of J with different weighting factors.

Table 7.5: Numerical results for the case of different weighting factors

J	$Q_1(\mathbf{a})$	$Q_2(\mathbf{a})$	$Q_3(\mathbf{a})$	$Q_4(\mathbf{a})$	$\sum Q_m(\mathbf{a})$	$\min_m \{w_m Q_m(\mathbf{a})\}$	U	V	W	$\mathbf{a} = [a_1, a_2, a_3, a_4]$
0	0.9115	0.0063	1.3134	0.0001	2.2313	0.0010	1.0000	0	0	[1.00, 0.46, 0.64, 0.14]
0.0617	0.9328	0.0313	1.2341	0.0089	2.2071	0.0625	0.9492	0.0666	0.0318	[1.00, 0.47, 0.60, 0.15]
0.1233	0.9285	0.0623	1.1746	0.0159	2.1813	0.1246	0.8952	0.1338	0.0610	[1.00, 0.49, 0.58, 0.16]
0.1850	0.8412	0.0941	1.1959	0.0232	2.1544	0.1853	0.8385	0.1994	0.0670	[1.00, 0.54, 0.63, 0.18]
0.2466	0.8522	0.1317	1.1082	0.0312	2.1233	0.2496	0.7733	0.2690	0.1117	[1.00, 0.56, 0.59, 0.19]
0.3083	0.8724	0.1553	1.0300	0.0401	2.0978	0.3106	0.7198	0.3351	0.1585	[1.00, 0.57, 0.55, 0.20]
0.3700	0.8759	0.1880	0.9567	0.0477	2.0683	0.3759	0.6579	0.4058	0.2096	[1.00, 0.59, 0.52, 0.21]
0.4316	0.8069	0.2199	0.9581	0.0546	2.0395	0.4364	0.5974	0.4713	0.2231	[1.00, 0.64, 0.55, 0.23]
0.4933	0.8754	0.2471	0.8272	0.0621	2.0118	0.4941	0.5394	0.5337	0.3208	[1.00, 0.63, 0.47, 0.23]
0.5550	0.8789	0.2786	0.7529	0.0698	1.9802	0.5573	0.4733	0.6021	0.3965	[1.00, 0.65, 0.44, 0.24]
0.6166	0.7774	0.3093	0.7837	0.0785	1.9489	0.6186	0.4076	0.6685	0.3887	[1.00, 0.72, 0.49, 0.27]
0.6783	0.7276	0.3426	0.7604	0.0856	1.9162	0.6844	0.3386	0.7397	0.4223	[1.00, 0.77, 0.50, 0.29]
0.7399	0.7910	0.3720	0.6265	0.0942	1.8837	0.7441	0.2708	0.8043	0.5729	[1.00, 0.76, 0.42, 0.29]
0.8016	0.8345	0.4045	0.5103	0.1004	1.8497	0.8035	0.1993	0.8686	0.7255	[1.00, 0.76, 0.36, 0.29]
0.8633	0.8969	0.4324	0.3678	0.1094	1.8065	0.8649	0.1086	0.9350	0.9124	[1.00, 0.75, 0.29, 0.29]
0.9249	0.9260	0.4625	0.2452	0.1211	1.7548	0.9249	0	1.0000	1.0000	[1.00, 0.76, 0.24, 0.30]

respectively. The curve of W, however, deviates more away from that of V in the case of different weighting factors. In fact, using W in this case is less effective to accurately indicate the improvement of weighted max-min fairness performance though the minimum weighted utility value, $\min_m \{w_m Q_m(\mathbf{a})\}$, increases with J, shown in Table 7.5. Nonetheless, the Jain's fairness index gauges network-wise fairness performance, instead of measuring the worst-link fairness performance, even though the general trend of the curve roughly matches the weighted max-min fairness performance. With different weighting factors, the allocation solution $\mathbf{a}$ in the GOP is not the same as that with the equal weighting factor. As mentioned, the smaller the value of w_m, the more eager is the m^{th} link to obtain more extra resources. At the last row of Table 7.5, the value of $Q_1(\mathbf{a})$ is the largest while that of $Q_4(\mathbf{a})$ is the smallest,

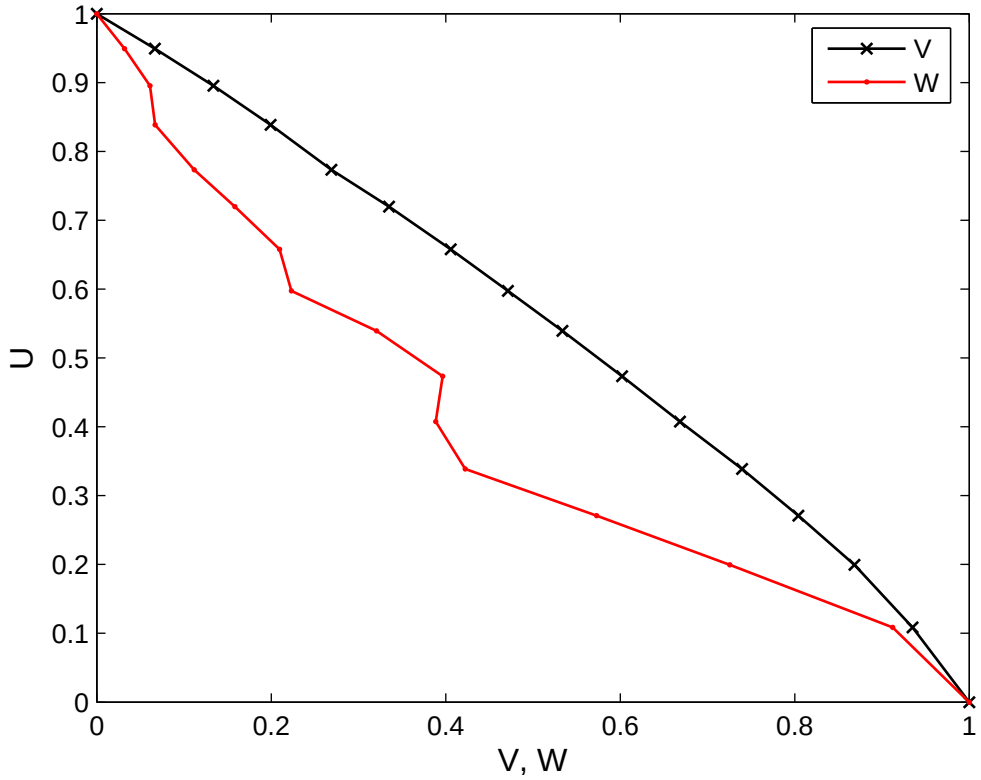


Figure 7.4: The relationship of system throughput and fairness with different weighting factors.

as $w_1 < w_2 < w_3 < w_4$. Therefore, different traffic classes can be differentiated with different weighting factors. As such, the use of w_m is crucial for effective service differentiation and QoS support in WMNs supporting heterogeneous traffic (e.g., voice, video, and data traffic).

7.6 Discussion

In this section, some issues on practical implementation are discussed. For WMNs with centralized control, a central controller (such as a base station) essentially collects requests from wireless nodes, makes decisions on call admission, and allocates resources to wireless links. The global network information (i.e., network resources, QoS requirements of calls, and channel conditions) are available at the central controller, thereby leading to an easier practical implementation to obtain (near-)optimal resource allocation by solving the GOP. In contrast, for WMNs with distributed control (without central coordination), acquiring global network information by message exchanges is not desired, causing a considerable amount of overhead. Thus, each node usually has its local network information only. As presented in Chapter 3, node clustering is a viable approach, where the whole network is divided into clusters. Within each cluster, a clusterhead merely deals with the network activity of its own neighborhood [119]. Each clusterhead can run the GOP locally by treating the inter-cluster interference as a portion of the intra-cluster interference.

Whether a resource allocation solution is optimal or suboptimal depends on the algorithm design. Although our framework is universal in the sense that it can be applied to any interference-limited system model, one drawback is that solving this optimization problem is mostly computationally expensive. Some (suboptimal) algorithms with low complexity are preferred for the sake of practical implementation. To tackle a non-trivial optimization problem such as the GOP, two approaches are widely used:

1) *Convex approximation* – An original optimization problem is relaxed or approximated

to a convex optimization problem. The optimal solution of the relaxed problem can be achieved by classical methods such as gradient methods and interior-point methods [9]. In [26], it is proved that under a high SINR approximation, the objective function of the GOP can be formulated to be a concave function (by verifying that the Hessian matrix is negative definite) and, therefore, the GOP becomes a convex optimization problem. A rich body of literature exists on the theory of convex optimization whereby fast, simple, and robust practical algorithms can be devised (e.g., gradient-based iterations) [9,26]; and

2) *Interpreting KKT conditions* – The necessary conditions for the optimal solution can be verified the by KKT conditions. Optimal and/or heuristic algorithms can be deduced based on the interpretations of the KKT conditions [9,42]. To handle a large amount of decision variables, data structure plays an important role in determining the complexity. For instance, tree implementation together with sorting and/or searching can facilitate to bring lower computational complexity to the algorithm [112].

In this Chapter, the goal is to develop a unified framework to balance throughput and fairness. Nonetheless, the actual system performance with practical implementation needs further investigation. Devising algorithms of low computational complexity is imperative; addressing this issue, however, is left for further investigation.

7.7 Chapter Summary

In this Chapter, we propose a unified optimization framework for interference-limited WMNs, whereby the optimal relationship curve of system throughput and fairness can be procured. Different degrees of performance tradeoff between system throughput and fairness can be attained, by suitably adjusting the value of our introduced bargaining floor. QoS support

is assured with the help of the CAC based on the feasibility of the solution of the GOP. Further, our resource allocation solutions achieve Pareto optimality, making efficient use of network resources.

Chapter 8

Conclusions and Further Research

8.1 Conclusions

The success of deploying WMNs to suburban and rural residential areas is highly contingent on the effectiveness of how the network is managed and how the radio resources are allocated. With the increasing demand of multimedia applications (e.g., voice, video, and data services), QoS provisioning and fairness assurance are imperative in future broadband wireless access. In this dissertation, we have proposed several novel and effective resource allocation approaches for WMNs with QoS support and service differentiation. Our results have showed that our proposed approaches are effective in packet-level QoS provisioning and system performance improvement. Performance comparison has also been carried out to demonstrate the merits of the proposed resource allocation approaches over their conventional counterparts suggested in the literature. Further, our resource allocation strategies are of low computational complexity, conducing to viable and preferred candidates for practical implementation. The accomplishments in this dissertation are summarized as follows:

- In Chapter 3, we address the joint problem of node clustering and subcarrier allo-

cation in order to foster frequency reuse and increase system capacity in WMNs for suburban/rural residential areas. We propose a novel node clustering algorithm with effective subcarrier allocation for WMNs. The proposed node clustering algorithm is QoS-aware, and the subcarrier allocation is optimality-driven and can be performed in a decentralized manner. Simulation results show that our novel tax-based subcarrier allocation approach outperforms a conventional interference conflict-graph approach in terms of frequency reuse ratio and system throughput. By carefully adjusting an upper bound of subcarriers allocated to each cluster, we can attain an improved system performance. Further, the proposed resource allocation achieves Pareto optimality, demonstrating efficient use of network resources. Our findings also reveal that how to allocate resources in a decentralized fashion can affect the solution space of a performance tradeoff between QoS provisioning and throughput maximization;

- In Chapter 4, we address the joint power-frequency-time resource allocation problem for WMNs supporting high-speed multimedia applications. We propose three QoS-aware resource allocation approaches, namely 1) a KKT-driven approach, 2) a GA-based approach, and 3) a hybrid KKT-GA approach. We compare our proposed approaches with some existing resource allocation counterparts. It is shown that all of the proposed approaches are effective in provisioning packet-level QoS and improving system throughput. In particular, our proposed hybrid KKT-GA resource allocation approach is demonstrated promising in achieving a desired balance between system performance and computational complexity. Our findings reveal that QoS provisioning and throughput maximization are two conflicting goals, where admitting more multimedia calls can reduce the system throughput;

- To further enhance the system performance, in Chapter 5, we study the problem of non-altruistic node cooperative resource allocation for WMNs with service differentiation. We propose two low-complexity node cooperative resource allocation approaches, tak-

ing QoS support, subcarrier allocation, power allocation, partner selection/allocation, and packet scheduling into account. Simulation results show that both proposed approaches are effective in packet-level QoS provisioning and system performance enhancement over their non-cooperative counterparts. Further, our study reveals a critical principle that whether node cooperation is beneficial depends upon the nature of node cooperation, the mode of network operation, and the traffic pattern;

- In Chapter 6, by employing optimal stopping, we propose a simple channel sensing order for secondary users in cognitive radio networks to alleviate the problem of low resource utilization. By sensing the channels according to the descending order of their achievable rates, we prove that a secondary user should stop at the first sensed free channel. We can further prove that the proposed instinctive channel sensing order is optimal in terms of throughput when the primary-free probabilities are equal. Besides, the probability of collision in a multi-secondary user scenario with respect to the proposed instinctive sensing is derived and further validated by extensive computer simulations. Our results show that, when the number of secondary users approaches the number of channels, resource utilization can be further improved at the cost of a higher collision probability;

- With limited available radio resources, balancing system throughput and fairness for WMNs with QoS support and high resource efficiency is vital. In Chapter 7, we derive a unified optimization framework so as to procure the optimal relationship (i.e., tradeoff curve) of system throughput and fairness with QoS support. A relationship curve between throughput and fairness can be obtained by solving the optimization problem iteratively, whereby different degrees of performance tradeoff can be acquired. Our numerical study indicates that the tradeoff curve of system throughput against fairness is concave in shape, meaning that a unit decrease in system throughput leads to a larger marginal improvement in fairness performance.

8.2 Further Research

In this dissertation, we have addressed some key research issues related to resource management tailored for WMNs. Not only our results demonstrate the effectiveness of our proposed resource management approaches, but our studies also reveal other important research directions to further improving the system performance of WMNs. Despite its accomplishments, this dissertation is only the necessary first step to understand the issues of radio resource management for WMNs supporting heterogeneous traffic. More research problems need to be addressed and solved before WMNs can be properly designed and successfully deployed to suburban and rural residential areas in practice. Further research should address the following challenging yet important research problems for WMNs:

- *Gateway Deployment in WMNs* – Employing multiple gateways (attached to the Internet backbone) is one of the effective ways to alleviate the problem of traffic congestion in WMNs [66]. With a single gateway only, traffic at and near the gateway is expected to be heavy, posing the issue of a single-point-of-failure. In the presence of multiple gateways, traffic load can be balanced more effectively and efficiently, thereby facilitating traffic routing, packet scheduling, and QoS provisioning. On the other hand, with better traffic distribution, co-channel interference can be reduced to a greater extent. Together with directional antennas, frequency reuse can be greatly fostered and, therefore, system capacity can be further increased. However, how to deploy such gateways in WMNs in order to achieve a desired system performance remains an open research problem;

- *Joint Routing and MAC-Layer Resource Allocation for WMNs* – The issues of routing and MAC-layer resource allocation are coupled, and studying these two issues separately generally results in suboptimal system performance [18]. In WMNs supporting heterogeneous traffic, route selection is necessary so as to guarantee delay and rate

requirements of each incoming call and the ongoing sessions already admitted to the system [17]. Recently, routing via network coding has been demonstrated promising in terms of packet delivery rate [111]. Incorporating the notion of network coding when designing a resource allocation scheme is certainly of great interest. However, obtaining globally optimal solutions for the joint problem of (network coding-based) routing and MAC-layer resource allocation usually requires exponential time complexity. Low-complexity approaches for the joint network coding-based routing and packet-level resource allocation problem with QoS assurance are strongly desired. Since end-to-end QoS support and CAC are closely related, call-level resource allocation and packet-level resource allocation should also be jointly considered;

- *QoS Provisioning in Cognitive Radio Mesh Networks* – In cognitive radio mesh networks (CRMNs), secondary users can access the spectrum only if there is no primary activity. Therefore, resource reservation for secondary users is hardly feasible, as channel availabilities (or primary activities) vary from time to time. In CRMNs with broadband wireless access, we expect that multimedia applications are widely supported for both licensed primary users and unlicensed secondary users. As a result, to effectively support their QoS, *soft* bandwidth reservation for secondary users is indispensable. In the presence of hidden terminals, effective MAC is also crucial so as to reduce the probability of collision. How to design a simple yet low-temperature MAC protocol tailored for CRMNs is an open research problem.

Appendix A

Supplementary

A.1 Tax Interpretation of the KKT Conditions in Chapter 3

Consider the following throughput maximization problem

$$\max_{\mathbf{c},\mathbf{p}} \left\{ \sum_{m} \sum_{n} \sum_{l} c_{m,n}^{l} r_{m,n}^{l} \right\} \tag{A-1}$$

$$\text{subject to } p_{m,n}^{l} \geq 0, \forall m, n, l \tag{A-2}$$

$$\sum_{n} p_{m,n}^{l} \leq P_{m}^{\max}, \forall m, l \tag{A-3}$$

$$c_{m,n}^{l} \in \{0, 1\}, \forall m, n, l. \tag{A-4}$$

By relaxing the constraint (A-4), i.e., $0 \leq c_{m,n}^{l} \leq 1, \forall m,n,l$, consider part of the KKT conditions of the relaxed problem with respect to $p_{m,n}^{l}$

$$\frac{\partial}{\partial p_{m,n}^{l}}(\sum_{m}\sum_{n}\sum_{l} c_{m,n}^{l} r_{m,n}^{l}) = \mu_{m}^{l} - \alpha_{m,n}^{l}, \forall m,n,l \tag{A-5}$$

$$\alpha_{m,n}^{l} p_{m,n}^{l} = 0, \forall m,n,l \tag{A-6}$$

$$\mu_{m}^{l}\left(\sum_{n} p_{m,n}^{l} - P_{m}^{\max}\right) = 0, \forall m,l \tag{A-7}$$

where $\alpha_{m,n}^{l}(\geq 0)$ and $\mu_{m}^{l}(\geq 0)$ are the Lagrange multipliers for the constraints (A-2) and (A-3), respectively. Condition (A-5) can be written as (A-8) and (A-9).

$$\frac{\partial}{\partial p_{m,n}^{l}}(\sum_{n}\sum_{l} c_{m,n}^{l} r_{m,n}^{l}) + \sum_{k\neq m}\frac{\partial}{\partial p_{m,n}^{l}}(\sum_{n}\sum_{l} c_{k,n}^{l} r_{k,n}^{l}) = \mu_{m}^{l} - \alpha_{m,n}^{l}, \forall m,n,l \tag{A-8}$$

$$\Rightarrow \quad c_{m,n}^{l}\frac{\partial r_{m,n}^{l}}{\partial p_{m,n}^{l}} + \sum_{k\neq m}\frac{\partial}{\partial p_{m,n}^{l}}(\sum_{n}\sum_{l} c_{k,n}^{l} r_{k,n}^{l} - R_{k}^{d}) = \mu_{m}^{l} - \alpha_{m,n}^{l}, \forall m,n,l. \tag{A-9}$$

Assuming $\sum_{n}\sum_{l} c_{k,n}^{l} r_{k,n}^{l} - R_{k}^{d} \geq \epsilon$, condition (A-9) can be written as

$$c_{m,n}^{l}\frac{\partial r_{m,n}^{l}}{\partial p_{m,n}^{l}} + \sum_{k\neq m}\frac{\partial}{\partial p_{m,n}^{l}}(U_{k}(\mathbf{c},\mathbf{p},R_{k}^{d})) = \mu_{m}^{l} - \alpha_{m,n}^{l}, \forall m,n,l. \tag{A-10}$$

Let $T_{km,n}^{l} = -\frac{\partial U_{k}(\mathbf{c},\mathbf{p},R_{k}^{d})}{\partial p_{m,n}^{l}}$. Viewing $T_{km,n}^{l}$ as tax paid by the m^{th} clusterhead for generating interference to the k^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot, condition (A-10) is one of the (necessary) conditions for optimality for the relaxed problem. In other words, each clusterhead optimizes its power allocation and subcarrier allocation to maximize its payoff function, defined as

$$S_{m,n}^{l}\left(c_{m,n}^{l}, p_{m,n}^{l}\right) = c_{m,n}^{l} r_{m,n}^{l} - p_{m,n}^{l}\sum_{k\neq m}^{M} T_{km,n}^{l}, \forall m,n,l. \tag{A-11}$$

As a result, for each subcarrier and each timeslot, each clusterhead is to maximize the difference between its throughput obtained minus its lump-sum tax paid to the other clusterheads in the mesh backbone due to the induced interference.

A.2 Upper Bound for Throughput in Chapter 4

Under the assumptions of no packet dropping for the real-time traffic and perfect statistical traffic multiplexing, we derive the upper bound for the throughputs obtained for all the traffic. Denote ρ_1 and ρ_2 as the voice activity factor and the video activity factor, respectively. The upper bound for the throughput obtained for each of the traffic is given as follows.

- Throughput obtained for voice traffic, T_{voice}:

$$T_{voice} = \sum_{i=0}^{N_{voice}} \binom{N_{voice}}{i} i\rho_1^i (1-\rho_1)^{N_{voice}-i} \times R_{voice} \times 5ms/20ms \tag{A-12}$$

$$= \rho_1 N_{voice} R_{voice}/4 \tag{A-13}$$

  where N_{voice} and R_{voice} are the number of voice sources in the system and the constant data rate of voice traffic, respectively.

- Throughput obtained for video traffic, T_{video} :

$$T_{video} = \sum_{i=0}^{N_{video}} \binom{N_{video}}{i} i\rho_2^i (1-\rho_2)^{N_{video}-i} \times R_{video} \times 5ms/5ms \tag{A-14}$$

$$= \rho_2 N_{video} R_{video} \tag{A-15}$$

  where N_{video} and R_{video} are the number of video sources in the system and the mean data rate of video traffic, respectively.

- Throughput obtained for background data traffic, T_{data_1}:

$$T_{data_1} = \sum_m R_m - T_{voice} - T_{video} \tag{A-16}$$

  where R_m is the achieved data rate for the m^{th} link. Equation (A-16) can be served as an upper bound for the throughput obtained by background data traffic.

- Throughput obtained for bursty data traffic, T_{data_2}:

$$T_{data_2} = \lambda N_{data} \sqrt{\pi} \tag{A-17}$$

where λ and N_{data} are the packet arrival rate and the number of data sources in the system, (i.e., $\lambda = 50$ packets/second), respectively. Note that the mean of a Weibull distribution (i.e., $Weibull(u, k)$) is $u\Gamma(1 + 1/k)$, where Γ is the Gamma function, and $\Gamma(1 + 1/2) = \sqrt{\pi}/2$.

A.3 Details on Non-Altruistic Node Cooperation in Chapter 5

Consider a wireless network consisting of three nodes, namely Node S, Node R, and Node D. All three nodes are equipped with a single antenna. Node S is to transmit data to Node D, while Node R is viewed as a relay to help Node S forward the data to Node D. We employ the Cooperation Protocol I suggested in [82] as our node cooperation strategy. In the first symbol interval, Node S transmits a symbol to both Node R and Node D. In the second symbol interval, Node R forwards the symbol received from Node S to Node D, while Node S transmits another symbol to Node D. In this work, we assume that Node D has complete knowledge of channel state information (CSI). After receiving the symbols in the two intervals, Node D then employs the maximum likelihood (ML) detection to decode the symbols originally transmitted from Node S [104]. Notice that Node R can choose the decode-and-forward (DF) mode or amplify-and-forward (AF) mode of cooperation in the second symbol interval. In this work, although we consider the DF mode of cooperation (i.e., regenerative nodes), we examine the outage performances of both the AF and DF modes for the sake of completeness.

Let x_1 and x_2 be the first symbol and the second symbol transmitted by Node S, respectively. We assume that the energy of a symbol is one and the mean of the value of a symbol is zero, i.e., $E[|x_i|^2] = 1$ and $E[x_i] = 0$, where $i = 1, 2$. Denote $y_{D,1}$ as the

signal received at Node D during the first symbol interval, $y_{R,1}$ the signal received at Node R during the first symbol interval, and $y_{D,2}$ the signal received at Node D during the second symbol interval.

A.3.1 AF Mode of Cooperation

Consider a frequency-flat slow Rayleigh fading environment where the carrier phase distortion due to the fading channel can be estimated at the receiver and removed. For the AF mode of cooperation, we have

$$y_{D,1} = \sqrt{E_{SD}} h_{SD} x_1 + \eta_{D,1} \tag{A-18}$$

$$y_{R,1} = \sqrt{E_{SR}} h_{SR} x_1 + \eta_{R,1} \tag{A-19}$$

$$y_{D,2} = \sqrt{E_{SD}} h_{SD} x_2 + \sqrt{E_{RD}} h_{RD} \frac{y_{R,1}}{\sqrt{E[|y_{R,1}|^2]}} + \eta_{D,2} \tag{A-20}$$

where $E_{XY} (\geq 0)$ is the average energy of a symbol received at Node Y from Node X, i.e., $X,\ Y \in \{S,\ R,\ D\}$, h_{XY} is the Rayleigh fading coefficient for the $X \rightarrow Y$ link modeled as an independent zero-mean complex Gaussian random variable with unit variance, and $\eta_{D,1}$, $\eta_{R,1}$, and $\eta_{D,2}$ are independent zero-mean complex Gaussian random variables with variance $\sigma^2/2$ per dimension. Here, we make a reasonable assumption that the fading channels are quasi-static over a period of two symbol intervals. Notice that the introduction of a factor of $\sqrt{E[|y_{R,1}|^2]}$ in (A-20) is to normalize the received signal at Node R during the first symbol interval so that the average energy of a symbol is one, where $\sqrt{E[|y_{R,1}|^2]} = E_{SR} + \sigma^2$. Thus, the received signals can be represented in a matrix form given by

$$\mathbf{y}_{AF} = \mathbf{H}_{AF} \mathbf{x} + \mathbf{Q}_{AF} \tag{A-21}$$

where $\mathbf{x} = [x_1 \ x_2]$, $\mathbf{y}_{AF} = [y_{D,1} \ y_{D,2}/w]$, $\mathbf{Q}_{AF} = [\eta_{D,1} \ \tilde{\eta}/w]$, and $\mathbf{H}_{AF}$ is the channel matrix given by

$$\mathbf{H}_{AF} = \begin{bmatrix} \sqrt{E_{SD}}h_{SD} & 0 \\ \frac{1}{w}\sqrt{\frac{\gamma E_{SR}E_{RD}}{\gamma E_{SR}+1}}h_{SR}h_{RD} & \frac{1}{w}\sqrt{E_{SD}}h_{SD} \end{bmatrix} \tag{A-22}$$

with $\gamma = 1/\sigma^2$, $\tilde{\eta} = \eta_{D,2} + \sqrt{\frac{E_{RD}}{E_{SR}+\sigma^2}}h_{RD}\eta_{R,1}$, and $w = \sqrt{(E_{SR}+E_{RD}+\sigma^2)/(E_{SR}+\sigma^2)}$. Notice that a factor of $1/w$ is multiplied to (A-20) to ensure the variance of the noise to be the same as that of $\eta_{D,1}$ [82].

A.3.2 DF Mode of Cooperation

For the DF mode of cooperation, we have

$$y_{D,1} = \sqrt{E_{SD}}h_{SD}x_1 + \eta_{D,1} \tag{A-23}$$

$$y_{R,1} = \sqrt{E_{SR}}h_{SR}x_1 + \eta_{R,1} \tag{A-24}$$

$$y_{D,2} = \sqrt{E_{SD}}h_{SD}x_2 + \sqrt{E_{RD}}h_{RD}x_1 + \eta_{D,2}. \tag{A-25}$$

In (A-25), we assume that Node R can decode the received symbol x_1 successfully, yet this assumption will be lifted when we perform the analysis of outage probability, to be discussed in Section A.3.3. The received signals can be represented in a matrix form given by

$$\mathbf{y}_{DF} = \mathbf{H}_{DF}\mathbf{x} + \mathbf{Q}_{DF} \tag{A-26}$$

where $\mathbf{y}_{DF} = [y_{D,1} \ y_{D,2}]$, $\mathbf{Q}_{DF} = [\eta_{D,1} \ \eta_{D,2}]$, and $\mathbf{H}_{DF}$ is the channel matrix given by

$$\mathbf{H}_{DF} = \begin{bmatrix} \sqrt{E_{SD}}h_{SD} & 0 \\ \sqrt{E_{RD}}h_{RD} & \sqrt{E_{SD}}h_{SD} \end{bmatrix}. \tag{A-27}$$

A.3.3 Outage Performance

Here, we study the outage probabilities of both the AF and DF modes of cooperation. Let C_{AF}, C_{DF}, and R_{tar} be the channel capacity achieved by the AF mode, the channel capacity achieved by the DF mode, and a target transmission rate, respectively. An outage event occurs when the channel capacity is smaller than the target transmission rate, e.g., $C_{AF} < R_{tar}$.

The channel capacity achieved by the AF mode is given by

$$\begin{aligned} C_{AF} &= \frac{1}{2}\log_2 \det\left(\mathbf{I_2} + \gamma \mathbf{H}_{AF}\mathbf{H}_{AF}^*\right) \\ &= \frac{1}{2}\log_2\left((1+\gamma\alpha_{SD})(1+\frac{1}{w^2}\gamma\alpha_{SD}) + \frac{1}{w^2}\left(\frac{\gamma^2}{\gamma E_{SR}+1}\right)\alpha_{SR}\alpha_{RD}\right) \end{aligned} \tag{A-28}$$

where $\alpha_{SR} = E_{SR}|h_{SR}|^2$, $\alpha_{SD} = E_{SD}|h_{SD}|^2$, and $\alpha_{RD} = E_{RD}|h_{RD}|^2$. Notice that α_{SR}, α_{SD}, and α_{RD} are exponential random variables. On the other hand, assuming that Node R can perfectly decode the symbols transmitted from Node S, the channel capacity achieved by the DF mode, denoted by $C_{\widetilde{DF}}$, is given by

$$\begin{aligned} C_{\widetilde{DF}} &= \frac{1}{2}\log_2 \det\left(\mathbf{I_2} + \gamma \mathbf{H}_{DF}\mathbf{H}_{DF}^*\right) \\ &= \frac{1}{2}\log_2\left((1+\gamma\alpha_{SD})^2 + \gamma\alpha_{RD}\right). \end{aligned} \tag{A-29}$$

Proposition 21 *In a three-node wireless network employing the cooperation protocol of interest, the AF mode of cooperation achieves the diversity order of two; however, the DF mode of cooperation can achieve the diversity order of two only if the decoding at the relay is perfect.*

Proof: Denote $P_{AF}^{out}(R_{tar})$ as the outage probability of the AF cooperation mode at a

target rate R_{tar}, which is given by

$$\begin{aligned} P_{AF}^{out}(R_{tar}) &= P(C_{AF} < R_{tar}) \\ &\leq P\left(\frac{1}{2}\log_2\left(1+\varepsilon^{\min}\left(\alpha_{SD}+\alpha_{SR}\alpha_{RD}\right)\right) < R_{tar}\right) \\ &= P\left(\alpha_{SD}+\alpha_{SR}\alpha_{RD} < \frac{2^{2R_{tar}}-1}{\varepsilon^{\min}}\right) \end{aligned} \tag{A-30}$$

where $\varepsilon^{\min} = \min\left\{\left(1+\frac{1}{w^2}\right)\gamma, \frac{1}{w^2}\left(\frac{\gamma^2}{\gamma E_{SR}+1}\right)\right\}$. Let $\delta = \frac{2^{2R_{tar}}-1}{\varepsilon^{\min}}$. Let λ_{SR}, λ_{SD}, and λ_{RD} be the parameters of the exponential random variables α_{SR}, α_{SD}, and α_{RD}, respectively. Equation (A-30) becomes

$$\begin{aligned} &\int_0^{\delta} P\left(\alpha_{SR}\alpha_{RD} < \delta - x\right)\lambda_{SD}e^{-\lambda_{SD}x}dx \\ \leq &\int_0^{\delta} P\left(\frac{\alpha_{SR}\alpha_{RD}}{\alpha_{SR}+\alpha_{RD}+1} < \delta - x\right)\lambda_{SD}e^{-\lambda_{SD}x}dx \\ = &\delta^2\int_0^1 \left[\frac{P\left(\frac{\alpha_{SR}\alpha_{RD}}{\alpha_{SR}+\alpha_{RD}+1} < \delta x'\right)}{\delta x'}\right] x'\lambda_{SD}e^{-\lambda_{SD}\delta(1-x')}dx' \end{aligned} \tag{A-31}$$

where $x' = 1 - x/\delta$. Let $h(\delta) = \delta x'$. As $\gamma \to \infty$, $\delta \to 0$ and $h(\delta) \to 0$. Applying the results obtained in [64], we have

$$\lim_{\gamma\to\infty}\frac{1}{h(\delta)}P\left(\frac{\alpha_{SR}\alpha_{RD}}{\alpha_{SR}+\alpha_{RD}+1} < h(\delta)\right) = \lambda_{SR}+\lambda_{RD}.$$

Therefore, in a high SNR regime, (A-31) can be re-written as

$$\delta^2\int_0^1 \left(\lambda_{SR}+\lambda_{RD}\right)x'\lambda_{SD}dx' = \frac{\lambda_{SD}(\lambda_{SR}+\lambda_{RD})}{2}\delta^2.$$

Then, the outage probability of the AF cooperation mode at a target rate R_{tar} is upper-bounded according to

$$P_{AF}^{out}(R_{tar}) \leq \frac{\lambda_{SD}(\lambda_{SR}+\lambda_{RD})}{2}\delta^2. \tag{A-32}$$

From (A-32), it is clear that the diversity order of two can be attained in the AF cooperation mode [104].

For the case of the DF cooperation mode, denote $P_{DF}^{out}(R_{tar})$ as the outage probability at a target rate R_{tar}, given by

$$P_{DF}^{out}(R_{tar}) = P(C_{DF} < R_{tar})$$

where [64]

$$C_{DF} = \min\left\{\frac{1}{2}\log_2\left(1+\gamma\alpha_{SR}\right), C_{\widetilde{DF}}\right\}. \tag{A-33}$$

In (A-33), the first term refers to the maximum transmission rate at which Node R can decode the symbols sent from Node S successfully, whereas the second term is given by (A-29). Thus, the outage probability of the DF mode is given by

$$\begin{aligned} P_{DF}^{out}(R_{tar}) &= P(C_{DF} < R_{tar}) \\ &= P\left(\min\left\{\alpha_{SR}, 2\alpha_{SD}+\alpha_{RD}+\gamma\alpha_{SD}^2\right\} < \frac{2^{2R_{tar}}-1}{\gamma}\right) \\ &\leq P\left(\min\left\{\alpha_{SR}, 2\alpha_{SD}+\alpha_{RD}\right\} < \frac{2^{2R_{tar}}-1}{\gamma}\right). \end{aligned} \tag{A-34}$$

Let $\delta = \frac{2^{2R_{tar}}-1}{\gamma}$, we have

$$P_{DF}^{out}(R_{tar}) \leq P\left(\alpha_{SR} < \delta\right) + P\left(\alpha_{SR} \geq \delta\right) P\left(2\alpha_{SD}+\alpha_{RD} < \delta\right).$$

As $\gamma \to \infty$, $\delta \to 0$. Applying the results obtained in [64], we have

$$\begin{aligned} \lim_{\gamma\to\infty} \frac{1}{\delta} P\left(\alpha_{SR} < \delta\right) &= \lambda_{SR} \\ \lim_{\gamma\to\infty} P\left(\alpha_{SR} \geq \delta\right) &= 1 \\ \lim_{\gamma\to\infty} \frac{1}{\delta^2} P\left(2\alpha_{SD}+\alpha_{RD} < \delta\right) &= \frac{\lambda_{SD}\lambda_{RD}}{4}. \end{aligned} \tag{A-35}$$

Thus, the outage probability of the DF cooperation mode at a target rate R_{tar} is upper-bounded according to

$$P_{DF}^{out}(R_{tar}) \leq \lambda_{SR}\delta + \frac{\lambda_{SD}\lambda_{RD}}{4}\delta^2. \tag{A-36}$$

From (A-36), the DF cooperation mode provides no diversity gain even in a high SNR regime [104]. The rationale is that the diversity benefit vanishes due to the decoding capability of Node R. However, given that the decoding is perfect at Node R, the diversity order of two can be achieved in the DF cooperation mode. Denote $P_{\widetilde{DF}}^{out}(R_{tar})$ as the outage probability of the DF cooperation mode with perfect decoding at a target rate R_{tar}. In a high SNR regime, we have

$$P_{\widetilde{DF}}^{out}(R_{tar}) = P\left(C_{\widetilde{DF}} < R_{tar}\right) \leq \frac{\lambda_{SD}\lambda_{RD}}{4}\delta^2.$$

■

A.3.4 Relay Selection

In a large-scale wireless network, it is likely that a source node can exploit the notion of *selection diversity*. With a number of potential relays, the source node can choose the best node as its relay to assist its data transmissions.

Corollary 5 *In a wireless network with m potential relays employing the cooperation protocol of interest with* relay selection*, both the AF cooperation mode and DF cooperation mode with perfect decoding achieves the full diversity order.*

Proof: Following the similar steps shown in the proof of Proposition 21, we consider the outage probability of the AF cooperation mode with relay selection at a target rate R_{tar}, denoted by $P_{AF,sel}^{out}(R_{tar})$. Denote E_{XR_i} as the average energy of a symbol received at Node R_i from Node X, where $i \in \{1, 2, \ldots m\}$, $\alpha_{SR_i} = E_{SR_i}|h_{SR_i}|^2$, and $\alpha_{R_iD} = E_{R_iD}|h_{R_iD}|^2$. We have

$$\begin{aligned} P_{AF,sel}^{out}(R_{tar}) &= P\left(\frac{1}{2}\log_2\left((1+\gamma\alpha_{SD})(1+\frac{\gamma\alpha_{SD}}{w^2}) + \max_i\left\{\frac{\alpha_{SR_i}\alpha_{R_iD}}{w^2}\left(\frac{\gamma^2}{\gamma E_{SR_i}+1}\right)\right\}\right) < R_{tar}\right) \\ &\leq P\left(\alpha_{SD} + \max_i\{\alpha_{SR_i}\alpha_{R_iD}\} < \frac{2^{2R_{tar}}-1}{\varepsilon^{\min}}\right) \end{aligned} \quad \text{(A-37)}$$

where $\varepsilon^{\min} = \min\left\{\left(1+\frac{1}{w^2}\right)\gamma, \min_i\left\{\frac{1}{w^2}\left(\frac{\gamma^2}{\gamma E_{SR_i}+1}\right)\right\}\right\}$. Let $\delta = \frac{2^{2R_{tar}}-1}{\varepsilon^{\min}}$. Let λ_{SR_i} and λ_{R_iD} be the parameters of the exponential random variables α_{SR_i} and α_{R_iD}, respectively. Inequality (A-37) becomes

$$\begin{aligned} P_{AF,sel}^{out}(R_{tar}) &\leq \int_0^{\delta} P\left(\alpha_{SR_i}\max_i\{\alpha_{R_iD}\} < \delta - x\right)\lambda_{SD}e^{-\lambda_{SD}x}dx \\ &= \int_0^{\delta}\prod_{i=1}^{m} P\left(\alpha_{SR_i}\alpha_{R_iD} < \delta - x\right)\lambda_{SD}e^{-\lambda_{SD}x}dx. \end{aligned}$$

By Proposition 21, in a high SNR regime, we have

$$P_{AF,sel}^{out}(R_{tar}) \leq \frac{\lambda_{SD}\prod_{i=1}^{m}\left(\lambda_{SR_i}+\lambda_{R_iD}\right)}{m+1}\delta^{m+1}.$$

Therefore, the full diversity order can be obtained with relay selection in the AF cooperation mode.

The outage probability of the DF cooperation mode with perfect decoding and relay selection at a target rate R_{tar}, denoted by $P_{\widetilde{DF},sel}^{out}(R_{tar})$, is given by

$$\begin{aligned} P_{\widetilde{DF},sel}^{out}(R_{tar}) &= P\left(2\gamma\alpha_{SD}+\gamma^2\alpha_{SD}^2+\max_i\{\gamma\alpha_{R_iD}\} < 2^{2R_{tar}}-1\right) \\ &\leq P\left(\alpha_{SD}+\max_i\{\alpha_{R_iD}\} < \delta\right) \\ &= \int_0^{\delta} P\left(\max_i\{\alpha_{R_iD}\} < \delta - x\right)\lambda_{SD}e^{-\lambda_{SD}x}dx \\ &= \int_0^{\delta}\prod_{i=1}^{m} P\left(\alpha_{R_iD} < \delta - x\right)\lambda_{SD}e^{-\lambda_{SD}x}dx \end{aligned}$$

where $\delta = \frac{2^{2R_{tar}}-1}{\gamma}$. Thus, by Proposition 21, in a high SNR regime, we have

$$P_{\widetilde{DF},sel}^{out}(R_{tar}) \leq \frac{\lambda_{SD}\prod_{i}^{m}\lambda_{R_iD}}{m+1}\delta^{m+1}.$$

With relay selection, the DF cooperation mode with perfect decoding achieves the full diversity order. ■

A.3.5 Cooperation versus Non-Cooperation

Consider a non-altruistic cooperative wireless network consisting of a number of nodes, including Node S, Node R, and Node D. Suppose Node D is a receiving node at the time period of interest, and both Node S and Node R have their own data to transmit. Here, we adopt the DF cooperation mode to illustrate the idea of whether and when node cooperation is beneficial. According to the discussion given in Section A.3.2, the DF cooperation mode should not be initiated unless the decoding at Node R is reliable. Therefore, the DF cooperation mode is employed only if Node R can reliably decode the symbols sent from Node S and vice versa. Assuming perfect decoding at Node R, we have

$$y_{D,1} = \sqrt{a_S E_{SD}} h_{SD} x_1 + \eta_{D,1} + I \tag{A-38}$$

$$y_{R,1} = \sqrt{a_S E_{SR}} h_{SR} x_1 + \eta_{R,1} + I \tag{A-39}$$

$$y_{D,2} = \sqrt{a_S E_{SD}} h_{SD} x_2 + \sqrt{a_R E_{RD}} h_{RD} x_1 + \eta_{D,2} + I \tag{A-40}$$

where a_X is the scaling factor for the transmit power of Node X, i.e., $0 \leq a_X \leq 1$ and I is the co-channel interference (e.g., generated by the transmissions from other clusters). With effective node clustering, the co-channel interference level can be strictly bounded [24] (see Chapter 3), and we denote σ^2 as the aggregate noise-plus-co-channel interference power. Thus, (A-38) and (A-40) can be represented in the following matrix form

$$\mathbf{y}_{DF,p} = \mathbf{H}_{DF,p}\mathbf{x} + \mathbf{Q}_{DF,p} \tag{A-41}$$

where $\mathbf{y}_{DF,p} = [y_{D,1}\ y_{D,2}]$, $\mathbf{Q}_{DF,p} = [\eta_{D,1} + I\ \eta_{D,2} + I]$, and $\mathbf{H}_{DF,p}$ is the channel matrix given by

$$\mathbf{H}_{DF,p} = \begin{bmatrix} \sqrt{a_S E_{SD}} h_{SD} & 0 \\ \sqrt{a_R E_{RD}} h_{RD} & \sqrt{a_S E_{SD}} h_{SD} \end{bmatrix}. \tag{A-42}$$

Denote C^c as the channel capacity achieved by means of node cooperation from Node S to Node D (with the help of Node R), which is given by

$$\begin{aligned} C^c &= \frac{1}{2}\log_2 \det\left(\mathbf{I_2} + \frac{1}{\sigma^2}\mathbf{H}_{DF,p}\mathbf{H}_{DF,p}^*\right) \\ &= \frac{1}{2}\log_2\left((1+a_S\gamma_{SD})^2 + a_R\gamma_{RD}\right) \end{aligned} \tag{A-43}$$

where $\gamma_{SD} = E_{SD}|h_{SD}|^2/\sigma^2$ and $\gamma_{RD} = E_{RD}|h_{RD}|^2/\sigma^2$.

Corollary 6 *Given perfect decoding at Node* R*, arbitrarily positive power allocation has no impact on the diversity performance in the DF cooperation mode.*

Proof: Similar to the proof of Proposition 21, we consider the outage probability of the DF cooperation mode with arbitrarily positive power allocation, denoted by $P_c^{out}(R_{tar})$, where $a_S, a_R > 0$. We have

$$\begin{aligned} P_c^{out}(R_{tar}) &= P(C^c < R_{tar}) \\ &= P\left(2a_S\gamma_{SD} + a_S^2\gamma_{SD}^2 + a_R\gamma_{RD} < 2^{2R_{tar}} - 1\right) \\ &\leq P\left(2a_S\gamma_{SD} + a_R\gamma_{RD} < 2^{2R_{tar}} - 1\right) \\ &\leq P\left(2\alpha_{SD} + \alpha_{RD} < \delta\right) \end{aligned}$$

where $\alpha_{SD} = E_{SD}|h_{SD}|^2$, $\alpha_{RD} = E_{RD}|h_{RD}|^2$, $\delta = \frac{2^{2R_{tar}}-1}{\varepsilon^{\min}}$, and $\varepsilon^{\min} = \min\{\frac{a_S}{\sigma^2}, \frac{a_S}{\sigma^2}\}$. As $\sigma^2 \to 0$, $\delta \to 0$. Therefore, by (A-35), we have

$$\lim_{\delta\to 0}\frac{1}{\delta^2}P_c^{out}(R_{tar}) \leq \frac{\lambda_{SD}\lambda_{RD}}{4}.$$

■

In fact, the results obtained in Corollary 6 can be extended to the case with relay selection.

Corollary 7 *Given perfect decoding at Node* R*, arbitrarily positive power allocation has no impact on the diversity performance in the DF cooperation mode with relay selection.*

Proof: It can be proved by following the same arguments given in the proof of Corollaries 5 and 6. ■

Denote C^{nc} as the channel capacity achieved from Node S to Node D without node cooperation (i.e., *ordinary direct transmissions*), which is given by

$$C^{nc} = \log_2 (1 + \gamma_{SD}) . \tag{A-44}$$

In the case of non-altruistic node cooperation, power allocation is imperative as part of the transmit power of a node is dedicated to transmitting their own data (i.e., *direct transmissions*) and the rest of the transmit power is dedicated to relaying the data from other nodes (i.e., *assisted transmissions*). The variables a_S and a_R in (A-43) capture the power allocation. Here, we derive a sufficient condition for a cooperative transmission being advantageous over an ordinary direct transmission, which is given by

$$\begin{aligned} & C^c \geq C^{nc} \\ \Rightarrow & (1 + a_S \gamma_{SD})^2 + a_R \gamma_{RD} \geq (1 + \gamma_{SD})^2 \\ \Rightarrow & \ a_S \geq \frac{-1 + \sqrt{(1 + \gamma_{SD})^2 - a_R \gamma_{RD}}}{\gamma_{SD}}. \end{aligned} \tag{A-45}$$

Notice that in the case where relay selection is considered, $R = \arg\max_{R_i} \{a_{R_i} \gamma_{R_i D}\}$. Consider $\gamma_{SD} = \gamma_{RD} = \hat{\gamma}$ and $a_R = \rho a_S$, where $\rho \geq 0$. The sufficient condition for a cooperative transmission being advantageous over an ordinary direct transmission is given by

$$a_S \geq \frac{-(2 + \rho) + \sqrt{(2 + \rho)^2 + 4\hat{\gamma}(\hat{\gamma} + 2)}}{2\hat{\gamma}}. \tag{A-46}$$

First, we consider the boundary case where $a_R = 0$. Since a_S has to be positive (i.e., $a_S > 0$) for a feasible cooperative transmission, $\rho = 0$. When $\rho = 0$, from (A-46), $a_S \geq 1$; however, since $0 \leq a_S \leq 1$, $a_S = 1$. Thus, for the case of $a_S = 1$ and $a_R > 0$, the sufficient

condition is (strictly) satisfied. In fact, this corresponds to the scenario of altruistic node cooperation (e.g., with pure relays in IEEE 802.16j). On the other hand, given the value of ρ, the value of a_S is lower-bounded by (A-46) in order for node cooperative transmissions to be beneficial. This scenario of interest corresponds to non-altruistic node cooperation in wireless networks where every node has its own data to transmit and only a portion of its transmit power can be dedicated to relaying data from other neighboring nodes. In fact, the condition (A-45) refers to the cooperative transmissions being beneficial to an individual node only. For the sake of overall system performance, cooperation among nodes should be considered in a holistic manner, as discussed in Chapter 5.

A.3.6 Numerical Results

In our numerical analysis, we compare the outage probabilities of non-altruistic cooperative transmissions and ordinary direct transmissions. Regarding non-altruistic node cooperation with regenerative wireless nodes, the nodes that can successfully decode some neighboring node's transmissions become its potential relays. Here, if the rate achieved by a source-relay link is larger than that by a cooperative transmission given in (A-43), we assume that the relay of interest can reliably decode the source node's data and become one of its potential relays [64]. Figure A.1 depicts the outage probabilities of cooperative transmissions using the DF cooperation mode and ordinary direct transmissions. As observed in Figure A.1, without considering perfect decoding, no diversity gain can be obtained even in the presence of multiple relays (i.e., Proposition 21). On the other hand, given perfect decoding, node cooperative transmissions with one relay achieve the diversity order of two (i.e., Corollary 6), while the ones with three relays achieve the diversity order of four (i.e., Corollary 7). The details of this supplementary chapter are reported in [25].

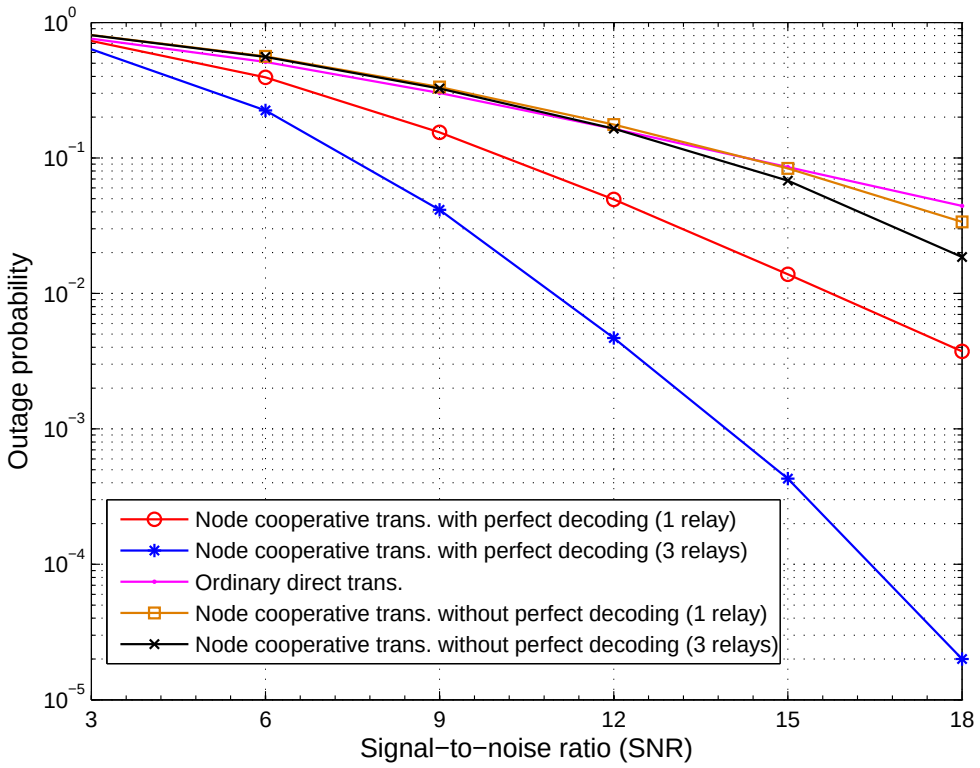


Figure A.1: Outage probabilities of cooperative transmissions using the DF cooperation mode and ordinary direct transmissions (where $E_{SD} = E_{RD}$, $a_S = 0.75$, $a_R = 0.25$, and $\sigma^2 = 10^{-3}$W).

Appendix B

Proofs of Propositions, Corollaries, Lemmas, and Theorems

Proposition 1 *Suppose all available subcarriers are chosen by the clusterheads at least once on every timeslot (i.e.,* $\sum_m c_{m,n}^l \geq 1, \forall n, l$*). The subcarrier allocation solution obtained from our proposed approach is Pareto optimal.*

Proof: Given the action profile or solution $\mathbf{c}^*$ (i.e., subcarrier allocation) obtained from (3.5), $U_m(\mathbf{c}^*, R_m^d)$ is the utility acquired by the m^{th} clusterhead. Consider another action profile $\tilde{\mathbf{c}}$. For some m, if $U_m(\tilde{\mathbf{c}}, R_m^d) > U_m(\mathbf{c}^*, R_m^d)$, then $U_n(\tilde{\mathbf{c}}, R_n^d) < U_n(\mathbf{c}^*, R_n^d)$ for some n, as either $\tilde{\mathbf{c}}_\mathbf{m} \succeq \mathbf{c}_m^*$ or $\tilde{\mathbf{c}}_\mathbf{n} \preceq \mathbf{c}_n^*$ or both. According to the definition of Pareto optimality, the subcarrier allocation solution $\mathbf{c}^*$ obtained from our proposed approach achieves the Pareto optimality. ■

Proposition 2 *Suppose all available subcarriers are active. If all the clusters are heavily loaded (i.e.,* $0 < \sum_l^L \sum_n^N c_{m,n}^l r_{m,n}^l - R_m^d < \epsilon, \forall m$*), frequency reuse is prohibited.*

Proof: If $0 < \sum_l^L \sum_n^N c_{m,n}^l r_{m,n}^l - R_m^d < \epsilon, \forall m$, $\varepsilon_{m,n}^l \rightarrow \infty, \forall m, n, l$, and hence $T_{mj,n}^l \rightarrow \infty, \forall j, m, n, l$. According to the subcarrier allocation criterion given in (3.5), a subcarrier

can be selected only once (i.e., $\sum_m c^l_{m,n} = 1, \forall n, l$). Therefore, frequency reuse is prohibited. ■

Proposition 3 *Suppose all available subcarriers are active and all the clusters are heavily loaded. Modeled by a round-robin game played by the clusterheads, the proposed subcarrier allocation solution attains an NE.*

Proof: Rewrite the payoff function of the m^{th} clusterhead over the n^{th} subcarrier on the l^{th} timeslot as $S^l_{m,n}(c^l_{m,n}, c^l_{-m,n})$, where $c^l_{-m,n} = (c^l_{1,n}, ..., c^l_{m-1,n}, c^l_{m+1,n}, ...c^l_{M,n})$. The proposed subcarrier allocation approach can be modeled by a round-robin game, where clusterheads (players) take turn to maximize their payoffs based on the subcarrier allocation criterion given in (3.5). Under the conditions that all available subcarriers are active and all the clusters are heavily loaded, all the clusterheads would have no intention to re-allocate the subcarriers, for $S^l_{m,n}(c^{*l}{}_{m,n}, c^{*l}{}_{-m,n}) \geq S^l_{m,n}(c^l_{m,n}, c^{*l}{}_{-m,n}), \forall m, n, l$, where $\mathbf{c}^*$ is the currently obtained subcarrier allocation solution and $\mathbf{c}$ is another solution. According to the definition of an NE, the subcarrier allocation solution $\mathbf{c}^*$ obtained from our proposed approach attains an NE. ■

Corollary 1 *If an NE is attained, all available subcarriers are in use.*

Proof: Suppose that there exists a subcarrier allocation solution $\tilde{\mathbf{c}}$ at an NE, where not all the subcarriers are in use, i.e., $\sum_m \tilde{c}^l_{m,n} = 0, \exists n, l$. We can always find another solution $\mathbf{c}^*$ where $\mathbf{c}^* \geq \tilde{\mathbf{c}}$ such that $S^l_{m,n}(c^{*l}{}_{m,n}, \tilde{c}^l_{-m,n}) > S^l_{m,n}(\tilde{c}^l_{m,n}, \tilde{c}^l_{-m,n}), \exists m, n, l$, which contradicts to the definition of an NE. Thus, no such solution $\tilde{\mathbf{c}}$ exists at an NE, and all available subcarriers are in use at an NE. ■

Proposition 4 *The ICRAOP is an NP-hard problem.*

Proof: We prove the NP-hardness by reducing the well-known NP-complete *number partitioning problem* (denoted by PARTITION[1]) [34] to the ICRAOP. We consider the special case of the ICRAOP by fixing the power allocation and letting $M_c = 2$, $r_{1,n}^l = r_{2,n}^l = r_n^l$, $\forall n$, and $R_1^d = R_2^d = \sum_{n=1}^{N_c} \sum_{l=1}^{L_c} \frac{r_n^l}{2}$. Thus, the "size" of the n^{th} subcarrier at the l^{th} timeslot is r_n^l. Since each subcarrier can only be allocated to one link at any timeslot (i.e., $c_{1,n}^l + c_{2,n}^l = 1$, $\forall n, l$), the solution to this special-case problem is exactly the same as that of the PARTITION problem. In other words, the PARTITION problem can be *polynomially transformed* into the special-case problem, and vice versa. Since the PARTITION problem is NP-complete, the special-case problem is also NP-complete. Further, the above special-case problem can be generalized by the ICRAOP. Thus, the ICRAOP is an NP-hard problem. ■

Lemma 1 *The utility function of the m^{th} node, given in (5.13), is an increasing function of its achievable data rate.*

Proof: Without loss of generality, assume $R_m(\tilde{\mathbf{c}}, \tilde{\mathbf{p}}, \tilde{\mathbf{a}}, \tilde{\mathbf{z}}) > R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z}), \exists_m$. Consider the following two cases:

Case 1: For $\mathbf{\Theta} = 1$, since $R_m(\tilde{\mathbf{c}}, \tilde{\mathbf{p}}, \tilde{\mathbf{a}}, \tilde{\mathbf{z}}) > R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})$, $U_m(R_m(\tilde{\mathbf{c}}, \tilde{\mathbf{p}}, \tilde{\mathbf{a}}, \tilde{\mathbf{z}})|\mathbf{\Theta}) = R_m(\tilde{\mathbf{c}}, \tilde{\mathbf{p}}, \tilde{\mathbf{a}}, \tilde{\mathbf{z}}) > R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z}) = U_m(R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})|\mathbf{\Theta})$;

Case 2: For $\mathbf{\Theta} = 2$, since $R_m(\tilde{\mathbf{c}}, \tilde{\mathbf{p}}, \tilde{\mathbf{a}}, \tilde{\mathbf{z}}) > R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})$,

$$\ln\left(\frac{R_m(\tilde{\mathbf{c}}, \tilde{\mathbf{p}}, \tilde{\mathbf{a}}, \tilde{\mathbf{z}})}{A}\right) > \ln\left(\frac{R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})}{A}\right) \tag{B-1}$$

$$\Rightarrow \left[-\ln\left(\frac{R_m(\tilde{\mathbf{c}}, \tilde{\mathbf{p}}, \tilde{\mathbf{a}}, \tilde{\mathbf{z}})}{A}\right)\right]^{\kappa} < \left[-\ln\left(\frac{R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})}{A}\right)\right]^{\kappa}, \text{ for } \kappa > 0 \tag{B-2}$$

$$\Rightarrow U_m(R_m(\tilde{\mathbf{c}}, \tilde{\mathbf{p}}, \tilde{\mathbf{a}}, \tilde{\mathbf{z}})|\mathbf{\Theta}) > U_m(R_m(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z})|\mathbf{\Theta}). \tag{B-3}$$

[1]Any instance of PARTITION is a finite set $\Omega = \{1, 2, \ldots, N\}$, and a size $s(n) \in Z^+$ for each $n \in \Omega$. Any yes-instance of PARTITION is a subset (i.e., partition) of Ω, $N' \subseteq \Omega$ such that $\sum_{n \in N'} s(n) = \sum_{n \in \Omega - N'} s(n)$ [34].

By combining the above two cases, the utility function of the m^{th} node increases with its achievable data rate. ■

Proposition 5 *The NCRAOP is an NP-hard problem.*

Proof: We omit the proof as it is similar to the one for Proposition 4. ■

Proposition 6 *(**Optimal Power Allocation**) Let*

$$f(p_{m,n}^{l}) = {\xi^{(2)}}_{m}^{l} - \sum_{u=1}^{M} \left(U_u' \left(R_u \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{u} \right) \frac{\partial R_u \left(\cdot \right)}{\partial p_{m,n}^{l}}, \forall_{m,n,l}. \tag{B-4}$$

The optimal power allocation, denoted by $\mathbf{p}^*$*, for the NCRAOP is given by* ${p^*}_{m,n}^{l} = \max\{\tilde{p}_{m,n}^{l}, 0\}, \forall_{m,n,l}$*, where* $\tilde{p}_{m,n}^{l}$ *is the solution of the function* $f(\tilde{p}_{m,n}^{l}) = 0$*.*

Proof: We substitute ${\xi^{(7)}}_{m,n}^{l}$ in (5.31) into (5.34), which leads to

$$p_{m,n}^{l} f(p_{m,n}^{l}) = 0, \forall_{m,n,l}. \tag{B-5}$$

Since ${\xi^{(7)}}_{m,n}^{l} \geq 0$, $f(p_{m,n}^{l}) \geq 0$. Thus, ${p^*}_{m,n}^{l} = 0$ if $f(p_{m,n}^{l}) > 0$ or ${p^*}_{m,n}^{l} = \tilde{p}_{m,n}^{l}$ if $f(\tilde{p}_{m,n}^{l}) = 0$, otherwise. The optimal power allocation for the NCRAOP is given by ${p^*}_{m,n}^{l} = \max\{\tilde{p}_{m,n}^{l}, 0\}, \forall_{m,n,l}$. ■

Proposition 7 *(**Subcarrier Allocation Criterion**) For the* n^{th} *subcarrier and the* l^{th} *timeslot, the necessary condition for* $c_{m,n}^{l}$ *being positive is*

$$m = \arg\max_{\tilde{m}} \left\{ \left(U_{\tilde{m}}' \left(R_{\tilde{m}} \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{\tilde{m}} \right) \frac{\partial R_{\tilde{m}} \left(\cdot \right)}{\partial c_{\tilde{m},n}^{l}} \right\}. \tag{B-6}$$

Proof: We substitute ${\xi^{(9)}}_{m,n}^{l}$ in (5.29) into (5.32), which leads to

$$c_{m,n}^{l} \left({\xi^{(3)}}_{n}^{l} - \left(U_m' \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{m} \right) \frac{\partial R_m \left(\cdot \right)}{\partial c_{m,n}^{l}} \right) = 0, \forall_{m,n,l}. \tag{B-7}$$

Since ${\xi^{(9)}}_{m,n}^{l} \geq 0$,

$$ {\xi^{(3)}}_{n}^{l} \geq \left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{m} \right) \frac{\partial R_m \left(\cdot \right)}{\partial c_{m,n}^{l}}. \tag{B-8} $$

Therefore, the necessary condition for $c_{m,n}^{l}$ being positive is

$$ {\xi^{(3)}}_{n}^{l} = \left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{m} \right) \frac{\partial R_m \left(\cdot \right)}{\partial c_{m,n}^{l}}. \tag{B-9} $$

Notice that ${\xi^{(3)}}_{n}^{l}$ can be interpreted as an upper bound of $\left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{m} \right) \frac{\partial R_m (\cdot)}{\partial c_{m,n}^{l}}$ which represents the (scaled) marginal increase in $U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right)$ when the n^{th} subcarrier is allocated to the m^{th} node on the l^{th} timeslot (i.e., $c_{m,n}^{l} = 1$). Thus, the criterion of subcarrier allocation can be given by

$$ m = \arg\max_{\tilde{m}} \left\{ \left(U'_{\tilde{m}} \left(R_{\tilde{m}} \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{\tilde{m}} \right) \frac{\partial R_{\tilde{m}} \left(\cdot \right)}{\partial c_{\tilde{m},n}^{l}} \right\}. \tag{B-10} $$

Assuming $R_m \left(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z} \right) \neq R_m^d$ (i.e., ${\xi^{(1)}}_{m} = 0$), the subcarrier allocation criterion can be further simplified as follows. For the n^{th} subcarrier and the l^{th} timeslot, choose m^* such that

$$ m^* = \arg\max_{m} \left\{ U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) \frac{\partial R_m \left(\cdot \right)}{\partial c_{m,n}^{l}} \right\} \tag{B-11} $$

and set $c_{m^*,n}^{l} = 1$. ■

Proposition 8 ***(Partner Selection Criterion)*** *The necessary condition for z_{mu} being positive is*

$$ u = \arg\max_{\tilde{u} \neq m} \left\{ \frac{1}{a_{m\tilde{u}}} \left(\left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{m} \right) \frac{\partial R_m \left(\cdot \right)}{\partial z_{m\tilde{u}}} - {\xi^{(4)}}_{\tilde{u}} - {\xi^{(5)}}_{m} \right) \right\}. \tag{B-12} $$

Proof: We substitute ${\xi^{(10)}}_{mu}$ in (5.30) into (5.33), which leads to

$$ z_{mu} \left(- \left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{m} \right) \frac{\partial R_m \left(\cdot \right)}{\partial z_{mu}} + {\xi^{(4)}}_{u} + {\xi^{(5)}}_{m} + {\xi^{(6)}}_{u} a_{mu} \right) = 0, \forall_{m,u}. \tag{B-13} $$

Since ${\xi^{(10)}}_{mu} \geq 0$,

$$ {\xi^{(6)}}_{u} a_{mu} \geq \left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + {\xi^{(1)}}_{m} \right) \frac{\partial R_m \left(\cdot \right)}{\partial z_{mu}} - {\xi^{(4)}}_{u} - {\xi^{(5)}}_{m}. \tag{B-14} $$

Therefore, the necessary condition for z_{mu} being positive is

$$\xi^{(6)}{}_u = \frac{1}{a_{mu}} \left(\left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + \xi^{(1)}{}_m \right) \frac{\partial R_m \left(\cdot \right)}{\partial z_{mu}} - \xi^{(4)}{}_u - \xi^{(5)}{}_m \right). \tag{B-15}$$

Here, $\xi^{(6)}{}_u$ can be interpreted as an upper bound of $\frac{1}{a_{mu}} \left(\left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + \xi^{(1)}{}_m \right) \frac{\partial R_m(\cdot)}{\partial z_{m\tilde{u}}} - \xi^{(4)}{}_u - \xi^{(5)}{}_m \right)$ which represents the (scaled) marginal increase in $U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right)$ when the u^{th} node becomes a partner of the m^{th} node (i.e., $z_{mu} = 1$). Thus, the criterion of partner selection can be given by

$$u = \arg\max_{\tilde{u} \neq m} \left\{ \frac{1}{a_{m\tilde{u}}} \left(\left(U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right) + \xi^{(1)}{}_m \right) \frac{\partial R_m \left(\cdot \right)}{\partial z_{m\tilde{u}}} - \xi^{(4)}{}_{\tilde{u}} - \xi^{(5)}{}_m \right) \right\}. \tag{B-16}$$

Assuming $R_m \left(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z} \right) \neq R_m^d$ (i.e., $\xi^{(1)}{}_m = 0$), and ignoring the constraints (5.18) and (5.19) (i.e., $\xi^{(4)}{}_u = \xi^{(5)}{}_m = 0$), the partner selection criterion can be deduced as follows. For the m^{th} node, choose u^* such that

$$u^* = \arg\max_{u \neq m} \left\{ \frac{U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right)}{a_{mu}} \frac{\partial R_m \left(\cdot \right)}{\partial z_{mu}} \right\} \tag{B-17}$$

and set $z_{mu^*} = 1$. ■

Corollary 2 *(**Partner Allocation Criterion**) The necessary condition for z_{mu} being positive is*

$$m = \arg\max_{\tilde{m} \neq u} \left\{ \frac{1}{a_{\tilde{m}u}} \left(\left(U'_{\tilde{m}} \left(R_{\tilde{m}} \left(\cdot \right) | \mathbf{\Theta} \right) + \xi^{(1)}{}_{\tilde{m}} \right) \frac{\partial R_{\tilde{m}} \left(\cdot \right)}{\partial z_{\tilde{m}u}} - \xi^{(4)}{}_u - \xi^{(5)}{}_{\tilde{m}} \right) \right\}. \tag{B-18}$$

Proof: The proof is the same as the one for Proposition 8. Assuming $R_m \left(\mathbf{c}, \mathbf{p}, \mathbf{a}, \mathbf{z} \right) \neq R_m^d$ (i.e., $\xi^{(1)}{}_m = 0$), and ignoring the constraints (5.18) and (5.19) (i.e., $\xi^{(4)}{}_u = \xi^{(5)}{}_m = 0$), the partner allocation criterion can be deduced as follows. For the u^{th} node, choose m^* such that

$$m^* = \arg\max_{m \neq u} \left\{ \frac{U'_m \left(R_m \left(\cdot \right) | \mathbf{\Theta} \right)}{a_{mu}} \frac{\partial R_m \left(\cdot \right)}{\partial z_{mu}} \right\} \tag{B-19}$$

and set $z_{m^*u} = 1$. ■

Proposition 9 *(**Power Allocation Criterion for Assisted Transmissions**) For the u^{th} node, the sufficient condition for a_{mu} being positive is $a_{uj} > 0$ (and $z_{uj} = 1$), $\exists m, j$.*

Proof: Since the total transmit power of the u^{th} node will be split into parts (i.e., $a_{uu} > 0$ and $a_{mu} > 0$) only if it becomes a partner of the m^{th} node, for some m, it is trivial that $a_{mu} > 0$ if $z_{mu} = 1$; otherwise, $a_{mu} = 0$. Also, from (5.20), $a_{mu} > 0$ if $a_{uu} < 1$. Similar to the discussion given in Appendix A.3.5, in order for a cooperative transmission being advantageous over an ordinary direct transmission, the lower bound on a_{uu} can be given by

$$a_{uu} \geq \max_{\forall n,l} \left\{ \frac{-1 + \sqrt{(1 + g^l_{uu,n} p^l_{u,n})^2 - z_{uj} a_{uj} g^l_{uj,n} p^l_{j,n}}}{g^l_{uu,n} p^l_{u,n}} \right\}. \tag{B-20}$$

Therefore, if $a_{mu} > 0, \exists m$, then $a_{uu} < 1$. In order for the condition (B-20) to be satisfied, $a_{uj} > 0$ (and $z_{uj} = 1$), $\exists j$. ■

Proposition 10 *With the subcarrier allocation solution, denoted by $\tilde{\mathbf{c}}$, the optimal power allocation solution, denoted by $\mathbf{p}^*$, for the optimization problem given in (5.56)-(5.59) is given*

$$p^{*l}_{m,n} = \tilde{c}^l_{m,n} \left[\frac{U'_m(R_m(\mathbf{p})|\mathbf{\Theta}) + \xi^{(1)}{}_m}{\xi^{(2)l}_m} - \frac{1}{g^l_{mm,n}} \right]^+, \forall_{m,n,l} \tag{B-21}$$

where $[x]^+ = \max\{0, x\}$.

Proof: Since $z_{mu} = 0, \forall_{m \neq u}$, the KKT condition with respect to $p^l_{m,n}$ given in (5.31) becomes

$$-\left(U'_m \left(R_m(\mathbf{p})|\mathbf{\Theta} \right) + \xi^{(1)}{}_m \right) \frac{\partial R_m(\mathbf{p})}{\partial p^l_{m,n}} + \xi^{(2)l}_m = \xi^{(7)l}_{m,n}, \forall m, n, l. \tag{B-22}$$

We substitute $\xi^{(7)l}_{m,n}$ in (B-22) into (5.34), which leads to

$$p^l_{m,n} \left(\xi^{(2)l}_m - \left(U'_m \left(R_m(\mathbf{p})|\mathbf{\Theta} \right) + \xi^{(1)}{}_m \right) \frac{\partial R_m(\mathbf{p})}{\partial p^l_{m,n}} \right) = 0, \forall_{m,n,l}. \tag{B-23}$$

If ${\xi^{(2)}}^l_m > \left(U'_m\left(R_m(\mathbf{p})|\mathbf{\Theta}\right) + {\xi^{(1)}}_m\right)\frac{\partial R_m(\mathbf{p})}{\partial p^l_{m,n}}$, then the condition (B-23) holds if $p^l_{m,n} = 0$; otherwise, $p^l_{m,n} = \frac{U'_m(R_m(\tilde{\mathbf{c}})|\mathbf{\Theta}) + {\xi^{(1)}}_m}{{\xi^{(2)}}^l_m} - \frac{1}{g^l_{mm,n}}$. Thus, the optimal power allocation solution is given by

$$ {p^*}^l_{m,n} = \tilde{c}^l_{m,n}\left[\nu^l_m - \frac{1}{g^l_{mm,n}}\right]^+ \tag{B-24} $$

where $\nu^l_m = \frac{U'_m(R_m(\tilde{\mathbf{c}})|\mathbf{\Theta}) + {\xi^{(1)}}_m}{{\xi^{(2)}}^l_m}$. Notice that ν^l_m is often referred to as the *water level* of the m^{th} node on the l^{th} timeslot such that $\sum_{n=1}^N {p^*}^l_{m,n} = P^{\max}_m$. ■

Lemma 2 *With the subcarrier allocation solution and the partner allocation solution, denoted by* $\tilde{\mathbf{c}}$ *and* $(\tilde{\mathbf{a}}, \tilde{\mathbf{z}})$*, respectively, the power allocation solution obtained from water filling, denoted by* $\mathbf{p}^*$*, is given by*

$$ {p^*}^l_{m,n} = \tilde{c}^l_{m,n}\left[\frac{-\left(2{\xi^{(2)}}^l_m - \Upsilon\tilde{a}^l_{mm}g^l_{mm,n}\right) + \sqrt{\Upsilon^2(\tilde{a}^l_{mm})^2(g^l_{mm,n})^2 - 4({\xi^{(2)}}^l_m)^2\Gamma_m}}{2{\xi^{(2)}}^l_m\tilde{a}^l_{mm}g^l_{mm,n}}\right]^+, \ \forall m,n,l \tag{B-25} $$

where $\Upsilon = U'_m(R_m(\mathbf{p})|\mathbf{\Theta}) + {\xi^{(1)}}_m$ *and* $\Gamma_m = \sum_{u\neq m}\tilde{z}_{mu}\tilde{a}_{mu}g^l_{mu,n}\hat{p}^l_{u,n}$.

Proof: We omit the proof as it is similar to the one for Proposition 10. ■

Proposition 11 *Modeled by a round-robin game played by the RG nodes with fixed subcarrier allocation and power allocation, the proposed partner selection solution in the distributed resource allocation achieves an NE.*

Proof: Rewrite the utility function of the m^{th} node as $U_m(\mathbf{z}_m, \mathbf{z}_{-m})$, where $\mathbf{z}_m = [z_{mu}]_{1\times M}$ and $\mathbf{z}_{-m} = \left([z_{mu}]_{1\times M}\ [z_{2u}]_{1\times M}\cdots\ [z_{(m-1)u}]_{1\times M}\ [z_{(m+1)u}]_{1\times M}\cdots\ [z_{Mu}]_{1\times M}\right)^T$. Viewing $U_m(\mathbf{z}_m, \mathbf{z}_{-m})$ as the payoff function of the m^{th} node, we model our distributed node cooperative resource allocation approach as a round-robin game played by the RG nodes. In specific, with fixed subcarrier allocation and power allocation, each RG node takes turn to maximize its own

payoff function by choosing the best partner. Since each RG node selects the best partner in the proposed distributed approach, the RG nodes would have no tendency to perform partner re-selection, for $U_m(\mathbf{z}^*{}_m, \mathbf{z}^*{}_{-m}) \geq U_m(\mathbf{z}_m, \mathbf{z}^*{}_{-m}), \forall_m$, where $\mathbf{z}^*$ is the solution currently obtained from our proposed distributed approach and $\mathbf{z}$ is another solution. According to the definition of an NE, the partner selection solution $\mathbf{z}^*$ obtained from our proposed distributed approach attains an NE. ■

Proposition 12 *Modeled by a round-robin game played by the potential partners with fixed subcarrier allocation and power allocation, the partner allocation solution from the proposed centralized approach achieves an NE.*

Proof: Let $S_j\left(\mathbf{z}^*{}_j, \mathbf{z}^*{}_{-j}\right)$ denote the payoff function of the j^{th} node with the partner allocation solution $\mathbf{z}^*$ obtained from our proposed centralized approach, where the payoff function here indicates the utility gain when the j^{th} node is chosen to be the partner of some source node. With fixed subcarrier allocation and power allocation, we model the proposed centralized node cooperative resource allocation approach as a round-robin game played by the potential partners. Each potential partner takes turn to maximize its own payoff function by choosing the best source node to assist. If a partner cannot find a favorable source node to assist (i.e., allocating this partner to any source node results in a negative payoff), that partner is disqualified from the game. In our proposed centralized approach, a potential partner is allocated to a source node such as the utility gain can be maximized. Once allocated, therefore, the potential partners would have no intention to change their partner allocation solutions, for $S_j(\mathbf{z}^*{}_j, \mathbf{z}^*{}_{-j}) \geq S_j(\mathbf{z}_j, \mathbf{z}^*{}_{-j}), \forall_j$, where $\mathbf{z}^*$ is the solution currently obtained from our proposed centralized approach and $\mathbf{z}$ is another solution. According to the definition of an NE, the partner allocation solution $\mathbf{z}^*$ obtained from our proposed centralized approach attains an NE. ■

Proposition 13 *With fixed subcarrier allocation and power allocation, the partner selection solution from the proposed distributed approach is Pareto optimal.*

Proof: Given the partner selection solution $\mathbf{z}^*$ obtained from the proposed distributed approach, $U_m(\mathbf{z}^*)$ is the utility function of the m^{th} node. Let $\tilde{\mathbf{z}}$ be another solution. Consider the following two cases:

Case 1: For $\sum_{j \neq m} {z^*}_{mj} = 0$ (i.e., the m^{th} node does not have a partner), if $U_m(\tilde{\mathbf{z}}) > U_m(\mathbf{z}^*), \exists_m$ as $\tilde{z}_{mu} = 1, \exists_u$, then either $U_u(\tilde{\mathbf{z}}) < U_u(\mathbf{z}^*)$ as $\sum_{i \neq u} {z^*}_{iu} = 0$ or $U_i(\tilde{\mathbf{z}}) < U_i(\mathbf{z}^*)$ as ${z^*}_{iu} = 1, \exists_i$. If the u^{th} node becomes the partner of the m^{th} node (i.e., $\tilde{z}_{mu} = 1$), the utility achieved by the u^{th} node will decrease due to the split of transmit power. On the other hand, if the u^{th} node has been selected by the i^{th} node (i.e., ${z^*}_{iu} = 1$), meaning that the u^{th} node is the best partner for the i^{th} node in $\mathbf{z}^*$, the utility achieved by the i^{th} node will decrease with a different partner selection solution (i.e., $\tilde{z}_{ij} = 1, \exists_{j \neq u}$ or $\sum_{j \neq i} \tilde{z}_{ij} = 0$);

Case 2: For ${z^*}_{mj} = 1, \exists_j$ (i.e., the m^{th} node has a partner, the j^{th} node), if $U_m(\tilde{\mathbf{z}}) > U_m(\mathbf{z}^*), \exists_m$ as $\tilde{z}_{mu} = 1, \exists_u$, it means that the u^{th} node is the better partner for the m^{th} node than the j^{th} node. Since the j^{th} node is the best partner for the m^{th} node in $\mathbf{z}^*$ according to the notion of our partner selection criterion, the only possibility is ${z^*}_{iu} = 1$ and $\tilde{z}_{iu} = 0, \exists_i$. Since the u^{th} node is the best partner for the i^{th} node in $\mathbf{z}^*$, the utility achieved by the i^{th} node will decrease with a different partner selection solution, i.e., $U_i(\tilde{\mathbf{z}}) < U_i(\mathbf{z}^*), \exists_i$.

According to the definition of Pareto optimality, with fixed subcarrier allocation and power allocation, the partner selection solution $\mathbf{z}^*$ obtained from our proposed distributed approach achieves the Pareto optimality. ■

Proposition 14 *With fixed subcarrier allocation and power allocation, the partner allocation solution from the proposed centralized resource allocation is Pareto optimal.*

Proof: Denote $S_j(\mathbf{z}^*)$ as the utility function of the j^{th} node. Denote $\mathbf{z}^*$ be the partner allocation solution obtained from the proposed centralized approach and $\tilde{\mathbf{z}}$ another solution. If $S_j(\tilde{\mathbf{z}}) > S_j(\mathbf{z}^*), \exists_j$ as $\tilde{z}_{mj} = 1, \exists_m$, then $S_u(\tilde{\mathbf{z}}) < S_u(\mathbf{z}^*), \exists_u$, as ${z^*}_{mu} = 1$ and $\tilde{z}_{mu} = 0$. Since the m^{th} node is the best source node for the u^{th} node to assist in $\mathbf{z}^*$, changing the partner allocation solution will decrease the utility gain of the u^{th} node and therefore, $S_u(\tilde{\mathbf{z}}) < S_u(\mathbf{z}^*), \exists_u$. Notice that the case of $\tilde{z}_{mj} = 1$ and $\sum_{u \neq m} {z^*}_{mu} = 0, \exists_m$, is not possible because, should the j^{th} node be the partner of the m^{th} node in $\tilde{\mathbf{z}}$, the j^{th} node can be the partner of the m^{th} node in $\mathbf{z}^*$ based on the notion of our partner allocation criterion. Therefore, according to the definition of Pareto optimality, with fixed subcarrier allocation and power allocation, the partner allocation solution $\mathbf{z}^*$ obtained from our proposed centralized approach achieves the Pareto optimality. ■

Proposition 15 *Provided that the channels are sensed according to the descending order of their achievable rates, a secondary user can achieve the maximal reward by stopping at the first sensed free channel.*

Proof: To prove Proposition 15 is equivalent to show $\Lambda_1 > \Lambda_2 > ... > \Lambda_N$. We prove $\Lambda_1 > \Lambda_2 > ... > \Lambda_N$ by (backward) mathematical induction.

For $k = N - 1$,

$$\Lambda_{N-1} = \theta_{s_{N-1}} c_{N-1} R_{s_{N-1}} + (1 - \theta_{s_{N-1}}) \Lambda_N \tag{B-26}$$

$$= \theta_{s_{N-1}} (c_{N-1} R_{s_{N-1}} - \Lambda_N) + \Lambda_N \tag{B-27}$$

$$= \theta_{s_{N-1}} (c_{N-1} R_{s_{N-1}} - \theta_{s_N} c_N R_{s_N}) + \Lambda_N \tag{B-28}$$

$$> \Lambda_N \quad (\because \theta_{s_N} \leq 1,\ c_{N-1} > c_N, \text{ and } R_{s_{N-1}} \geq R_{s_N}) \tag{B-29}$$

Therefore, the statement is true for $k = N - 1$.

Assuming that the statement is true for $k = i$, i.e., $\Lambda_i > \Lambda_{i+1}$ (and hence $c_i R_{s_i} > \Lambda_{i+1}$).

For $k = i - 1$, we have

$$\begin{aligned}
&\Lambda_{i-1} - \Lambda_i \\
&= \theta_{s_{i-1}}(c_{i-1}R_{s_{i-1}} - \Lambda_i) && \text{(B-30)} \\
&= \theta_{s_{i-1}}\left[c_{i-1}R_{s_{i-1}} - (\theta_{s_i}c_iR_{s_i} + (1-\theta_{s_i})\Lambda_{i+1})\right] && \text{(B-31)} \\
&= \theta_{s_{i-1}}\left(c_{i-1}R_{s_{i-1}} - \theta_{s_i}c_iR_{s_i} - (1-\theta_{s_i})\Lambda_{i+1}\right) && \text{(B-32)} \\
&> \theta_{s_{i-1}}\left(c_iR_{s_i} - \theta_{s_i}c_iR_{s_i} - (1-\theta_{s_i})\Lambda_{i+1}\right) \\
&\qquad\qquad (\because c_{i-1} > c_i \text{ and } R_{s_{i-1}} \geq R_{s_i}) && \text{(B-33)} \\
&= \theta_{s_{i-1}}(1-\theta_{s_i})(c_iR_{s_i} - \Lambda_{i+1}) && \text{(B-34)} \\
&> 0 \quad (\because \text{by the assumption}) && \text{(B-35)}
\end{aligned}$$

Therefore, the statement is also true for $k = i - 1$.

By mathematical induction, the statement $\Lambda_i > \Lambda_{i+1}$ is true for $1 \leq i < N$ and therefore, $\Lambda_1 > \Lambda_2 > ... > \Lambda_N$. Given that the channels are sensed according to the descending order of their achievable rates, a secondary user can achieve the maximal reward by stopping at the first sensed free channel. ■

Proposition 16 *Provided that $\theta_{s_i} = \theta_{s_j}, \forall i, j$, and the channels are sensed according to the descending order of their achievable transmission rates, the proposed instinctive sensing is optimal in terms of throughput (i.e., Λ_1).*

Proof: We prove this proposition by contradiction. Let $(\tilde{s}_1, ..., \tilde{s}_i, ..., \tilde{s}_j, ..., \tilde{s}_N)$ be the proposed sensing order with the expected reward $\tilde{\Lambda}_1$. Suppose there is another sensing order $(\hat{s}_1, ..., \hat{s}_i, ..., \hat{s}_j, ..., \hat{s}_N)$ with the expected reward $\hat{\Lambda}_1$, where $\tilde{s}_k = \hat{s}_k, \forall k \neq i, j$, $\tilde{s}_i = \hat{s}_j$, $\tilde{s}_j = \hat{s}_i$,

and $\hat{\Lambda}_1 > \tilde{\Lambda}_1$. Let $\theta_{s_k} = \theta, \forall k > 0$. Here, consider

$$\begin{aligned}
&\hat{\Lambda}_1 - \tilde{\Lambda}_1 \\
&= \sum_{k=1}^{N} \left[\prod_{m=1}^{k} \left(1 - \theta_{\hat{s}_{m-1}}\right) \right] \theta_{\hat{s}_k} c_k R_{\hat{s}_k} - \sum_{k=1}^{N} \left[\prod_{m=1}^{k} \left(1 - \theta_{\tilde{s}_{m-1}}\right) \right] \theta_{\tilde{s}_k} c_k R_{\tilde{s}_k} && \text{(B-36)} \\
&= \left[\prod_{m=1}^{i} (1-\theta) \right] \theta c_i \left(R_{\hat{s}_i} - R_{\tilde{s}_i}\right) + \left[\prod_{m=1}^{j} (1-\theta) \right] \theta c_j \left(R_{\hat{s}_j} - R_{\tilde{s}_j}\right) && \text{(B-37)} \\
&= (1-\theta)^i \theta c_i \left(R_{\tilde{s}_j} - R_{\tilde{s}_i}\right) + (1-\theta)^j \theta c_j \left(R_{\tilde{s}_i} - R_{\tilde{s}_j}\right) && \text{(B-38)} \\
&= (1-\theta)^i \theta \left(R_{\tilde{s}_i} - R_{\tilde{s}_j}\right) \left((1-\theta)^{j-i} c_j - c_i\right). && \text{(B-39)}
\end{aligned}$$

Since $\hat{\Lambda}_1 > \tilde{\Lambda}_1$, $0 \leq \theta \leq 1$, and $c_i > c_j$, $R_{\tilde{s}_i} < R_{\tilde{s}_j}$ for $i < j$. However, it contradicts to our proposed instinctive sensing that the channels are sensed according to the descending order of their achievable transmission rates. Therefore, no such a sensing order $(\hat{s}_1, ..., \hat{s}_i, ..., \hat{s}_j, ..., \hat{s}_N)$ with the expected reward $\hat{\Lambda}_1$ exists. In other words, the throughput or expected reward $\tilde{\Lambda}_1$ procured according to the proposed instinctive sensing order is indeed maximal. Therefore, our instinctive channel sensing achieves optimality in terms of throughput. ■

Proposition 17 *The set of feasible weighted utilities (i.e., $w_m Q_m(\mathbf{a}), \forall m$) in the WMMFOP has the solidarity property [93].*

Proof: Suppose that the feasible resource allocation solution is $\mathbf{a}^*$. Without loss of generality, we assume that there exists an n value such that $Q_n(\mathbf{a}^*) > 0$ (i.e., $a_n^* > 0$), where $n \neq m$ [93]. For a particular timeslot t, over which the m^{th} and n^{th} links are to be active, given the resource allocation solution $\mathbf{a}^*$, the values of their weighted utilities are to be $w_m Q_m(\mathbf{a}^*)$ and $w_n Q_n(\mathbf{a}^*)$, respectively. Let l_t denote the length of the timeslot t, i.e., $l_t > 0$. Similar to [93], it is possible to partition the slot into three minislots, namely t_1, t_2, and t_3 with positive durations l_{t_1}, l_{t_2}, and l_{t_3}, respectively, such that $l_{t_1} + l_{t_2} + l_{t_3} = l_t$. Note that the choice of how to determine these values is arbitrary. During t_1, the resource allocation $\mathbf{a}$ is chosen to be the same as that in timeslot t. During t_2, the same allocation solution is kept

as in t_1, except $a_n = 0$. During t_3, we set $a_n = 0$, and a_m is adjusted until the interference experienced by other active links is larger than that in the original timeslot t, if possible, otherwise, we set $a_m = 1$.

In this new resource allocation, compared with the weighted utilities obtained in the original resource allocation in timeslot t, all the links excluding the m^{th} and n^{th} links have the same or higher weighted utilities in t_1 and t_2, respectively. In t_3, their weighted utilities can be higher, the same, or lower, depending on the value of a_m. Since the partitioning into t_1, t_2, and t_3 is entirely arbitrary, it is possible to choose their lengths l_{t_1}, l_{t_2}, and l_{t_3} so that there exist small ϵ_m (> 0) and ϵ_n (> 0) such that the weighted utility of the m^{th} link increases by at most ϵ_m and the weighted utility of the n^{th} link decreases by at most ϵ_n, while the rest of the active links have the same or higher weighted utilities.

Let $w_m Q_m(\tilde{\mathbf{a}})$ and $w_n Q_n(\tilde{\mathbf{a}})$ denote the newly obtained weighted utilities of the m^{th} link and the n^{th} link, respectively. We can now acquire the following inequalities: $w_m Q_m(\mathbf{a}^*) < w_m Q_m(\tilde{\mathbf{a}}) < w_m Q_m(\mathbf{a}^*) + \epsilon_m$ and $w_n Q_n(\mathbf{a}^*) - \epsilon_n < w_n Q_n(\tilde{\mathbf{a}}) < w_n Q_n(\mathbf{a}^*)$. Therefore, the value of $w_m Q_m(\mathbf{a})$ can be increased by at most ϵ_m by decreasing the value of $w_n Q_n(\mathbf{a})$ by at most ϵ_n. And,

$$\begin{bmatrix} \vdots \\ w_m Q_m(\tilde{\mathbf{a}}) \\ w_n Q_n(\tilde{\mathbf{a}}) \\ \vdots \\ w_k Q_k(\tilde{\mathbf{a}}) \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ w_m Q_m(\mathbf{a}^*) \\ w_n Q_n(\mathbf{a}^*) \\ \vdots \\ w_k Q_k(\mathbf{a}^*) \\ \vdots \end{bmatrix} - \begin{bmatrix} \vdots \\ 0 \\ \epsilon_n \\ \vdots \\ 0 \\ \vdots \end{bmatrix} + \begin{bmatrix} \vdots \\ \epsilon_m \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix} + \sum_{k \neq m,n} \begin{bmatrix} \vdots \\ 0 \\ 0 \\ \vdots \\ \epsilon_k \\ \vdots \end{bmatrix} \tag{B-40}$$

where $\epsilon_k > 0, \forall k \neq m, n$. Since all $w_m Q_m(\tilde{\mathbf{a}})$, $\forall_m$, belong to the feasible set (i.e., $Q_m(\tilde{\mathbf{a}}) \geq 0$, $\forall_m$), by the definition of solidarity [93], the set of feasible weighted utilities in the WMMFOP has the solidarity property. ■

Proposition 18 *The optimal solution* $\hat{\mathbf{a}}$ *obtained from the WMMFOP is also the optimal solution for the modified STOP.*

Proof: Suppose that there exists another solution $\tilde{\mathbf{a}}$ such that $\sum_{m=1}^{M} Q_m(\tilde{\mathbf{a}}) > \sum_{m=1}^{M} Q_m(\hat{\mathbf{a}})$. It means that there exists some m such that $w_m Q_m(\tilde{\mathbf{a}}) > J^*$ within the feasible region. In the WMMFOP, if for some m, $w_m Q_m(\tilde{\mathbf{a}}) > J^*$, J^* can be increased by decreasing the value of $\tilde{a}_m$ or increasing the value of $\tilde{a}_n$ or both, for $n \neq m$, within the feasible region until it reaches the maximal value, say $\tilde{J}$. However, it contradicts to the statement that J^* is the maximal value obtained from the WMMFOP. Therefore, no such a solution $\tilde{\mathbf{a}}$ exists. The optimal solution $\hat{\mathbf{a}}$ obtained from the WMMFOP is also the optimal solution for the modified STOP. ■

Proposition 19 *The system throughput (i.e.,* $\sum_{m=1}^{M} Q_m(\mathbf{a}^*)$*) is a non-increasing function of bargaining floor* J*.*

Proof: When J increases (decreases), the feasible region of $\mathbf{a}$ in the GOP shrinks (expands). For $0 \leq J_1 \leq J_2 \leq J^*$, let $\mathbf{a_1^*}$ and $\mathbf{a_2^*}$ denote the optimal solutions of the GOP with J_1 and J_2, respectively. The feasible region of $\mathbf{a}$ of the GOP with J_2 is only a subset of that with J_1. Thus, $\sum_{m=1}^{M} Q_m(\mathbf{a_1^*}) \geq \sum_{m=1}^{M} Q_m(\mathbf{a_2^*})$ and hence the system throughput does not increase with the value of J. ■

Corollary 3 *The minimum value of* $w_m Q_m(\mathbf{a}^*)$ *(i.e.,* $\min_m \{w_m Q_m(\mathbf{a}^*)\}$*) is a non-decreasing function of* J*.*

Proof: Let J^* be the solution of WMMFOP and $\mathbf{a_1^*}$ be the optimal solution of the GOP with J_1, where $0 \leq J_1 \leq J^*$. For $0 \leq J_1 < J_2 \leq J^*$, consider the following two cases:

Case 1: If the solution $\mathbf{a_1^*}$ is feasible for the GOP with J_2, from Proposition 19, $\mathbf{a_1^*}$ is also the optimal solution for the GOP with J_2. Hence, the minimum value of $w_m Q_m(\mathbf{a}^*)$ is the same for the GOP with J_1 and J_2;

Case 2: If the solution $\mathbf{a_1^*}$ is not feasible for the GOP with J_2, it means that there exists some m such that $w_m Q_m(\mathbf{a_1^*}) < J_2$ and hence $\min_m \{w_m Q_m(\mathbf{a_1^*})\} < J_2$. Thus, $\mathbf{a_1^*}$ is an infeasible solution for the GOP with J_2. In addition, suppose that $\mathbf{a_2^*}$ is the optimal solution for the GOP with J_2, i.e., $w_m Q_m(\mathbf{a_2^*}) \geq J_2, \forall m$. Hence, $\min_m \{w_m Q_m(\mathbf{a_2^*})\} \geq J_2 > \min_m \{w_m Q_m(\mathbf{a_1^*})\}$.

By combining the above two cases, $\min_m \{w_m Q_m(\mathbf{a_2^*})\} \geq \min_m \{w_m Q_m(\mathbf{a_1^*})\}$ and hence the minimum value of $w_m Q_m(\mathbf{a}^*)$ (i.e., $\min_m \{w_m Q_m(\mathbf{a}^*)\}$) does not decrease with the value of J. ■

Theorem 1 *The system throughput (i.e., $\sum_{m=1}^{M} Q_m(\mathbf{a}^*)$) does not increase with J, but the minimum value of $w_m Q_m(\mathbf{a}^*)$ (i.e., $\min_m \{w_m Q_m(\mathbf{a}^*)\}$) does not decrease with J.*

Proof: By Proposition 19 and Corollary 3, it is proved. ■

Corollary 4 *A relationship between the system throughput and weighted max-min fairness performance can be achieved by solving the GOP with different values of J.*

Proof: From Theorem 1, the solution obtained from the GOP with $J = 0$ corresponds to the maximal system throughput, while the solution obtained from the GOP with $J = J^*$ corresponds to the maximal weighted max-min fairness performance. When J increases from zero, the solution obtained from the GOP can refer to decreased system throughput and increased fairness performance. Therefore, the performance tradeoff and hence a desired relationship between the system throughput and weighted max-min fairness performance can be achieved by solving the GOP with different values of J, i.e., $J \in [0, J^*]$. ■

Proposition 20 *The optimal solution* $\mathbf{a}$ *of the GOP is Pareto optimal.*

Proof: Given the action profile or optimal solution $\mathbf{a}^*$ (i.e., resource allocation) obtained from the GOP, denote $Q_m(\mathbf{a}^*)$ or $w_m Q_m(\mathbf{a}^*)$ as the utility function of the m^{th} link. Consider another action profile $\tilde{\mathbf{a}}$. For some m, if $Q_m(\tilde{\mathbf{a}}) > Q_m(\mathbf{a}^*)$, then $Q_n(\tilde{\mathbf{a}}) < Q_n(\mathbf{a}^*)$ for some n, as either $\tilde{a}_m > a_m^*$ or $\tilde{a}_n < a_n^*$ or both. According to the definition of Pareto optimality, the optimal solution $\mathbf{a}^*$ obtained from the GOP is Pareto optimal. ■

Appendix C

List of Acronyms

AF	amplify-and-forward
ATM	asynchronous transfer mode
BER	bit error rate
CAC	call admission control
CBR	constant-bit-rate
CDMA	code division multiple access
CRN	cognitive radio network
DCA	dynamic channel allocation
DF	decode-and-forward
DSL	digital subscriber line
FCA	fixed channel allocation
FIFO	first-in-first-out

GA	genetic algorithm
GPS	generalized processor sharing
HCA	hybrid channel allocation
IPTV	Internet Protocol Television
ISI	inter-symbol interference
ITU	International Telecommunications Union
KKT	Karush-Kuhn-Tucker
MAC	medium access control
MANET	mobile ad hoc network
MIMO	multiple-input-multiple-output
NE	Nash equilibrium
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
PLR	packet loss rate
QoS	quality-of-service
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
UWB	ultra-wideband
VoIP	voice over IP

WLAN wireless local area network

WMN wireless mesh network

Bibliography

[1] A. Abdrabou and W. Zhuang, "A position-based QoS routing scheme for UWB mobile ad hoc networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 4, pp. 850–856, Apr. 2006.

[2] A. Adas, "Traffic models in broadband networks," *IEEE Commun. Mag.*, vol. 35, no. 7, pp. 82–89, July 1997.

[3] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, "Analysis of a mixed-use urban WiFi network: when metropolitan becomes neapolitan," in *Proc. ACM SIGCOMM*, pp. 85–98, 2008.

[4] I. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Computer Networks*, vol. 47, no. 4, pp. 445–487, Mar. 2005.

[5] M. Alicherry, R. Bhatia, and L. E. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 1960-1971, Nov. 2006.

[6] S. Banerjee and S. Khuller, "A clustering scheme for hierarchical control in multi-hop wireless networks," in *Proc. IEEE INFOCOM '01*, vol. 2, pp. 1028–1037, 2001.

[7] BelAir Networks. [Online]. Available: http://www.belairnetworks.com

[8] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

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

[10] V. Brik, S. Rayanchu, S. Saha, S. Sen, V. Shrivastava, and S. Banerjee, "A measurement study of a commercial-grade urban WiFi mesh," in *Proc. ACM SIGCOMM*, pp. 111–124, 2008.

[11] R. Bruno, M. Conti, and E. Gregori, "Mesh networks: commodity multihop ad hoc networks," *IEEE Commun. Mag.*, vol. 43, no. 3, pp. 123–134, Mar. 2005.

[12] J. Cai, X. Shen, J. W. Mark, and A. S. Alfa, "Semi-distributed user relaying algorithm for amplify-and-forward wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1348–1357, Apr. 2008.

[13] Y.-Y. Chen, S.-C. Liu, and C. Chen, "Channel assignment and routing for multichannel wireless mesh networks using simulated annealing," in *Proc. IEEE Globecom '06*, pp. 1–5, Nov. 2006.

[14] Y. Chen and S. Kishore, "A game-theoretic analysis of decode-and-forward user cooperation," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1941–1951, May 2008.

[15] H. T. Cheng and T. M. Lok, "Detection schemes for distributed space-time block coding in time-varying wireless cooperative systems," in *Proc. IEEE Tencon '05*, pp. 289–293, Nov. 2005.

[16] H. T. Cheng, H. Mheidat, M. Uysal, and T. M. Lok, "Distributed space-time block coding with imperfect channel estimation," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 1, pp. 583–587, May 2005.

[17] H. T. Cheng, A. Abdrabou, and W. Zhuang, "Novel resource management approach for end-to-end QoS support in wireless mesh networks," accepted for presentation in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2010.

[18] H. T. Cheng, H. Jiang, and W. Zhuang, "Distributed medium access control for wireless mesh networks," *Wireless Communications and Mobile Computing*, vol. 6, no. 6, pp. 845–864, Sept. 2006.

[19] H. T. Cheng and W. Zhuang, "Simple channel sensing order in cognitive radio networks," *IEEE J. Select. Areas Commun.*, submitted.

[20] ——, "QoS-Driven MAC-Layer resource allocation for wireless mesh networks with non-altruistic node cooperation and service differentiation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6089–6103, Dec. 2009.

[21] ——, "Joint power-frequency-time resource allocation in clustered wireless mesh networks," *IEEE Network*, vol. 22, no. 1, pp. 45–51, Jan.-Feb. 2008.

[22] ——, "An optimization framework for balancing throughput and fairness in wireless networks with QoS support," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 584–593, Feb. 2008.

[23] ——, "Novel packet-level resource allocation with effective QoS provisioning for wireless mesh networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 694–700, Feb. 2009.

[24] ——, "Pareto optimal resource management for wireless mesh networks with QoS assurance: joint node clustering and subcarrier allocation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1573–1583, Mar. 2009.

[25] ——, "On packet-level non-altruistic node cooperation in wireless networks," accepted for presentation in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2010.

[26] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 104–116, Jan. 2005.

[27] M. Clouqueur and W. D. Grover, "Availability analysis of span-restorable mesh networks," *IEEE J. Select. Areas Commun.*, vol. 20, no. 4, pp. 810–821, May 2002.

[28] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[29] J. Damnjanovic, A. Jain, T. Chen, and S. Sarkar, "Scheduling the CDMA2000 reverse link," in *Proc. IEEE Vehicular Technology Conference (VTC)*, vol. 1, pp. 386–390, 2002.

[30] A. K. Das, H. M. K. Alazemi, R. Vijayakumar, and S. Roy, "Optimization models for fixed channel assignment in wireless mesh networks with multiple radios," in *Proc. IEEE SECON '2005*, pp. 463– 474, Sept. 2005.

[31] M. K. Denko, "Using mobile internet gateways in wireless mesh networks," in *Proc. AINA '08*, pp. 1086–1092, Mar. 2008.

[32] T. Ferguson, "Optimal stopping and applications." [Online]. Available: www.math.ucla.edu/∼tom/Stopping/Contents.html/

[33] L. Gao and S. Cui, "Efficient subcarrier, power, and rate allocation with fairness consideration for OFDMA uplink," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1507–1511, May 2008.

[34] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman & Co., 1979.

[35] R. Guérin and V. Peris, "Quality-of-service in packet networks: basic mechanisms and directions," *Computer Networks*, vol. 31, no. 3, pp. 169–189, 1999.

[36] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[37] A. Hasan, K. Yang, and J. Andrews, "Clustered CDMA ad hoc networks without closed-loop power control," in *Proc. IEEE MILCOM '03*, vol. 2, pp. 1030–1035, Oct. 2003.

[38] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Select. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[39] D. P. Heyman, "The GBAR source model for VBR videoconferences," *IEEE/ACM Trans. Networking*, vol. 5, no. 4, pp. 554–560, Aug. 1997.

[40] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.

[41] J. Huang, Z. Han, M. Chiang, and H. V. Poor, "Auction-based resource allocation for cooperative communications," *IEEE J. Select. Areas Commun.*, vol. 26, no. 7, pp. 1226–1237, Sept. 2008.

[42] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[43] C. Ibars, A. del Coso, Y. Grunenberger, F. Theoleyre, and F. Rousseau, "Increasing the throughput of wireless mesh networks with cooperative techniques," in *Proc. IST Mobile and Wireless Communications Summit*, pp. 1–5, 2007.

[44] IEEE 802.16 Broadband Wireless Access Working Group, "Channel models for fixed wireless applications," 2003. [Online]. Available: http://www.ieee802.org/16

[45] J. Ishmael, S. Bury, D. Pezaros, and N. Race, "Deploying rural community wireless mesh networks," *IEEE Internet Computing*, vol. 12, no. 4, pp. 22–29, Jul.-Aug. 2008.

[46] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless Networks*, vol. 11, no. 4, pp. 471–487, July 2005.

[47] R. Jain, A. Durresi, and G. Babic, "Throughput fairness index: an explanation," *ATM Forum Document Number: ATM Forum/990045*, Feb. 1999.

[48] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 362–371, Feb. 2004.

[49] J. Janssen, D. De Vleeschauwer, M. Buchli, and G. H. Petit, "Assessing voice quality in packet-based telephony," *IEEE Internet Comput.*, vol. 6, pp. 1135–1140, May.-Jun. 2002.

[50] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Select. Areas Commun.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.

[51] H. Jiang, L. Lai, R. Fan, and H. Poor, "Optimal selection of channel sensing order in cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 297–307, Jan. 2009.

[52] B. Johanson, A. Fox, and T. Winograd, "The interactive workspaces project: experiences with ubiquitous computing rooms," *IEEE Pervasive Computing*, vol. 1, no. 2, pp. 67–74, Apr.-Jun. 2002.

[53] P. Johri, "An insight into dynamic channel assignment in cellular mobile communications systems," *Euro J. Operational Research*, vol. 74, pp. 70–77, 1994.

[54] D. Julian, M. Chiang, D. O'Neill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks," in *Proc. IEEE INFOCOM '02*, vol. 2, pp. 477–486, June 2002.

[55] R. Jurdak, C. V. Lopes, and P. Baldi, "A survey, classification and comparative analysis of medium access control protocols for ad hoc networks," *IEEE Communications Surveys and Tutorials*, vol. 6, no. 1, 2004.

[56] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[57] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunicationsystems: a comprehensive survey," *IEEE Personal Communications*, vol. 3, no. 3, pp. 10–31, 1996.

[58] F. P. Kelly, "Notes on effective bandwidths," in *Stochastic Networks: Theory and Applications*. Oxford University Press, 1996.

[59] ——, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[60] F. P. Kelly, A. Maulloo, and D. K. H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[61] B. S. Kim, Y. Fang, T. F. Wong, and Y. Kwon, "Throughput enhancement through dynamic fragmentation in wireless LANs," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1415–1425, July 2005.

[62] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM MobiCom '05*, pp. 73–87, Sept. 2005.

[63] L. Lai, K. Liu, and H. El Gamal, "The three-node wireless networks: achievable rates and cooperation strategies," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 805–828, Mar. 2006.

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

[65] K. Lee, "Supporting mobile multimedia in integrated services networkings," *Wireless Networks*, vol. 2, no. 3, pp. 205–217, 1996.

[66] F. Li, Y. Wang, and X.-Y. Li, "Gateway placement for throughput optimization in wireless mesh networks," in *Proc. IEEE International Conference on Communications (ICC)*, pp. 4955–4960, June 2007.

[67] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[68] C. R. Lin and M. Gerla, "Adaptive clustering for mobile wireless networks," *IEEE J. Select. Areas Commun.*, vol. 15, no. 7, pp. 1265–1275, Sept. 1997.

[69] S. H. Low and D. E. Lapsely, "Optimization flow control I: basic algorithm and convergence," *IEEE/ACM Trans. Networking*, vol. 7, no. 6, pp. 861–874, Dec. 1999.

[70] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Trans. Networking*, vol. 7, no. 4, pp. 473–489, Aug. 1999.

[71] H. Luo, J. Cheng, and S. Lu, "Self-coordinating localized fair queueing in wireless ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 1, pp. 86–98, Jan.-Mar. 2004.

[72] V. Mahinthan, L. Cai, J. W. Mark, and X. Shen, "Partner selection based on optimal power allocation in cooperative-diversity systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 511–520, Jan. 2008.

[73] J. Mark and W. Zhuang, *Wireless Communications and Networking*. Prentice Hall, 2003.

[74] MeshDynamics. [Online]. Available: http://www.meshdynamics.com

[75] Mesh@Purdue. [Online]. Available: https://engineering.purdue.edu/MESH

[76] Microsoft. [Online]. Available: http://research.microsoft.com/mesh

[77] G. Miklos and S. Molna, "Fair allocation of elastic traffic for a wireless base station," in *Proc. IEEE Globecom '99*, vol. 3, pp. 1673–1678, 1999.

[78] MIT Roofnet. [Online]. Available: http://pdos.csail.mit.edu/roofnet

[79] M. Mitchell, *An Introduction to Genetic Algorithms*. The MIT Press, Cambridge, MA, 1998.

[80] A. H. Mohsenian Rad and V. W. S. Wong, "Joint logical topology design, interface assignment, channel allocation, and routing for multi-channel wireless mesh networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 12, p. 4432-4440, Dec. 2007.

[81] A. Mukherjee and H. M. Kwon, "Robust auction-theoretic partner selection in cooperative diversity wireless networks," in *Proc. ACSSC '07*, pp. 443–447, Nov. 2007.

[82] R. U. Nabar, H. Blcskei, and F. W. Kneubhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[83] K. Navaie, D. Y. Montuno, and Y. Q. Zhao, "Fairness of resource allocation in cellular networks: A survey," *Resource Allocation in Next Generation Wireless Networks*, vol. 5, 2005.

[84] M. Neufeld, J. Fifield, C. Doerr, A. Sheth, and D. Grunwald, "SoftMAC–flexible wireless research platform," in *Proc. ACM SIGCOMM*, pp. 1–5, Nov. 2005.

[85] C. Y. Ng and C. W. Sung, "Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1667–1675, May 2008.

[86] Nokia LMA, "Sharing study: Part A - last mile access at 5GHz spectrum co-existence study," Oct. 2003.

[87] A. Nosratinia and T. E. Hunter, "Grouping and partner selection in cooperative wireless networks," *IEEE J. Select. Areas Commun.*, vol. 25, no. 2, pp. 369–378, Feb. 2007.

[88] G. Owen, *Game Theory, 3rd Ed.* Academic Press, 2001.

[89] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single-node case," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 344–357, June 1993.

[90] ——, "A generalized processor sharing approach to flow control in integrated services networks: the multiple node case," *IEEE/ACM Trans. Networking*, vol. 2, no. 2, pp. 137–150, 1994.

[91] V. Paxson and S. Floyd, "Wide area traffic: the failure of poisson modeling," *IEEE/ACM Trans. Networking*, vol. 3, no. 3, pp. 226–244, June 1995.

[92] J. B. Punt, D. Sparreboom, and F. Brouwer, "Mathematical models for the analysis of dynamic channel selection for indoor mobile wireless communication systems," in *Proc. IEEE PIMRC*, vol. 4, pp. 1081–1085, Sept. 1994.

[93] B. Radunović and J. Y. Le Boudec, "Rate performance objectives of multihop wireless networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 334–349, Oct.-Dec. 2004.

[94] H. Rasouli, S. Sadr, and A. Anpalagan, "A fair subcarrier allocation algorithm for cooperative multiuser OFDM systems with grouped users," in *Proc. IEEE Globecom '08*, 2008.

[95] A. Sabharwal, A. Khoshnevis, and E. Knightly, "Opportunistic spectral usage: Bounds and a multi-band CSMA/CA protocol," *IEEE/ACM Trans. Networking*, vol. 15, no. 3, pp. 533–545, June 2007.

[96] H. Shan, W. Zhuang, and Z. Wang, "Distributed cooperative MAC for multi-hop wireless networks," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 126–133, Feb. 2009.

[97] C. E. Shannon, "A mathematical theory of communications," *Bell Labs Technical Journal*, vol. 27, pp. 623–656, 1948.

[98] J. Shi and A. Hu, "Maximum utility-based resource allocation algorithm in the IEEE 802.16 OFDMA system," in *Proc. IEEE Globecom '08*, pp. 311–316, May 2008.

[99] Y. Shi, S. Sharma, Y. T. Hou, and S. Kompella, "Optimal relay assignment for cooperative communications," in *Proc. ACM MobiHoc '08*, pp. 3–12, 2008.

[100] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-Part I: Theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614–624, Mar. 2005.

[101] ——, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Commun. Mag.*, vol. 43, no. 12, pp. 127–134, Dec. 2005.

[102] X. Song, Z. He, K. Niu, and W. Wu, "A hierarchical resource allocation for OFDMA distributed wireless communication systems," in *Proc. IEEE Globecom '07*, pp. 5195–5199, Nov. 2007.

[103] G. Staple and K. Werbach, "The end of spectrum scarcity," *IEEE Spectr.*, vol. 41, no. 3, pp. 48–52, 2004.

[104] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge University Press, 2005.

[105] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications.* Artech House, Inc. Norwood, MA, USA, 2000.

[106] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[107] H. M. Wagner, *Principles of Operations Research.* Prentice-Hall, Inc., 1975.

[108] B. Wang, K. C. Chua, and H.-P. Schwefel, "Resource allocation for non-realtime data users in UMTS uplink," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, vol. 2, pp. 1055–1059, Mar. 2003.

[109] B. Wang, Z. Han, and K. J. R. Liu, "Distributed relay selection and power control for multiuser cooperative communication networks using buyer/seller game," in *Proc. IEEE INFOCOM '07*, pp. 544–552, May 2007.

[110] T. C. Wang, S. Y. Chang, and H.-C. Wu, "Novel cooperative transmission using matching algorithm for wimax mesh networks," in *Proc. IWCMC*, pp. 655–660, Aug. 2008.

[111] Y. Wang, S. Jain, M. Martonosi, and K. Fall, "Erasure-coding based routing for opportunistic networks," in *Proc. ACM SIGCOMM WDTN*, pp. 229–236, 2005.

[112] M. A. Weiss, *Data Structures and Algorithm Analysis in C*. Addision Wesley, 1997.

[113] K. Whitehouse, A. Woo, F. Jiang, J. Polastre, and D. Culler, "Exploiting the capture effect for collision detection and recovery," *IEEE EmNetS-II*, pp. 45–52, May 2005.

[114] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[115] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, July 2003.

[116] A. Yener and S. Kishore, "Distributed power control and routing for clustered CDMA wireless ad hoc networks," in *Proc. IEEE Vehicular Technology Conference (VTC2004-Fall)*, vol. 4, pp. 2951–2955, Sept. 2004.

[117] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM-based broadband wireless systems," in *Proc. IEEE Globecom '00*, vol. 1, pp. 103–107, 2000.

[118] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: recent developments and deployment challenges," *IEEE Network*, vol. 20, no. 3, pp. 20–25, May-June 2006.

[119] J. Y. Yu and P. Chong, "A survey of clustering schemes for mobile ad hoc networks," *IEEE Communications Surveys and Tutorials*, vol. 7, no. 1, pp. 32–48, First Qtr. 2005.

[120] Q.-Q. Zhang, W.-D. Gao, M.-G. Peng, and W.-B. Wang, "Partner selection strategies in cooperative wireless networks with optimal power distribution," *The Journal of China Universities of Posts and Telecommunications*, vol. 15, no. 3, pp. 47–50,58, 2008.

[121] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1566–1575, Sept. 2004.

[122] Z. Zhang, J. Shi, H.-H. Chen, M. Guizani, and P. Qiu, "A cooperation strategy based on nash bargaining solution in cooperative relay networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2570–2577, July 2008.

[123] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Select. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.

[124] Y. Zhao, R. S. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," in *Proc. IEEE ISIT*, pp. 1234–1238, July 2006.