Algorithms and Optimization for Wireless Networks

Yi Shi

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Engineering

Prof. Y. Thomas Hou, Chair
Prof. R. Michael Buehrer
Prof. Scott F. Midkiff
Prof. Hanif D. Sherali
Prof. Yaling Yang

October 25, 2007
Blacksburg, Virginia

Keywords: Wireless networks, optimization, algorithm, cross-layer design

Algorithms and Optimization for Wireless Networks

Yi Shi

ABSTRACT

Recently, many new types of wireless networks have emerged for both civil and military applications, such as wireless sensor networks, ad hoc networks, among others. To improve the performance of these wireless networks, many advanced communication techniques have been developed at the physical layer. For both theoretical and practical purposes, it is important for a network researcher to understand the performance limits of these new wireless networks. Such performance limits are important not only for theoretical understanding, but also in that they can be used as benchmarks for the design of distributed algorithms and protocols. However, due to some unique characteristics associated with these networks, existing analytical technologies may not be applied directly. As a result, new theoretical results, along with new mathematical techniques, need to be developed. In this dissertation, we focus on the design of new algorithms and optimization techniques to study theoretical performance limits associated with these new wireless networks.

In this dissertation, we mainly focus on sensor networks and ad hoc networks. Wireless sensor networks consist of battery-powered nodes that are endowed with a multitude of sensing modalities. A wireless sensor network can provide in-situ, unattended, high-precision, and real-time observation over a vast area. Wireless ad hoc networks are characterized by the absence of infrastructure support. Nodes in an ad hoc network are able to organize themselves into a multi-hop network. An ad hoc network can operate in a stand-alone fashion or could possibly be connected to a larger network such as the Internet (also known as mesh networks).

For these new wireless networks, a number of advanced physical layer techniques, e.g., ultra wideband (UWB), multiple-input and multiple-output (MIMO), and cognitive radio (CR), have been employed. These new physical layer technologies have the potential to improve network performance. However, they also introduce some unique design challenges. For example, CR is

capable of reconfiguring RF (on the fly) and switching to newly-selected frequency bands. It is much more advanced than the current multi-channel multi-radio (MC-MR) technology. MC-MR remains hardware-based radio technology: each radio can only operate on a single channel at a time and the number of concurrent channels that can be used at a wireless node is limited by the number of radio interfaces. While a CR can use multiple bands at the same time. In addition, an MC-MR based wireless network typically assumes there is a set of "common channels" available for all nodes in the network. While for CR networks, each node may have a different set of frequency bands based on its particular location. These important differences between MC-MR and CR warrant that the algorithmic design for a CR network is substantially more complex than that under MC-MR.

Due to the unique characteristics of these new wireless networks, it is necessary to consider models and constraints at multiple layers (e.g., physical, link, and network) when we explore network performance limits. The formulations of these cross-layer problems are usually in very complex forms and are mathematically challenging. We aim to develop some novel algorithmic design and optimization techniques that provide optimal or near-optimal solutions.

The main contributions of this dissertation are summarized as follows.

1. **Node lifetime and rate allocation** We study the sensor node lifetime problem by considering not only maximizing the time until the first node fails, but also maximizing the lifetimes for all the nodes in the network. For fairness, we maximize node lifetimes under the *lexicographic max-min* (LMM) criteria. Our contributions are two-fold. First, we develop a polynomial-time algorithm based on a parametric analysis (PA) technique, which has a much lower computational complexity than an existing state-of-the-art approach. We also present a polynomial-time algorithm to calculate the flow routing schedule such that the LMM-optimal node lifetime vector can be achieved. Second, we show that the same approach can be employed to address a different but related problem, called LMM rate allocation problem. More important, we discover an elegant *duality* relationship between the LMM node lifetime problem and the LMM rate allocation problem. We show that it is suf-

ficient to solve only one of the two problems and that important insights can be obtained by inferring the duality results.

2. **Base station placement** Base station location has a significant impact on sensor network lifetime. We aim to determine the best location for the base station so as to maximize the network lifetime. For a multi-hop sensor network, this problem is particularly challenging as data routing strategies also affect the network lifetime performance. We present an approximation algorithm that can guarantee $(1-\varepsilon)$-optimal network lifetime performance with any desired error bound $\varepsilon > 0$. The key step is to divide the continuous search space into a finite number of subareas and represent each subarea with a "fictitious cost point" (FCP). We prove that the largest network lifetime achieved by one of these FCPs is $(1-\varepsilon)$-optimal. This approximation algorithm offers a significant reduction in complexity when compared to a state-of-the-art algorithm, and represents the best known result to this problem.

3. **Mobile base station** The benefits of using a mobile base station to prolong sensor network lifetime have been well recognized. However, due to the complexity of the problem (time-dependent network topology and traffic routing), theoretical performance limits and provably optimal algorithms remain difficult to develop. Our main result hinges upon a novel transformation of the joint base station movement and flow routing problem from the time domain to the space domain. Based on this transformation, we first show that if the base station is allowed to be present only on a set of pre-defined points, then we can find the optimal sojourn time for the base station on each of these points so that the overall network lifetime is maximized. Based on this finding, we show that when the location of the base station is un-constrained (i.e., can move to any point in the two-dimensional plane), we can develop an approximation algorithm for the joint mobile base station and flow routing problem such that the network lifetime is *guaranteed* to be at least $(1-\varepsilon)$ of the maximum network lifetime, where ε can be made arbitrarily small. This is the first theoretical result with performance guarantee on this problem.

4. **Spectrum sharing in CR networks** Cognitive radio is a revolution in radio technology that

promises unprecedented flexibility in radio communications and is viewed as an enabling technology for dynamic spectrum access. We consider a cross-layer design of scheduling and routing with the objective of minimizing the required network-wide radio spectrum usage to support a set of user sessions. Here, scheduling considers how to use a pool of unequal size frequency bands for concurrent transmissions and routing considers how to transmit data for each user session. We develop a near-optimal algorithm based on a *sequential fixing* (SF) technique, where the determination of scheduling variables is performed iteratively through a sequence of linear programs (LPs). Upon completing the fixing of these scheduling variables, the value of the other variables in the optimization problem can be obtained by solving an LP.

5. **Power control in CR networks** We further consider the case of variable transmission power in CR networks. Now, our objective is minimizing the total required bandwidth footprint product (BFP) to support a set of user sessions. As a basis, we first develop an interference model for scheduling when power control is performed at each node. This model extends existing so-called protocol models for wireless networks where transmission power is deterministic. As a result, this model can be used for a broad range of problems where power control is part of the optimization space. An efficient solution procedure based on the branch-and-bound framework and convex hull relaxations is proposed to provide $(1-\varepsilon)$-optimal solutions. This is the first theoretical result on this important problem.

Acknowledgments

I would like to take this opportunity to acknowledge all the people that made a difference in my Ph.D. life at Virginia Tech.

First and foremost, I express my sincere gratitude to Prof. Y. Thomas Hou, for all his help, guidance, support, and encouragement. I can say without hesitation that he has made a profound influence on me as a teacher and researcher. I thank him for the countless hours he has spent with me, helping me to identify many interesting and important research topics during the past five years, sharpening my skills as a competent researcher, and helping me out in my research writing.

I also acknowledge the contributions and help given by the rest of my thesis committee. I especially thank Prof. Hanif D. Sherali for taking a genuine interest in my work and working with me throughout my Ph.D. years. Without his invaluable help and advice on my research, I could not have achieved so much progress over these years. I have learnt so much from him, and he is a role model for me in many ways. I also want to thank Prof. Scott F. Midkiff, for many advice and useful materials that he has given me. I thank Prof. R. Michael Buehrer for discussions on advanced wireless communication techniques, which helped me better understand physical layer issues. I thank Prof. Yaling Yang for her comments and suggestions on my thesis.

I want to thank Dr. Shiwen Mao for all his friendship and discussions. In addition, I would like to thank Xiaolin Cheng, Sastry Kompella, Jia Liu, Tong Liu, Xiaojun Wang, and many other graduate students for all their assistance. Their friendship has made my Ph.D. experience both fun and rewarding.

I thank my parents Qingxiao Shi and Tingyao Zhu for everything they have offered me in my life. They taught me the value of knowledge, the joy of love, and the importance of family. I always believe that it is important to have an early good start and right attitude on study even during childhood. I also want to thank my brother, Lei Shi, for all his support. Finally, I would like to thank my wife, Meiyu Zheng. Although she entered my life in my last year as Ph.D. student, her impact on my personal life cannot be simply stated in words. I thank her for all she has done to make my life full of joy.

Contents

1 Introduction **1**

1.1 Motivation and Objective 1

1.2 Summary of Contributions 3

1.2.1 Node Lifetime and Rate Allocation for Sensor Networks 3

1.2.2 Base Station Placement for Sensor Networks 4

1.2.3 Mobile Base Station for Sensor Networks 5

1.2.4 Optimal Spectrum Sharing for CR Networks 6

1.2.5 Optimal Power Control for CR Networks 8

1.3 Dissertation Outline 8

2 Node Lifetime and Rate Allocation Problems for Wireless Sensor Networks **11**

2.1 Introduction 11

2.2 System Modeling and Problem Formulation 14

2.2.1 Reference Network Architecture 14

2.2.2 Energy Model 17

2.2.3 The Lexicographic Max-Min Node Lifetime Problem 17

2.3 An Efficient Algorithm Based on Parametric Analysis 20
2.3.1 Link-based Formulation 22
2.3.2 Minimum Node Set Determination with Parametric Analysis 23
2.3.3 LMM-Optimal Routing Solution 27
2.3.4 Computational Complexity Analysis 31
2.4 Extension to LMM Rate Allocation Problem 33
2.4.1 The LMM Rate Allocation Problem 33
2.4.2 Solution 34
2.5 Duality Theorem 35
2.6 Numerical Investigation 37
2.6.1 Network Configurations and Parameter Settings 38
2.6.2 SLP-PA Algorithm to the LMM Node Lifetime Problem 38
2.6.3 Duality Results 45
2.7 Related Work 47
2.8 Conclusions 48

3 Base Station Placement for Wireless Sensor Networks 49

3.1 Introduction 49
3.2 Network Model and Problem Description 51
3.2.1 Network Model 51
3.2.2 Problem Description 52
3.3 An Approximation Algorithm 53

3.3.1 Our Approach . . . 53
3.3.2 Subarea Division and Fictitious Cost Points . . . 55
3.3.3 Summary of Algorithm and Example . . . 59
3.3.4 Proof and Complexity Analysis . . . 63
3.4 Numerical Results . . . 66
3.5 Related Work . . . 70
3.6 Conclusions . . . 71

4 Mobile Base Station for Wireless Sensor Networks 72

4.1 Introduction . . . 72
4.2 Network Model and Problem Formulation . . . 74
4.2.1 Network Model . . . 74
4.2.2 Problem Description . . . 75
4.3 From Time Domain to Space Domain . . . 77
4.4 Optimal Solution to the C-MB Problem . . . 83
4.5 An Approximation Algorithm to the U-MB Problem . . . 85
4.5.1 Subareas and Fictitious Cost Points . . . 85
4.5.2 Design of Approximation Algorithm . . . 86
4.5.3 Summary of Algorithm and Example . . . 91
4.5.4 Numerical Results . . . 94
4.6 Related Work . . . 101
4.7 Conclusions . . . 102

5 Optimal Spectrum Sharing for CR Networks **103**

5.1 Introduction . . . 103

5.2 CR Network Model and Problem Formulation . . . 105

5.2.1 Modeling of Multi-layer Characteristics . . . 106

5.2.2 Problem Formulation . . . 112

5.3 A Lower Bound for the Objective Function . . . 114

5.4 A Near-Optimal Algorithm Based on Sequential Fixing . . . 115

5.4.1 Basic Algorithm . . . 115

5.4.2 An Iteration-Speedup Technique . . . 117

5.5 Simulation Results . . . 117

5.6 Related Work . . . 122

5.7 Conclusions . . . 124

6 Optimal Power Control for CR Networks **125**

6.1 Introduction . . . 125

6.2 Related Work . . . 128

6.3 Power Control . . . 129

6.3.1 Transmission and Interference Ranges . . . 129

6.3.2 Necessary and Sufficient Condition for Successful Transmission . . . 130

6.3.3 Impact of Power Control . . . 132

6.4 Mathematical Modeling . . . 135

6.4.1 Scheduling and Power Control . . . 135

6.4.2 Flow Routing . . . 139
6.5 Problem Formulation . . . 140
6.6 A Solution Procedure . . . 142
6.6.1 Overview . . . 142
6.6.2 Linear Relaxation . . . 146
6.6.3 Local Search Algorithm . . . 148
6.6.4 Selection of Partition Variables . . . 148
6.7 Numerical Results . . . 150
6.7.1 Power Control and Level of Granularity on BFP . . . 153
6.7.2 Results on Scheduling Feasibility and Bandwidth Efficiency . . . 154
6.8 Conclusions . . . 156

7 Summary and Future Work **157**

7.1 Summary . . . 157
7.2 Future Research Direction . . . 160

Bibliography **162**

Vita **172**

List of Figures

2.1 Reference architecture for a two-tier wireless sensor network. 15

2.2 Node lifetime curves under the three approaches. 42

3.1 A schematic diagram showing that optimal base station location must be within SED. 54

3.2 A sequence of circles with increasing costs at centering node 4. 56

3.3 An example of subareas within disk $\mathcal{A}$ that are obtained by intersecting arcs from different circles. 58

3.4 The SED is divided into 16 subareas in the example sensor network. 62

3.5 A schematic showing the routing solution for the small 10-node network. 67

3.6 A schematic showing the routing solution for the 20-node network. 68

3.7 Network topology for 50-node and 100-node networks. 69

4.1 A simple example illustrating time-varying flow routing under φ in Lemma 5. . . . 79

4.2 A space-dependent flow routing under $\bar{\varphi}$ in Lemma 5. 80

4.3 Comparison of network lifetimes under different solutions used to construct $(1-\varepsilon)$-optimal solution. 86

4.4 The subareas for the example sensor network. 93

4.5 Results for the small 10-node network. 96

4.6 A 20-node network. 98

4.7 A 50-node network used in numerical investigation. 99

4.8 A 100-node network used in numerical investigation. 100

5.1 A schematic illustrating bands and sub-bands concepts in spectrum sharing. 107

5.2 An example illustrating interference among links. 110

5.3 Sequential fixing (SF) algorithm. 116

5.4 Normalized cost (with respect to lower bound) for 100 data sets of 20-node networks. 119

5.5 Normalized cost (with respect to lower bound). 121

6.1 A 3-link example network. 131

6.2 Illustration of branch-and-bound solution procedure. 143

6.3 A general framework of the branch-and-bound solution procedure. 144

6.4 Illustration of convex hull for a discrete term. 146

6.5 A local search algorithm. 149

6.6 A 20-node ad hoc network. 151

6.7 Total cost as a function of discretization levels of power control. 153

6.8 Routing topology for the 5 communication sessions in the 20-node network. 155

List of Tables

1.1 General notation. 9

2.1 Notation in Chapter 2. 21

2.2 Duality relationship between LMM node lifetime problem $\mathcal{P}_L$ and LMM rate allocation problem $\mathcal{P}_R$. 36

2.3 Locations (in meters) for each AFN in a 10-node network. 38

2.4 Locations (in meters) for each AFN in a 20-node network. 39

2.5 Node lifetime (in days) under the three approaches for the 10-node network. 39

2.6 Node lifetime (in days) under the three approaches for the 20-node network. 41

2.7 Numerical results verifying the duality relationship $T \cdot r_i = R \cdot t_i$ between the LMM rate allocation problem $\mathcal{P}_R$ and the LMM node lifetime problem $\mathcal{P}_L$ for the 10-node network. 45

2.8 Numerical results verifying the duality relationship $T \cdot r_i = R \cdot t_i$ between the LMM rate allocation problem $\mathcal{P}_R$ and the LMM node lifetime problem $\mathcal{P}_L$ for the 20-node network. 46

3.1 Notation in Chapter 3. 51

3.2 Each node's location, data rate, and initial energy of the example sensor network. . 61

3.3 Each node's location, data rate and initial energy for a small 10-node network. . . . 66

3.4 Each node's location, data rate and initial energy for a 20-node network. 68

4.1 Notation in Chapter 4. 75

4.2 Sensor locations, data rate, and initial energy of the example sensor network 92

4.3 Each node's location, data rate, and initial energy for a small 10-node network. . . 95

4.4 $(1-\varepsilon)$-optimal result for the small 10-node network with $\varepsilon = 0.05$. 97

4.5 Each node's location, data rate, and initial energy for a 20-node network. 97

4.6 $(1-\varepsilon)$-optimal result for the 20-node network with $\varepsilon = 0.05$. 98

4.7 $(1-\varepsilon)$-optimal result for the 50-node network with $\varepsilon = 0.05$. 98

4.8 $(1-\varepsilon)$-optimal result for the 100-node network with $\varepsilon = 0.05$. 100

5.1 Notation in Chapter 5. 106

5.2 Available bands among all nodes in the network in the simulation study. 118

5.3 Simulation results for the first 40 data sets of 20-node networks. 120

6.1 Notation in Chapter 6. 134

6.2 Each node's location and available frequency bands for the 20-node network. . . . 152

6.3 Source node, destination node, and rate requirement of the 5 sessions. 152

Chapter 1

Introduction

1.1 Motivation and Objective

In recent years, there has been significant advances in wireless networking. Such advances are fueled by the demand of new civil and military applications as well as advances in wireless communication technology at the physical layer. These new advances at upper and lower layers have brought many new research problems for networking researchers. Among these problems, a fundamental problem is to determine performance limits and to design a system to achieve these limits. Due to new requirements (performance metrics) by these new applications and unique characteristics (constraints) associated with these wireless networks, traditional analytical approaches are no longer adequate. In this dissertation, we will develop several new approaches to study the performance limits for wireless sensor networks and ad hoc networks.

Wireless ad hoc networks can be used to quickly build a network for peer-to-peer communication without first establishing a fixed infrastructure. Nodes in an ad hoc network are able to organize themselves into a multi-hop network for wireless communications. Ad hoc networks can be used where communication infrastructure is not available. On the other hand, wireless sensor networks are dedicated to in-situ, unattended, and real-time surveillance and monitoring appli-

cations. Such network consists of battery-powered nodes that are endowed with a multitude of sensing modalities [1]. Although there have been significant improvements in processor design and computing, advances in battery technology still lag behind, making energy resource the fundamental challenge in wireless sensor networking. As a consequence of the energy constraint, a new performance metric, namely, *network lifetime,* has become a vitally important performance metric for wireless sensor networks.

Recent advances in physical layer technologies, e.g., ultra wideband (UWB), multiple-input and multiple-output (MIMO), and cognitive radio (CR), have made profound impact on wireless networks. In this dissertation, we consider CR, which is a revolution in radio technology. CR is enabled by recent advances in RF design, signal processing, and communication software [83]. Fundamental characteristics of CR are that transmitted waveforms are defined by software and that received waveforms are demodulated by software. This is in contrast to traditional hardware based radios in which processing is done entirely in custom-made hardware circuitry. CR promises unprecedented flexibility in radio communications and is viewed as an enabling technology for dynamic spectrum access (DSA). For CR networks, we need to consider spectrum sharing among CR nodes. A new metric, called *bandwidth footprint product* (BFP) is proposed to measure CR nodes' resource usage in both spectrum and space.

To explore the performance limits of these new wireless networks, it is necessary to consider characteristics and constraints at multiple layers (i.e., power control at the physical layer, scheduling at the link layer, and routing at the network layer). Such problems are typically very complex, involving nonlinear, possibly non-convex relationship or constraints. As a result, developing theoretical results for these problems are very challenging and previous work are mostly heuristics without any performance guarantee. In this dissertation, we will develop several efficient algorithms to provide optimal or near-optimal solutions for these problems. These results are important not only for theoretical understanding, but also for providing guidelines to measure the performance of any proposed distributed algorithm and protocol.

1.2 Summary of Contributions

In this dissertation, I will present solutions to a number of challenging problems in wireless sensor networks and CR-based ad hoc networks.

1.2.1 Node Lifetime and Rate Allocation for Sensor Networks

We consider an overarching problem that encompasses both *network lifetime* (e.g., [17, 18]) and *network capacity* (e.g., [56]). There have been active research efforts recently at the networking layer on devising flow routing algorithms to maximize network lifetime [10, 11, 13, 17–19, 55, 115]. However, the network lifetime objective in most of these efforts has been centered around maximizing the time until the first node fails. For wireless sensor networks that are primarily designed for environmental monitoring or surveillance, the loss of a single node will only affect the coverage of one particular area and will not affect the monitoring or surveillance capabilities of the remaining nodes in the network. Consequently, it is important to investigate how to maximize the lifetime for, not only the first node, but also all the nodes in the network. For fairness, we maximize node lifetimes under the *lexicographic max-min* (LMM) criteria. Informally, the LMM node lifetime problem attempts to maximize the time until a set of nodes drain up their energy (which we call "drop point") while minimizing the number of nodes that drain up their energy at each drop point.

Recently, Brown et al. [15] studied the so-called "maximum node lifetime curve" problem, which is equivalent to the LMM node lifetime problem. A key step in their solution procedure is to use multiple independent linear programs (LPs) to determine the minimum set of nodes at each drop point. We call this approach "serial LP with slack variable analysis" (SLP-SV). Although this approach can solve the LMM node lifetime problem, its computational complexity is shown to be exponential.

Motivated by Brown et al.'s work, we develop a polynomial-time algorithm to derive the LMM-optimal node lifetime vector. We demonstrate that, for any given network configuration and initial

condition, our approach is always computationally more efficient than the slack variable (SV) based approach in [15]. The computational effectiveness of our approach accrues from two important techniques. First, we employ a link-based problem formulation, which significantly reduces the problem size in comparison with a flow-based formulation in [15]. Second, we exploit the so-called *parametric analysis* (PA) technique at each drop point to determine the minimum set of nodes that use up their energy. PA is extremely simple and has a linear time complexity per node in contrast to the SV-based approach in [15], which requires solving multiple additional LPs at each drop point.

In addition to providing an efficient polynomial-time algorithm for the LMM-optimal node lifetime vector computation, we also show how to obtain a corresponding flow routing solution among the remaining active nodes at each stage, such that the LMM-optimal node lifetime vector can indeed be achieved.

Another significant contribution in this work is that we extend our solution to the LMM node lifetime problem to study the LMM rate allocation problem. In particular, we study the network capacity (or rate allocation) problem under a given network lifetime requirement and the LMM criterion. We show how to apply the PA technique to solve the LMM rate allocation problem. Further, we show that there exists a simple and elegant *duality* relationship between the LMM node lifetime problem and the LMM rate allocation problem. As a result, it is sufficient to solve only one of these two problems. Important insights can be obtained by inferring duality results for the other problem.

1.2.2 Base Station Placement for Sensor Networks

Network lifetime is highly dependent upon the physical topology of the wireless sensor network. This is because energy expenditure at a node to transmit data to another node not only depends on the data bit rate, but also on the physical distance between these two nodes. Consequently, it is important to understand the impact of location related issues on network lifetime performance and

to optimize topology during network deployment stage.

We propose an approximation algorithm for base station placement problem to maximize the network lifetime. The main idea in our approximation algorithm is to discretize cost parameters (associated with energy consumption) with tight bounds under a given small error bound ε. As a result, we can divide the continuous search space into a finite number of subareas. By further exploiting the cost property of each subarea, we conceive a novel idea to represent each subarea with a so-called "fictitious cost point" (FCP), which is a cost vector with each component representing the upper bound of cost to a sensor node in the network. Based on these ideas, we have successfully reduced an infinite search space for base station location into a finite set of "points." For each point, we can apply an LP to find the corresponding achievable network lifetime and data routing solution. By comparing the maximum achievable network lifetime among all the FCPs, we choose the largest one and prove that locating the base station at any point in the subarea corresponding to this best FCP is $(1 - \varepsilon)$-optimal. We analyze the complexity of our approximation algorithm and show that it is significantly lower than a state-of-the-art algorithm proposed in [32].

1.2.3 Mobile Base Station for Sensor Networks

The benefits of using mobile base station to prolong sensor network lifetime have been well recognized [62, 110]. A mobile base station can alleviate the traffic aggregation burden from a fixed set of sensor nodes near the base station to the rest of sensor nodes in the network, and it is possible to extend the network lifetime significantly.

Although the potential benefit of using a mobile base station to prolong sensor network lifetime is significant, the theoretical difficulty of this problem is tremendous. There are two components that are tightly coupled in this problem. First, the location of the base station is now a function of time. That is, at different time instances, we have a different physical network topology, with the sink node being at different positions. Second, the traffic (or flow) routing behavior may change with both time as well as the location of the base station. As a result, to maximize the network

lifetime, we need to consider both base station movement and time-dependent flow routing.

We offer an in-depth study on network lifetime performance when a mobile base station is employed. As a first step, we show that as far as network lifetime objective is concerned, flow routing only needs to be dependent on the base station location, regardless of *when* the base station is present at this location. Further, the specific time instances for the base station to visit a location is not important, as long as the total sojourn time for the base station to be present at this location is fixed. With this finding, we make a novel transformation from a time-dependent problem formulation to location (space)-dependent problem formulation. As a second step, we show that when base station is only allowed to be present at a finite set of pre-determined points (called *constrained mobile base station* (C-MB) problem), we can find the optimal time duration for the base station to stay on each of these points (as well as the corresponding flow routing solution) via a single LP.

Building upon these results, we show that for the *un-constrained mobile base station* (U-MB) problem, i.e., the base station can be present at any point in the two dimensional plane, we can develop a polynomial-time approximation algorithm to provides a $(1-\varepsilon)$-optimal solution, where ε can be made arbitrarily small depending on required precision. The main idea in this approximation algorithm is again to divide the search space into subareas and to represent each subarea by an FCP. As a result, we can apply the LP approach developed for the C-MB problem on these FCPs and develop provably $(1-\varepsilon)$-optimal solution. This is the first theoretical result on mobile base station problem.

1.2.4 Optimal Spectrum Sharing for CR Networks

A cognitive radio (CR) is a frequency-agile data communication device with a rich control and monitoring (spectrum sensing) interface. It capitalizes advances in signal processing and radio technology, as well as recent advancements in spectrum policy [69, 83]. A frequency-agile radio module is capable of reconfiguring RF and switching to newly-selected frequency bands. Thus, a CR can be programmed to tune to and operate on specific frequency bands over a wide range of

spectrum [83].

We focus on the multi-hop networking problem for a CR-based wireless ad hoc network. In such a network, each node has a set of spectrum bands that it can use. Due to the unequal size of spectrum bands, it may be necessary to further divide each band into sub-bands for transmission and reception. Suppose there are a set of user sessions in the network that is characterized by a set of source-destination pairs each having certain rate requirement. We consider how to perform spectrum allocation, scheduling and interference avoidance, and multi-hop multi-path routing such that the required network-wide radio spectrum resource is minimized.

There are two reasons that we propose to investigate this problem. First, the cross-layer constraints that we will present is general and characterize common agreed behaviors of packet radio networks. Therefore, the solution procedure can be easily modified to address other performance objectives (e.g., rates or capacity). Second, it has been shown in [61] that so-called space bandwidth product is an extremely useful performance metric in the context of multi-hop CR networks. In the absence of power control, this metric degenerates into our performance objective.

To formulate the problem mathematically, we characterize the behavior and constraints from multiple layers for a general multi-hop CR network. Special attention is given to modeling of spectrum sharing and (unequal size) sub-band division, scheduling and interference modeling, and multi-path routing. We formulate an optimization problem with the objective of minimizing the required network-wide radio spectrum resource for a set of source-destination pair rate requirements. Since such a problem formulation is a mixed integer nonlinear program, which is NP-hard in general [40], we aim to develop a near-optimal solution. We first develop a lower bound by using a linear relaxation. This lower bound can be used as a measure for the quality of any solution. The proposed algorithm is based on a novel *sequential fixing* (SF) procedure where the determination of integer variables is performed iteratively through a sequence of LPs. Upon completing the fixing of the integer variables, the value of the other variables in the optimization problem can be obtained by solving an LP. Simulations show that the results obtained by the SF algorithm are very close to the lower bound, thus suggesting that the solutions obtained by the SF algorithm are even

closer to the optimum and thus near-optimal.

1.2.5 Optimal Power Control for CR Networks

We further consider how to support a set of user communication sessions (each with a rate requirement) by jointly optimizing power control, scheduling, and routing such that the required radio resource and interference area in the network are minimized. We follow the so-called "protocol model" [44] for interference modeling. Since power control directly affects the signal power at the receiving node and the interference power at other nodes, we find that it has profound impact on scheduling feasibility, bandwidth efficiency, and problem complexity.

We develop a formal mathematical model for scheduling when power control is employed. Along the line of protocol interference modeling, we develop a formal mathematical model for power control and scheduling. This model extends existing deterministic interference model and can be used for a range of problems where power control is part of the optimization space.

Based on this model, we formulate a joint power control, scheduling, and routing problem to support a set of user communication sessions with the objective of minimizing radio resource requirement and interference area. Subsequently, we develop an efficient solution procedure based on the branch-and-bound framework and convex hull relaxations to solve this cross-layer optimization problem. The solution obtained via our approach yields a $(1-\varepsilon)$-optimal solution, where ε is a small positive constant reflecting our desired accuracy in the final solution. By applying the solution procedure on sample random networks, we demonstrate quantitatively that power control has significant impact on scheduling feasibility, bandwidth efficiency, and BFP.

1.3 Dissertation Outline

The remainder of this dissertation is organized as follows. Chapter 2 presents efficient serial LP algorithm based on parametric analysis (SLP-PA) for the LMM node lifetime problem and the

Table 1.1: General notation.

Symbol	Definition	Symbol	Definition
$\mathcal{A}$	The search space for the base station	$\mathcal{N}$	The set of nodes in the network
$\mathcal{A}_m$	The m-th subarea in the search space	$O_{\mathcal{A}}$	The center of $\mathcal{A}$
B	The base station in a sensor network	r_i	Data rate at node i
d_{ij}	Distance from node i to node j	$r(l)$	Data rate of session l
$d(l)$	Destination node of session l	$R_{\mathcal{A}}$	The radius of $\mathcal{A}$
e_i	Initial energy at node i	$s(l)$	Source node of session l
f_{ij}	Data rate from on link $i \to j$	T	Network lifetime
$f_{ij}(l)$	Data rate that is attributed to session l	$\mathcal{T}_i^m$	The set of nodes that can use band m
	on link $i \to j$		to transmit to node i
g_{ij}	Propagation gain from node i to node j	$\mathcal{T}_i$	$= \bigcup_{m \in \mathcal{M}_i} \mathcal{T}_i^m$, the set of nodes that can
H_i	Number of circles at sensor node i		transmit to node i
	under a given ε	V_{ij}	The total bit volume from node i to j
$\mathcal{I}_j^m$	The set of nodes that can use band m	α	Path loss index
	and make non-neglectable interference	β_{rec}	Power consumption coefficient
	at node j		for receiving data
$\mathcal{L}$	The set of active sessions in the network	β_1, β_2	Constant terms in power consumption
M	The number of pre-determined locations		coefficient for transmitting data
	or the number of subareas	β_{ij}	Coefficient of power consumption
$\mathcal{M}_i$	The set of available bands at a cognitive		for transmitting data from node i to j
	radio (CR) node i	$\beta_{iB}^{\min}, \beta_{iB}^{\max}$	Lower and upper bounds of $\beta_{iB}(p)$
$\mathcal{M}$	$= \bigcup_{i \in \mathcal{N}} \mathcal{M}_i$, the set of available bands	$\beta[h]$	$= \beta_1(1+\varepsilon)^h$, the transmission cost
	in a CR network		for the h-th circle
$\mathcal{M}_{ij}$	$= \mathcal{M}_i \bigcap \mathcal{M}_j$, the set of available bands	ε	Desired small error bound, $\varepsilon > 0$
	for link $i \to j$	η	Ambient Gaussian noise density
N	The number of nodes in the network	ψ_ε	A $(1-\varepsilon)$-optimal solution

LMM rate allocation problem. We also show the duality relationship between these two problems. In Chapter 3, we present an approximation algorithm based on fictitious cost point (FCP) for the base station placement problem such that the network lifetime is at least $(1-\varepsilon)$ of the optimum. Chapter 4 presents an approximation algorithm for the mobile base station problem that provides a $(1-\varepsilon)$-optimal network lifetime. In Chapter 5, we present a sequential fixing (SF) algorithm for a multi-hop CR network to minimize the total required spectrum for a given set of communication sessions. In Chapter 6, we first develop a detailed mathematical model for power control, scheduling, and routing. Then, we develop a branch-and-bound solution procedure to minimize network-wide BFP for the given set of communication sessions. Table 1.1 lists the general notation used in this dissertation.

Chapter 2

Node Lifetime and Rate Allocation Problems for Wireless Sensor Networks

2.1 Introduction

Wireless sensor networks [1] consist of battery-powered nodes that are endowed with a multitude of sensing modalities including multimedia (e.g., video, audio) and scalar data (e.g., temperature, pressure, light, magnetometer, infrared). The demand for sensor networks is spurred by numerous applications that require in-situ, unattended, high-precision, and real-time observations over a vast area. Although there have been significant improvements in processor design and computing, advances in battery technology still lag behind, making energy resource the fundamental challenge in wireless sensor networking. Consequently, there have been active research efforts on performance limits of wireless sensor networks. These performance limits include, among others, *network lifetime* (e.g., [17,18]) and *network capacity* (e.g., [56]). Network lifetime refers to the maximum time limit that nodes in the network remain alive until one or more nodes drain up their energy, while network capacity typically refers to the maximum amount of bit volume that can be successfully delivered to the base station ("sink node") by all the nodes in the network.

In this chapter, we consider an overarching problem that encompasses both performance metrics. There have been active research efforts recently at the networking layer on devising flow routing algorithms to maximize network lifetime [10, 11, 13, 17–19, 55, 115]. However, the network lifetime objective in most of these efforts has been centered around maximizing the time until the first node fails. Although the time until the first node fails is an important measure from the complete network coverage point of view, this performance metric alone cannot measure the lifetime performance behavior for all nodes in the network. For wireless sensor networks that are primarily designed for environmental monitoring or surveillance, the loss of a single node will only affect the coverage of one particular area and will not affect the monitoring or surveillance capabilities of the remaining nodes in the network. This is because the remaining nodes in the network can adjust their transmission power (via power control) and reconfigure themselves into a new network routing (relay) topology so that information collected at the remaining nodes can still be delivered successfully to the base station. Consequently, it is important to investigate how to maximize the lifetime for, not only the first node, but also *all* the other nodes in the network. For fairness, we maximize node lifetimes under the lexicographic max-min (LMM) criteria, which will be formally defined in Section 2.2.3. We call this the *LMM node lifetime* problem.

Recently, Brown et al. [15] studied this problem under the so-called "maximum node lifetime curve" problem, which is equivalent to the LMM node lifetime problem. Informally, the maximum node life curve attempts to maximize the time until a set of nodes drain up their energy (which we call the *drop point*) while minimizing the number of nodes that drain up their energy at each drop point. The main contribution by Brown et al. [15] is the development of a procedure to solve the maximum node lifetime curve problem. A key step in their procedure is to use multiple independent linear programming (LP) calculations to determine the minimum set of nodes at each drop point, which we call "serial LP with slack variable analysis" (SLP-SV). Although this approach can solve the LMM node lifetime problem, its computational complexity is shown to be exponential, which could be a potential problem for large-scale networks.

Inspired by Brown et al.'s work on the LMM node lifetime problem, in this chapter, we develop a polynomial-time algorithm to derive the LMM-optimal node lifetime vector. In addition,

we demonstrate that, for any given network configuration and initial condition, our approach is always significantly computationally more efficient than the slack variable (SV) based approach in [15]. Consequently, this leads to an even stronger performance guarantee than the commonly used worst case complexity criteria. The computational effectiveness of our approach accrues from two important techniques. First, we employ a *link-based* problem formulation, which significantly reduces the problem size in comparison with a flow-based formulation used in [15]. Second, which is also the most significant contribution in this chapter, we exploit the so-called *parametric analysis* (PA) technique at each drop point to determine the minimum set of nodes that use up their energy. We show that this technique is a powerful tool in determining the minimum node set for each drop point. When the problem is non-degenerate, it is extremely simple and has a linear time complexity per node in contrast with the SV-based approach proposed in [15], which requires solving multiple additional LPs at each drop point. Even for the rare case, when the problem is degenerate, using the PA technique still is more efficient than the SV-based approach as it decreases the number of additional LPs that need to be solved at each drop point.

In addition to providing an efficient polynomial-time algorithm for the LMM-optimal node lifetime vector computation, we also develop a simple polynomial-time algorithm that provides a corresponding flow routing solution among the remaining alive nodes at each stage such that the LMM-optimal node lifetime vector can indeed be achieved. A nice property about this algorithm is that it can be executed in parallel (instead of in serial) for all the stages.

We also extend the PA technique for the LMM node lifetime problem to address the LMM rate allocation problem. In particular, we study the network capacity (or rate allocation) problem under a given network lifetime requirement. Again, for fairness, we use of the LMM criterion in rate allocation. In addition to solve the LMM rate allocation problem, we show that there exists a simple and elegant *duality* relationship between the LMM node lifetime problem and the LMM rate allocation problem. As a result, it is sufficient to solve only one of these two problems. Important insights can be obtained by inferring duality results for the other problem.

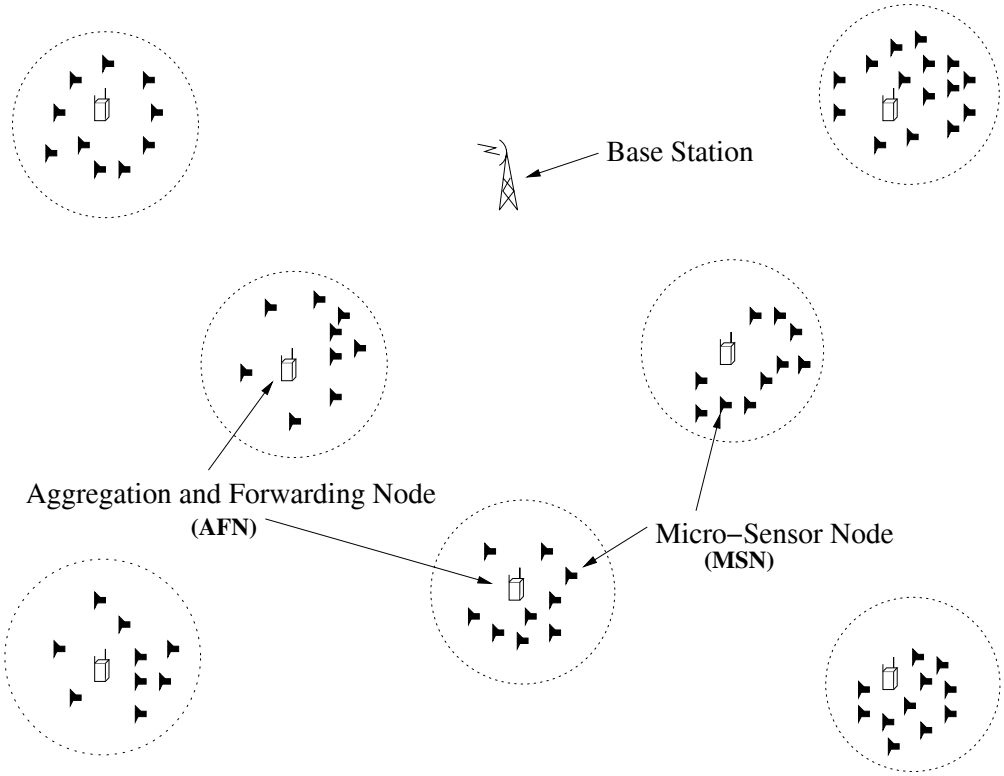
The remainder of this chapter is organized as follows. In Section 2.2, we describe the system

model and problem statement for this research, including the reference network architecture, nodal power dissipation behavior, and the LMM node lifetime problem description. We also describe a naive approach to address this problem and discuss why it usually gives an incorrect solution. Section 2.3 presents the link-based LMM problem formulation and our efficient serial LP algorithm based on parametric analysis, which we call SLP-PA. We also analyze the complexity of our algorithm and compares it with that in [15]. In Section 2.4, we introduce the LMM rate allocation problem and apply the SLP-PA algorithm to solve it. Section 2.5 shows an interesting duality relationship between the LMM node lifetime problem and the LMM rate allocation problem. Numerical results using the SLP-PA approach, the corresponding flow routing solution, and the duality relationship are given in Section 2.6. Section 2.7 reviews related work and Section 2.8 concludes this chapter.

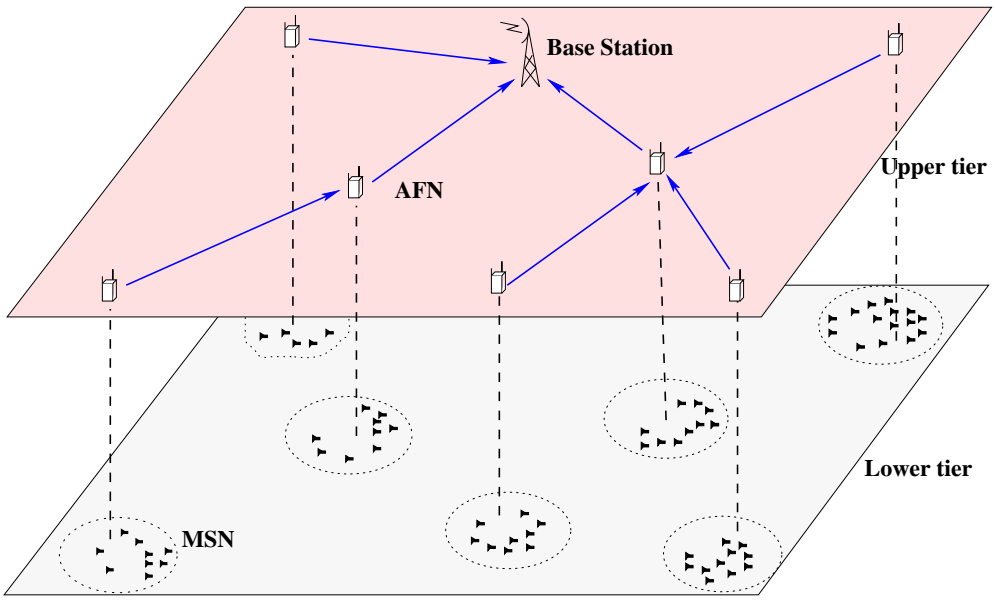
2.2 System Modeling and Problem Formulation

2.2.1 Reference Network Architecture

We focus on a two-tier architecture for wireless sensor networks. The two-tier network architecture is motivated by recent advances in distributed source coding [21, 75, 80]. Figures 2.1(a) and (b) show the physical and hierarchical network topology for such a network, respectively. Here, we have three types of nodes in the network: micro-sensor nodes (MSNs), aggregation and forwarding nodes (AFNs), and a base station. The MSNs can be application-specific sensor nodes (e.g., temperature sensor nodes, pressure sensor nodes, and video sensor nodes) and they constitute the lower tier of the network. They are deployed in groups (or clusters) at a strategic location for surveillance or monitoring applications. The MSNs are small and low-cost; they are densely deployed within a small geographical area. The objective of an MSN is very simple: Once triggered by an event, (e.g., the detection of motion or biological/chemical agents), it starts to capture live information (e.g., video), which it sends directly to the local AFN in one hop. It is worth pointing out that multi-hop routing among the MSNs may not be necessary due to the small distance between an



(a) Physical network topology.



(b) An instance of routing topology among AFNs.

Figure 2.1: Reference architecture for a two-tier wireless sensor network.

MSN and its AFN. By deploying these inexpensive MSNs in clusters, and within proximity of a strategic location, it is possible to obtain a comprehensive view of the area situation by exploring the correlation among the scenes collected at each MSN [21]. Furthermore, the reliability of area surveillance capability can also be improved through redundancy among the MSNs in the same cluster.

For each cluster of MSNs, there is one AFN, which is different from an MSN in terms of both its physical properties and functions. The primary functions of an AFN are: (1) data aggregation (or "fusion") for information flows coming from the local cluster of MSNs, and (2) forwarding (or relaying) the aggregated information to the next hop AFN toward the base station. For data fusion, an AFN analyzes the content of each data stream (e.g., video) it receives, from which it composes a complete scene by exploiting the correlation among each individual data stream from the MSNs. An AFN also serves as a relay node for other AFNs to carry traffic toward the base station. Although an AFN is expected to be provisioned with much more energy than an MSN, it also consumes energy at a substantially higher rate (due to wireless communication over large distances). Consequently, an AFN has limited lifetime. Upon the depletion of energy at an AFN, we expect that the coverage for the particular area under surveillance will be lost, despite the fact that some of the MSNs within the cluster may still have remaining energy.[1]

The third component in the two-tier architecture is the base station. The base station is, essentially, the sink node for all the AFNs in the network. A base station may be assumed to have a sufficient battery resource provision, or its battery may be re-provisioned during its course of operation. Therefore, its power dissipation is not a concern in our investigation.

In summary, the main functions of the lower tier MSNs are data acquisition and compression while the upper-tier AFNs are used for data fusion and relaying the information to the base station. The routing topology can be controlled by the power level of a node's transmitter [43, 78, 86, 105], which in turn controls the distance coverage of an AFN. Consequently, by adjusting the power level of an AFN's transmitter, we can form different network routing topologies.

[1] We assume that each MSN can only forward information to its local AFN for processing (e.g., video fusion).

2.2.2 Energy Model

For an AFN, the power consumption by data communication (i.e., receiving and transmitting) is the dominant factor [1]. The power dissipation at the transmitter can be modeled as:

$$p_{ik}^{\mathrm{t}} = \beta_{ik} \cdot f_{ik} \ , \tag{2.1}$$

where p_{ik}^{t} is the power dissipated at node i when it is transmitting to node k, f_{ik} is the bit rate transmitted from node i to node k, and β_{ik} is the power consumption cost of radio link $i \rightarrow k$ and is given by

$$\beta_{ik} = \beta_1 + \beta_2 \cdot d_{ik}^{\alpha} \ , \tag{2.2}$$

where β_1 and β_2 are two constant terms, d_{ik} is the distance between these two nodes, and α is the path loss index, with $2 \leq \alpha \leq 4$ [82]. The power dissipation at a receiver can be modeled as [82]:

$$p_i^{\mathrm{r}} = \beta_{\mathrm{rec}} \cdot \sum_{m \in \mathcal{N}}^{m \neq i} f_{mi} \ , \tag{2.3}$$

where β_{rec} is a constant.

The above transmission and reception energy model assumes a contention-free MAC protocol, where interference from simultaneous transmissions can be effectively avoided. For deterministic rate traffic pattern model in this chapter, a contention-free MAC protocol is fairly easy to design (e.g., [93]) and its discussion is beyond the scope of this chapter.

2.2.3 The Lexicographic Max-Min Node Lifetime Problem

For a network having a set $\mathcal{N}$ of AFNs, suppose that AFN $i \in \mathcal{N}$ generates data stream at a rate r_i, and that the initial energy at this node is given by e_i. For an AFN, the power consumption by data communication (i.e., receiving and transmitting) is the dominant factor [1]. Then, it is straightforward to use a linear programming (LP) approach to find an optimal routing solution

such that the time until any AFN runs out of energy is maximized [17, 18]. That is, we can solve

$$\begin{aligned}
\text{Max} \quad & T \\
\text{s.t.} \quad & f_{iB} + \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} f_{ik} - \sum_{m\in\mathcal{N}}^{m\neq i} f_{mi} = r_i && (i \in \mathcal{N}) && (2.4)\\
& \textstyle\sum_{m\in\mathcal{N}}^{m\neq i} \beta_{\text{rec}} f_{mi} T + \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{ik} f_{ik} T + \beta_{iB} f_{iB} T \leq e_i && (i \in \mathcal{N}) && (2.5)\\
& T, f_{ik}, f_{iB} \geq 0 && (i, k \in \mathcal{N}, i \neq k)\ .
\end{aligned}$$

where T represents the maximum time until the first node fails, f_{ik} and f_{iB} are data rates transmitted from $\beta_{\text{rec}}, \beta_{ik}$, and β_{iB} are constants. Eq. (2.4) states that, the total data transmitted from AFN i is equal to the total data received from other AFNs, plus the data generated locally by AFN i. Eq. (2.5) states that the energy required to receive and transmit all these data cannot exceed its initial energy.

Note that since f_{mi}, f_{ik}, f_{iB}, and T are all variables, the above optimization problem is not linear (due to the product terms $f_{mi}T$, etc). To transform it into an LP, we denote $V_{ik} = f_{ik}T$ and $V_{iB} = f_{iB}T$. Then, we may equivalently solve:

$$\begin{aligned}
\text{Max} \quad & T \\
\text{s.t.} \quad & V_{iB} + \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} V_{ik} - \sum_{m\in\mathcal{N}}^{m\neq i} V_{mi} - r_i T = 0 && (i \in \mathcal{N}) && (2.6)\\
& \textstyle\sum_{m\in\mathcal{N}}^{m\neq i} \beta_{\text{rec}} V_{mi} + \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{ik} V_{ik} + \beta_{iB} V_{iB} \leq e_i && (i \in \mathcal{N})\\
& T, V_{ik}, V_{iB} \geq 0 && (i, k \in \mathcal{N}, i \neq k)\ , && (2.7)
\end{aligned}$$

where (2.6) follows by multiplying (2.4) by T. We now have an LP problem and the optimal solution for T represents the maximum time until the first node fails.

Although it is important to maximize the time until first AFN runs out of energy (also know as the network lifetime in [10, 17]), it is even more important to concurrently maximize also the time that the second, third, and all subsequent AFNs run of energy. That is, it is important to find a routing solution among the AFNs such that the lifetimes of all AFNs in the network can achieve the optimal *lexicographic max-min* (LMM) vector. A formal definition for the LMM-optimal node lifetime vector is hereby given as follows.

Definition 1. (LMM-optimal Node Lifetime) *A sorted node lifetime vector* $[\tau_1, \tau_2, \cdots, \tau_N]$ *with* $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_N$ *is LMM-optimal if and only if for any other sorted node lifetime vector* $[\hat{\tau}_1, \hat{\tau}_2, \cdots, \hat{\tau}_N]$ *with* $\hat{\tau}_1 \leq \hat{\tau}_2 \leq \cdots \leq \hat{\tau}_N$*, there exists a* $k \leq N$*, such that* $\tau_i = \hat{\tau}_i$ *for* $i = 1, 2, \cdots, k-1$ *but* $\tau_k > \hat{\tau}_k$.

A naive approach to the LMM node lifetime problem would be to apply a max-min like iterative procedure to find the sequence of node lifetime for all AFNs in the network, by considering the energy of each AFN as the bottleneck resource. Under this approach, an iterative LP for alive nodes in the form of (2.6) to (2.7) could be employed to find the maximum time until the next node fails. By calculating the remaining energy at each node at the end of the iteration, we can move on to the next iteration, until all the nodes drain their energy. Although this approach seems appealing and intuitive, we now show that it usually gives an incorrect solution.

We first must realize that there is a fundamental difference between the LMM node lifetime problem here and the classical max-min rate allocation problem described in [9, 47]. That is, the LMM node lifetime problem implicitly embeds (or couples) a routing problem within the LMM node lifetime problem, while under the classical max-min rate allocation, there is no routing problem involved since the routes for all flows are fixed.

Due to this coupling of routing and LMM node lifetime optimization, we find that *any iterative LMM node lifetime algorithm requiring energy reservation among the nodes during each iteration is incorrect.* This is because, unlike max-min (which addresses only the rate allocation problem under fixed routes), starting from the first iteration, there usually exist non-unique routing solutions corresponding to the same next lifetime level. Consequently, each of these routing solutions, once chosen, will yield different remaining energy at the AFNs for future iterations and so forth, leading to a different node lifetime vector, which may not be the same as the LMM-optimal node lifetime vector. Numerical results demonstrating the incorrectness of this naive approach (which we call *serial LP with energy reservation* (SLP-ER)) will be given in Section 2.6.

Recently, Brown et al. [15] studied the LMM node lifetime problem under the notion of "node lifetime curve". In their approach, they first identified the uniqueness of the LMM-optimal node

lifetime vector. Based on this property, they developed an iterative procedure to solve the LMM node lifetime problem. In particular, they developed a revised simplex method to calculate the maximum node lifetime curve (equivalent to the LMM-optimal node lifetime vector). A key step in their procedure is to use multiple independent LPs to maximize the sum of slack variables in order to determine the minimum node set at each lifetime level. During each iteration, only the lifetime level and the corresponding minimum set of nodes are determined, and there is no resource reservation process among the nodes. Although their proposed approach solves the LMM node lifetime problem, there still remain significant issues to be addressed. Among others, the problem formulation proposed in [15] is shown to be of exponential computational complexity, which could become problematic when the scale of the network becomes large.

2.3 An Efficient Algorithm Based on Parametric Analysis

In this section, we present an efficient algorithm for the LMM node lifetime problem. Unlike the serial LP with slack variable analysis (SLP-SV) approach in [15], our approach results in a polynomial running time. Moreover, for any given network configuration and initial condition, our approach is much simpler than the slack variable (SV) based approach in [15]. The computational effectiveness of our approach hinges upon two important techniques. First, we employ a link-based problem formulation that significantly reduces the problem size in comparison with a flow-based formulation adopted in [15]. Second, we invoke a *parametric analysis* (PA) procedure at each stage to determine the minimum node set at each lifetime level. For non-degenerate problems, this parametric analysis results in only a linear time computational complexity per node, while the SV-based approach in [15] requires the solution of multiple independent LPs to determine the minimum set of nodes at each lifetime level. Even for the rare case when the problem is degenerate, using our parametric analysis technique is still more efficient than the SV-based approach because it decreases the number of additional LPs that needs to be solved at each lifetime level. In the remainder of this section, we elaborate on the details of our serial LP algorithm based on parametric analysis (SLP-PA). Table 2.1 shows the notation that are used only in this chapter.

Table 2.1: Notation in Chapter 2.

Symbol	Definition
$\mathbf{b}$	Right-hand-side (RHS) of LMM-Lifetime or LMM-Rate
$\mathcal{B}$ (or $\mathcal{Z}$)	The columns corresponding to the basic (or non-basic) variables of LMM-Lifetime or LMM-Rate
$\mathbf{c}_{\mathcal{B}}$	The parameters in the objective function corresponding to the basic (or non-basic) variables
(or $\mathbf{c}_{\mathcal{Z}}$)	of LMM-Lifetime or LMM-Rate
g_i	The i-th rate in the sorted LMM-optimal rate vector, i.e., $g_1 \leq g_2 \leq \cdots \leq g_N$
$\mathbf{I}_i$	A vector having a single 1 corresponding to the index i of (2.8) or (2.20) and 0 elsewhere
L	The number of distinct levels in the sorted LMM-optimal lifetime/rate vector
$\mathcal{N}_l$	The minimum set of nodes that uses up energy at l-th level
$\hat{\mathcal{N}}_l$	The set of all possible AFNs which may use up energy at l-th level, $\mathcal{N}_l \subseteq \hat{\mathcal{N}}_l$
t_i	AFN i's lifetime under the LMM-optimal node lifetime vector
$\mathbf{v}$	The optimal solution for the dual problem of LMM-Lifetime or LMM-Rate
$\mathbf{x}$	The optimal solution for LMM-Lifetime or LMM-Rate
$\mathbf{x}_{\mathcal{B}}$	Part of optimal solution corresponding to the basic (or non-basic) variables of LMM-Lifetime
(or $\mathbf{x}_{\mathcal{Z}}$)	or LMM-Rate
λ_l	The l-th distinct rate in the sorted LMM-optimal rate vector, i.e., $\lambda_1(=g_1) < \lambda_2 < \cdots < \lambda_L(=g_N)$
δ_l	$= \lambda_l - \lambda_{l-1}$
τ_i	The i-th node lifetime in the sorted LMM-optimal node lifetime vector, i.e., $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_N$
μ_l	The l-th distinct lifetime in the sorted LMM-optimal node lifetime vector,
	i.e., $\mu_1(=\tau_1) < \mu_2 < \cdots < \mu_L(=\tau_N)$
ζ_l	$= \mu_l - \mu_{l-1}$

2.3.1 Link-based Formulation

Suppose that $[\tau_1, \tau_2, \cdots, \tau_N]$ with $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_N$ is LMM-optimal. If $\tau_k = \tau_{k+1}$, then the respective nodes corresponding to τ_k and τ_{k+1} have the same node lifetime, i.e., the corresponding nodes for τ_k and τ_{k+1} use up their energy at the same time. To keep a track of distinct node lifetimes, we remove all repetitive elements in the vector and rewrite it as $[\mu_1, \mu_2, \cdots, \mu_L]$ such that $\mu_1 < \mu_2 < \cdots < \mu_L$, where $\mu_1 = \tau_1$, $\mu_L = \tau_N$, and $L \leq N$. Corresponding to these lifetime levels, denote $\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_L$ as the sets of nodes that drain their energy at the lifetime levels $\mu_1, \mu_2, \cdots, \mu_L$, respectively. Clearly, $\bigcup_{l=1}^{L} \mathcal{N}_l = \mathcal{N}$. The problem is to find the LMM-optimal values of $\mu_1, \mu_2, \cdots, \mu_L$ and the corresponding sets $\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_L$.

To formulate this problem into an iterative form, we define $\mu_0 = 0$ and $\mathcal{N}_0 = \emptyset$. Furthermore, denote $\zeta_l = \mu_l - \mu_{l-1}$. Then, the iterative optimization problem (starting with $l = 1$) for the LMM node lifetime problem becomes,

LMM-Lifetime: Max ζ_l

$$\text{s.t.} \quad V_{iB} + \sum_{k \in \mathcal{N}}^{k \neq i} V_{ik} - \sum_{m \in \mathcal{N}}^{m \neq i} V_{mi} - \zeta_l r_i = \mu_{l-1} r_i \quad (i \notin \bigcup_{j=0}^{l-1} \mathcal{N}_j) \tag{2.8}$$

$$V_{iB} + \sum_{k \in \mathcal{N}}^{k \neq i} V_{ik} - \sum_{m \in \mathcal{N}}^{m \neq i} V_{mi} = \mu_h r_i \quad (i \in \mathcal{N}_h, 1 \leq h < l) \tag{2.9}$$

$$\sum_{m \in \mathcal{N}}^{m \neq i} \beta_{\text{rec}} V_{mi} + \sum_{k \in \mathcal{N}}^{k \neq i} \beta_{ik} V_{ik} + \beta_{iB} V_{iB} \leq e_i \quad (i \notin \bigcup_{j=0}^{l-1} \mathcal{N}_j) \tag{2.10}$$

$$\sum_{m \in \mathcal{N}}^{m \neq i} \beta_{\text{rec}} V_{mi} + \sum_{k \in \mathcal{N}}^{k \neq i} \beta_{ik} V_{ik} + \beta_{iB} V_{iB} = e_i \quad (i \in \mathcal{N}_h, 1 \leq h < l) \tag{2.11}$$

$$V_{ik}, V_{iB}, \zeta_l \geq 0 \qquad (i, k \in \mathcal{N}, i \neq k).$$

For each node that still has remaining energy at time μ_{l-1}, (2.8) states that the total in-coming and local data bit volumes are equal to the total out-going data bit volumes, while (2.10) states that the total energy consumed for receiving and transmitting data bit volumes is no more than the initial energy. For each node that no longer has any remaining energy at time μ_h, $1 \leq h < l$, (2.9) states that the total in-coming and local data volumes are equal to the out-going data bit volume at time μ_h, while (2.11) states that the total energy consumed for receiving and transmitting data is equal to the initial energy.

The above LP formulation can be rewritten in the form: Max $\mathbf{cx}$, s.t. $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq 0$, the dual problem for which is given by: Min $\mathbf{vb}$, s.t. $\mathbf{vA} \geq \mathbf{c}$ and $\mathbf{v}$ unrestricted [6]. Both can be solved simultaneously by standard LP techniques in polynomial-time (e.g., [6]). Although solving LMM-Lifetime gives the optimal value for ζ_l, we need yet to determine the minimum set of nodes corresponding to this ζ_l, which is the main task in this investigation. We will exploit PA technique [6] to determine the minimum node set for each lifetime level.

2.3.2 Minimum Node Set Determination with Parametric Analysis

Denote $\hat{\mathcal{N}}_l \neq \emptyset$ to be the set of nodes that achieve equality in (2.10). Although at least one of the nodes in $\hat{\mathcal{N}}_l$ must belong to $\mathcal{N}_l$ (the minimum node set at μ_l), some of the nodes in $\hat{\mathcal{N}}_l$ may still be further "stretched" to live longer under alternative routing solutions. In the special case, if $|\hat{\mathcal{N}}_l| = 1$, then $\mathcal{N}_l = \hat{\mathcal{N}}_l$; otherwise, we need to determine the minimum set of $\mathcal{N}_l$ ($\subseteq \hat{\mathcal{N}}_l$) that achieves the LMM-optimal solution.

We find the so-called *parametric analysis* technique [6] is most effective in addressing this type of problems. The main idea of parameter analysis is to find how a small perturbation of some component in the LMM-Lifetime will affect the solution. In particular, consider a small increase in the right-hand-side (RHS) of (2.8), i.e., changing b_i to $b_i + \varepsilon_i$, where $\varepsilon_i > 0$. Then this node i belongs to $\mathcal{N}_l$ *if and only if* $\frac{\partial^+ \zeta_l}{\partial \varepsilon_i}(0) < 0$, i.e., a small increase in node i's lifetime (in terms of total bit volume generated at node i) leads to a decrease in the next lifetime level.

To compare $\frac{\partial^+ \zeta_l}{\partial \varepsilon_i}(0)$ with 0, we resort to an important duality relationship in LP theory. If $\mathbf{x}$ and $\mathbf{v}$ are the respective optimal solutions to the primal and dual problems, then based on the parametric duality property [6], we have

$$\frac{\partial^+ \zeta_l}{\partial \varepsilon_i}(0) = \frac{\partial^+ (cx)}{\partial b_i}(b_i) \leq v_i \ .$$

Note that by the nature of the problem, we have $v_i \leq 0$ for an optimal dual solution. Recall that these v_i can be easily obtained at the same time when we solve the primal LP problem. Therefore, if $v_i < 0$, then we can determine immediately that $i \in \mathcal{N}_l$. On the other hand, if we find that $v_i = 0$,

it is not clear whether $\frac{\partial^+ \zeta_l}{\partial \varepsilon_i}(0)$ is strictly negative or 0 and further analysis is thus needed.

For each node i with $v_i = 0$, we must perform a complete PA to see whether this RHS can be further increased without changing the objective value of LMM-Lifetime. If there is no change, then we can determine that node $i \notin \mathcal{N}_l$; otherwise, $i \in \mathcal{N}_l$.

Assume that the optimal solution is $(\mathbf{x}_\mathcal{B}, \mathbf{x}_\mathcal{Z})$, where $\mathbf{x}_\mathcal{B}$ and $\mathbf{x}_\mathcal{Z}$ denote the set of basic and non-basic variables; $\mathcal{B}$ and $\mathcal{Z}$ denote the columns corresponding to the basic and non-basic variables; $\mathbf{c}_\mathcal{B}$ and $\mathbf{c}_\mathcal{Z}$ denote the objective function coefficient vectors for the basic and non-basic variables; and q denotes the objective value. Then the corresponding canonical equations yield [6]

$$\begin{aligned} q + (\mathbf{c}_\mathcal{B}^t \mathcal{B}^{-1} \mathcal{Z} - \mathbf{c}_\mathcal{Z}^t)\mathbf{x}_\mathcal{Z} &= \mathbf{c}_\mathcal{B}^t \mathcal{B}^{-1} \mathbf{b} \ , \\ \mathbf{x}_\mathcal{B} + \mathcal{B}^{-1} \mathcal{Z} \mathbf{x}_\mathcal{Z} &= \mathcal{B}^{-1} \mathbf{b} \ . \end{aligned}$$

If $\mathbf{b}$ is replaced by $\mathbf{b} + \varepsilon_i \mathbf{I}_i$, where the column vector $\mathbf{I}_i$ has a single 1 corresponding to node i in the set of constraints (2.8) and has 0 elements otherwise, then the only change in the constraints due to this perturbation is that $\mathcal{B}^{-1}\mathbf{b}$ will be replaced by $\mathcal{B}^{-1}(\mathbf{b} + \varepsilon_i \mathbf{I}_i)$. Consequently, the objective value for the current basis becomes $\mathbf{c}_\mathcal{B}^t \mathcal{B}^{-1}(\mathbf{b} + \varepsilon_i \mathbf{I}_i)$. Furthermore, as long as $\mathcal{B}^{-1}(\mathbf{b} + \varepsilon_i \mathbf{I}_i)$ is nonnegative, the current basis remains optimal. Denote $\bar{\mathbf{b}} = \mathcal{B}^{-1}\mathbf{b}$ and $\mathcal{B}_i^{-1} = \mathcal{B}^{-1}\mathbf{I}_i$ and let $\hat{\varepsilon}_i$ be an upper bound for ε_i such that the current basis remains optimal, we have

$$\hat{\varepsilon}_i = \min_j \{ \frac{\bar{b}_j}{-\mathcal{B}_{ij}^{-1}} : \mathcal{B}_{ij}^{-1} < 0 \} \ . \tag{2.12}$$

If $\hat{\varepsilon}_i > 0$, the optimal objective value varies according to $\mathbf{c}_\mathcal{B}^t \mathcal{B}^{-1}(\mathbf{b} + \varepsilon_i \mathbf{I}_i)$ for $0 < \varepsilon_i \leq \hat{\varepsilon}_i$. Since $\mathbf{v} = \mathbf{c}_\mathcal{B}^t \mathcal{B}^{-1}$ and $v_i = 0$, we have $\mathbf{c}_\mathcal{B}^t \mathcal{B}^{-1} \mathbf{I}_i = v_i = 0$. Thus, the objective value will not change for $\varepsilon_i \in (0, \hat{\varepsilon}_i]$, and consequently, the lifetime for node i can be "stretched" to last longer beyond current lifetime level μ_l. That is, node i does not belong to the minimum node set $\mathcal{N}_l$.

For most practical problems, this directly yields whether $i \in \mathcal{N}_l$ or $i \notin \mathcal{N}_l$ for all $i \in \hat{\mathcal{N}}_l$. But in the rare event where $\hat{\varepsilon}_i = 0$, the problem is degenerate. To develop a polynomial-time algorithm, denote $\mathcal{V}_l$ as the set of all nodes with $v_i < 0$ and $\mathcal{U}_l$ the set of all nodes with $v_i = 0$ and $\hat{\varepsilon}_i = 0$.

Then we solve the following LP to maximize the slack variables for nodes in $\mathcal{U}_l$.

$$
\begin{aligned}
\textbf{MSV: Max} \quad & \textstyle\sum_{i\in\mathcal{U}_l} \varepsilon_i \\
\text{s.t.} \quad & V_{iB} + \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} V_{ik} - \sum_{m\in\mathcal{N}}^{m\neq i} V_{mi} - \varepsilon_i r_i = \mu_l r_i, \quad (i \in \mathcal{U}_l) \\
& V_{iB} + \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} V_{ik} - \sum_{m\in\mathcal{N}}^{m\neq i} V_{mi} = \mu_h r_i \ , \quad (i \in \mathcal{N}_h, 1 \le h < l) \\
& V_{iB} + \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} V_{ik} - \sum_{m\in\mathcal{N}}^{m\neq i} V_{mi} = \mu_l r_i, \quad (i \notin \mathcal{U}_l \bigcup_{h=1}^{l-1} \mathcal{N}_h) \\
& \textstyle\sum_{m\in\mathcal{N}}^{m\neq i} \beta_{\text{rec}} V_{mi} + \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{ik} V_{ik} + \beta_{iB} V_{iB} = e_i \ , \quad (i \in \mathcal{U}_l \bigcup \mathcal{V}_l \bigcup_{h=1}^{l-1} \mathcal{N}_h) \\
& \textstyle\sum_{m\in\mathcal{N}}^{m\neq i} \beta_{\text{rec}} V_{mi} + \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{ik} V_{ik} + \beta_{iB} V_{iB} \le e_i \ , \quad (i \notin \mathcal{U}_l \bigcup \mathcal{V}_l \bigcup_{h=1}^{l-1} \mathcal{N}_h) \\
& V_{ik}, V_{iB}, \varepsilon_i \ge 0 \ , \quad (i,k \in \mathcal{N}, i \neq k).
\end{aligned}
$$

If the optimal objective value is 0, then no node in $\mathcal{U}_l$ can have a positive ε_i, i.e., these nodes should all belong to $\mathcal{N}_l$. That is, these nodes should all belong to $\mathcal{N}_l$ and we have $\mathcal{N}_l = \mathcal{V}_l + \mathcal{U}_l$. On the other hand, if the optimal objective value is positive, then some nodes in $\mathcal{U}_l$ must have positive ε_i, i.e., these nodes should not belong to $\mathcal{N}_l$. Consequently, we remove these nodes from $\mathcal{U}_l$ and if $\mathcal{U}_l \neq \emptyset$, we solve another MSV. This procedure will terminate when the optimal objective value is 0 or $\mathcal{U}_l = \emptyset$.

In a nutshell, the complete PA procedure for the determination of whether or not a node $i \in \hat{\mathcal{N}}_l$ belongs to the minimum node set $\mathcal{N}_l$ can be summarized as follows.

Algorithm 1. (Minimum Node Set Determination with PA)

1. *Initialize $\mathcal{V}_l = \emptyset$ and $\mathcal{U}_l = \emptyset$.*
2. *For each node $i \in \hat{\mathcal{N}}_l$,*
   (a) *if $v_i < 0$, add node i to $\mathcal{V}_l$;*
   (b) *otherwise, using $\mathcal{B}^{-1}$ (which is readily available after solving an LMM-Lifetime), compute $\bar{\mathbf{b}} = \mathcal{B}^{-1}\mathbf{b}$, $\mathcal{B}_i^{-1} = \mathcal{B}^{-1}\mathbf{I}_i$, and $\hat{\varepsilon}_i$ according to (2.12). If $\hat{\varepsilon}_i = 0$, add i to $\mathcal{U}_l$.*

3. *If $\mathcal{U}_l = \emptyset$, let $\hat{\mathcal{N}}_l = \mathcal{V}_l$ and stop, else build and solve an MSV.*

4. *If the optimal objective value is 0, let $\hat{\mathcal{N}}_l = \mathcal{V}_l + \mathcal{U}_l$ and stop. Otherwise, remove all nodes with $\varepsilon_i > 0$ from $\mathcal{U}_l$ and go to Step 3.*

The following lemma establishes an important property for the minimum node set obtained at each lifetime level.

Lemma 1. *Under the LMM-optimal solution, the set of nodes in the minimum node set at each lifetime level is unique.*

Proof. By the definition of LMM-optimal node lifetime vector (Definition 1), the optimal node lifetimes (τ_l values) are unique and the corresponding number of nodes in the minimum node set ($|\hat{\mathcal{N}}_l|$ values) are also unique. To show that the group of physical nodes in each $\hat{\mathcal{N}}_l$ is also unique, we employ the parametric simplex approach to determine the minimum node set as follows.

In essence, the parametric simplex approach solely relies on PA technique without resorting the SV approach even when the problem is degenerate. That is, when the problem is degenerate, i.e., for some node $i \in \hat{\mathcal{N}}_l$, we have $v_i = 0$ and $\hat{\varepsilon}_i = 0$, then the basis can change while the optimal objective value remains unchanged. We can analyze v_i and ε_i under the new basis to determine whether or not node i belongs to the minimum node set $\hat{\mathcal{N}}_l$. If we still have $v_i = 0$ and $\hat{\varepsilon}_i = 0$, the basis can change again with the same optimal objective value. To prevent cycling back to a previous basis, we can use a de-cycling rule [6]. Thus, this procedure is guaranteed to terminate within a finite number of steps and we can determine whether or not node i indeed belongs to the minimum node set $\hat{\mathcal{N}}_l$.

Note that in the above parametric simplex approach, the set of physical nodes corresponding to $\hat{\mathcal{N}}_l$ is uniquely determined since the analysis is conducted independently for each node. Therefore, upon the completion of all stages, the group of physical nodes in each minimum node set is unique.

□

2.3.3 LMM-Optimal Routing Solution

Non-uniqueness of Routing Solution. The solution to the LMM node lifetime problem would not be complete without a corresponding routing solution. The first question to ask is whether such a routing solution is unique. We show that, although the set of physical nodes corresponding to the minimum node set at each stage is unique (Lemma 1), the routing solution is non-unique. This is because upon the completion of the last LMM-Lifetime in SLP-PA, there usually exist non-unique bit volume solutions (V_{ik} and V_{iB} values along each radio link), all of which can achieve the same unique objective values (ζ_i values). As the bit volumes (V_{ik} and V_{iB} values) along each radio link are non-unique, the corresponding routing solution is, as a result, also non-unique. This observation is formally stated in the following lemma.

Lemma 2. *The routing solution corresponding to the LMM-optimal node lifetime vector can be non-unique.*

Incidentally, this result corrects an error in [15] (Lemma 3.2), which incorrectly states that such a routing solution is unique.

An Optimal Routing Solution. Given that the optimal routing solution is non-unique, there are potentially many possible routing solutions that achieve the LMM-optimal node lifetime vector. In this section, we present a simple approach that provides an LMM-optimal routing solution.

The main task in this algorithm is to define flows from the bit volumes (V_{ik} and V_{iB} values), which are obtained upon the completion of the last LMM-Lifetime in our SLP-PA approach. Note that the bit volumes obtained here represent the total amount of bit volume being transported between the nodes during $[0,\ \mu_L]$, where $\mu_L = \tau_N$ is the time that the last set of nodes drain their energy. The main result here is that if we let the total amount of out-going flow at a node be distributed *proportionally to the bit volumes* on each out-going link for all the remaining alive nodes at each stage, then we can achieve the drop points $\mu_1,\ \mu_2,\ \cdots, \mu_L$ as well as the corresponding minimum node sets $\mathcal{N}_1,\ \mathcal{N}_2,\ \cdots, \mathcal{N}_L$. The algorithm is formally described as follows.

Algorithm 2. (An Optimal Flow Routing Solution) *Upon the completion of the SLP-PA algorithm for the LMM node lifetime vector, we have the drop points (in strictly increasing order) $\mu_1, \mu_2, \cdots, \mu_L$, the corresponding minimum physical node sets $\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_L$, and the total amount of bit volume on each radio link (i.e., V_{ik} and V_{iB}). The following algorithm gives an LMM-optimal flow routing solution for the corresponding time interval $(\mu_{l-1}, \mu_l]$, where $\mu_0 = 0$ and $l = 1,\ 2, \cdots,\ L$.*

1. *Denote $\mathcal{U}_l = \mathcal{N} - \bigcup_{j=0}^{l-1} \mathcal{N}_j$, with $\mathcal{N}_0 = \emptyset$. Initialize all flows to zero, i.e., $f_{ik}^{(l)} = 0$, $f_{iB}^{(l)} = 0$ for $i, k \in \mathcal{N}, i \neq k$.*

2. *If $\mathcal{U}_l = \emptyset$, then stop, else choose a node i from $\mathcal{U}_l$ such that*[2]

   - *either node i does not receive data from any other node, or*
   - *all nodes from which node i receives data are not in $\mathcal{U}_l$.*

3. *The flow routing at node i during $(\mu_{l-1}, \mu_l]$ is then defined as*

$$f_{ik}^{(l)} = \frac{V_{ik}}{V_{iB} + \sum_{k \in \mathcal{U}_l}^{k \neq i} V_{ik}} \left(\sum_{m \in \mathcal{U}_l}^{m \neq i} f_{mi}^{(l)} + r_i \right) \quad (k \in \mathcal{U}_l, k \neq i)$$

$$f_{iB}^{(l)} = \frac{V_{iB}}{V_{iB} + \sum_{k \in \mathcal{U}_l}^{k \neq i} V_{ik}} \left(\sum_{m \in \mathcal{U}_l}^{m \neq i} f_{mi}^{(l)} + r_i \right)$$

   where the $f_{mi}^{(l)}$ values, if not zero, have all been defined before calculating the flow routing for node i.

4. *Let $\mathcal{U}_l = \mathcal{U}_l - \{i\}$ and go to Step 2.*

As shown in this algorithm, for each time interval $(\mu_{l-1}, \mu_l]$, $l = 1,\ 2, \cdots,\ L$, we initialize $\mathcal{U}_l$ as the set of remaining alive nodes at this stage, which is represented by $\mathcal{U}_l = \mathcal{N} - \bigcup_{j=0}^{l-1} \mathcal{N}_j$. For these nodes, we compute a flow routing by starting with the "boundary" nodes and then move

[2]Such a node must exist when $\mathcal{U}_l \neq \emptyset$, since there is no cycle in an LMM-optimal solution, i.e., we do not have $V_{i_1,i_2}, \cdots, V_{i_{k-1},i_k}, V_{i_k,i_2} > 0$. Otherwise, by reducing these volumes a little further, we can increase the corresponding nodes' lifetimes.

to the "interior" nodes. More precisely, we will calculate the flow routing for a node i if and only if we have calculated the flow routing for each node m that has traffic coming into node i. The out-going flow at node i is calculated by distributing the aggregated in-coming flow proportionally according to the overall bit volume along its out-going radio links. As an example, suppose that during $(\mu_4, \mu_5]$, node 5 receives an aggregated flow of rate 2 kb/s and generates 0.4 kb/s amount of data locally. Assume that $V_{24} = 100$ kb, $V_{25} = 200$ kb, and $V_{2B} = 300$ kb over $[0, \mu_L]$. Then the out-going flow at node 2 is routed as follows: $f_{24}^{(5)} = 0.4$ kb/s, $f_{25}^{(5)} = 0.8$ kb/s, and $f_{2B}^{(5)} = 1.2$ kb/s.

We now give a formal proof that the flow routing solution defined by Algorithm 2 will indeed give the LMM-optimal node lifetime vector.

Proof. For $t \in (\mu_{l-1}, \mu_l]$, denote $f_{ik}(t) = f_{ik}^{(l)}$ and $f_{iB}(t) = f_{iB}^{(l)}$, $l = 1,\ 2,\ \cdots,\ L$; $\gamma_i(t) = r_i$ for $t \leq t_i$ and $\gamma_i(t) = 0$ for $t > t_i$. To show that the flow routing solution defined in Algorithm 2 indeed gives the LMM-optimal node lifetime vector, it is sufficient to show that each node $i \in \mathcal{N}$ has lifetime t_i under this flow routing solution, i.e.,

$$\int_{t=0}^{t_i} \left[\sum_{m \in \mathcal{N}}^{m \neq i} \beta_{\text{rec}} f_{mi}(t) + \sum_{k \in \mathcal{N}}^{k \neq i} \beta_{ik} f_{ik}(t) + \beta_{iB} f_{iB}(t) \right] dt = e_i \ . \tag{2.13}$$

To show that (2.13) is true, it is sufficient to show that $\int_{t=0}^{t_i} f_{mi}(t)dt = V_{mi}$, $\int_{t=0}^{t_i} f_{ik}(t)dt = V_{ik}$, and $\int_{t=0}^{t_i} f_{iB}(t)dt = V_{iB}$ hold. This is equivalent to showing that

$$\int_{t=0}^{T} f_{mi}(t)dt = V_{mi} \ , \tag{2.14}$$

$$\int_{t=0}^{T} f_{ik}(t)dt = V_{ik} \ , \tag{2.15}$$

$$\int_{t=0}^{T} f_{iB}(t)dt = V_{iB} \ , \tag{2.16}$$

and

$$f_{mi}(t) = 0 \quad \text{for } t > t_i, \tag{2.17}$$

$$f_{ik}(t) = 0 \quad \text{for } t > t_i, \tag{2.18}$$

$$f_{iB}(t) = 0 \quad \text{for } t > t_i. \tag{2.19}$$

We first prove (2.17), (2.18), and (2.19) hold. For $t > t_i$, suppose that $t \in (\mu_{j-1}, \mu_j]$. Then we have $\mu_{j-1} \geq t_i$. Since under Algorithm 2, positive flow routing for $f_{ik}^{(l)}$ and $f_{iB}^{(l)}$ are only defined for $t_i > \mu_{l-1}$, we have $f_{ik}^{(j)} = 0$ and $f_{iB}^{(j)} = 0$. Consequently, $f_{ik}(t) = 0$ and $f_{iB}(t) = 0$ for $t > t_i$, i.e., (2.18) and (2.19) both hold. Now we show that (2.17) also holds. If $V_{mi} = 0$, then $f_{mi}(t) = 0$ for $t > t_i$ holds trivially. If $V_{mi} > 0$, we must have $t_i \geq t_m$ under the LMM-optimal solution, otherwise we can obtain a better solution by decreasing V_{mi} while increasing V_{mB}. Since we have $\mu_{j-1} \geq t_i$, then $\mu_{j-1} \geq t_m$. Consequently, $f_{mi}(t) = 0$ for $t > t_i$ by the flow routing construction in Algorithm 2.

To show that (2.14) holds, it is sufficient to show that (2.15) holds for $i, k \in \mathcal{N}, i \neq k$. We now show that this is true.

1. Suppose that node i is a "boundary" node that does not receive any flow from other nodes. Hence, we have $V_{mi} = 0$, and so, $f_{mi}(t) = 0$. Consequently,

$$\begin{aligned}
\int_{t=0}^{T} f_{ik}(t)dt &= \sum_{l=1}^{L} \int_{t=\mu_{l-1}}^{\mu_l} \frac{V_{ik}}{V_{iB} + \sum_{k\in\mathcal{U}_l}^{k\neq i} V_{ik}} \cdot \gamma_i(t)dt \\
&= \sum_{l=1}^{L} \int_{t=\mu_{l-1}}^{\mu_l} \frac{V_{ik}}{V_{iB} + \sum_{k\in\mathcal{N}}^{k\neq i} V_{ik}} \cdot \gamma_i(t)dt \\
&= \frac{V_{ik}}{V_{iB} + \sum_{k\in\mathcal{N}}^{k\neq i} V_{ik}} \cdot r_i t_i \\
&= \frac{V_{ik}}{V_{iB} + \sum_{k\in\mathcal{N}}^{k\neq i} V_{ik}} \left(V_{iB} + \sum_{k\in\mathcal{N}}^{k\neq i} V_{ik} \right) = V_{ik} \ .
\end{aligned}$$

   The second equality holds since for $k \notin \mathcal{U}_l$, we have $t_k \leq \mu_{l-1} < t_i$. Thus, we have $V_{ik} = 0$. The third equality holds since $\gamma_i(t) = r_i$ for $t \in [0,\ t_i]$ and $\gamma_i(t) = 0$ otherwise. The fourth equality follows since the bit volumes V_{ik} and V_{iB} must meet the volume balance property at node i.

2. Now, let us suppose that node i is not a "boundary" node and thus will receive flow from some nodes m. Based on the selection of node i, node m's out-going flows have already been defined. Moreover, it is supposed that node m has already met the criteria in (2.14),

particularly, $\int_{t=0}^{T} f_{mi}(t)dt = V_{mi}$. Based on the definition for flow routing in Algorithm 2, we have

$$\begin{aligned}
\int_{t=0}^{T} f_{ik}(t)dt &= \sum_{l=1}^{L} \int_{t=\mu_{l-1}}^{\mu_l} \frac{V_{ik}}{V_{iB} + \sum_{k\in\mathcal{U}_l}^{k\neq i} V_{ik}} \left[\sum_{m\in\mathcal{U}_l}^{m\neq i} f_{mi}(t) + \gamma_i(t) \right] dt \\
&= \sum_{l=1}^{L} \int_{t=\mu_{l-1}}^{\mu_l} \frac{V_{ik}}{V_{iB} + \sum_{k\in\mathcal{N}}^{k\neq i} V_{ik}} \left[\sum_{m\in\mathcal{N}}^{m\neq i} f_{mi}(t) + \gamma_i(t) \right] dt \\
&= \frac{V_{ik}}{V_{iB} + \sum_{k\in\mathcal{N}}^{k\neq i} V_{ik}} \left(\sum_{m\in\mathcal{N}}^{m\neq i} V_{mi} + r_i t_i \right) \\
&= \frac{V_{ik}}{V_{iB} + \sum_{k\in\mathcal{N}}^{k\neq i} V_{ik}} \left(V_{iB} + \sum_{k\in\mathcal{N}}^{k\neq i} V_{ik} \right) = V_{ik} \ .
\end{aligned}$$

The second equality holds since for $m \notin \mathcal{U}_l$, we have $f_{mi} = 0$ by (2.18) and for $k \notin \mathcal{U}_l$, we have $V_{ik} = 0$. The third equality holds since $\int_{t=0}^{T} f_{mi}(t)dt = V_{mi}$ (which we have proved) and $\int_{t=0}^{T} \gamma_i(t)dt = r_i t_i$. The fourth equation holds since the bit volumes V_{ik} and V_{iB} must meet the volume balance property at node i.

Combining (i) and (ii), we have proved that (2.15) holds for $i, k \in \mathcal{N}, i \neq k$. Following the same argument, we can prove that (2.16) also holds for $i \in \mathcal{N}$. □

An important property is that the flow routing solution during each interval $(\mu_{l-1}, \mu_l]$, $l = 1,\ 2, \cdots,\ L$, can be calculated independently. Consequently, this property enables a parallel computation of the flow routing solution, i.e., for computing the flow routing for all L intervals at the same time.

2.3.4 Computational Complexity Analysis

Complexity of SLP-PA. We now analyze the complexity of our SLP-PA approach to solve the LMM node lifetime problem. First we consider the complexity of finding each node's lifetime and the total bit volume transmitted along each link. At each stage, we solve an LP problem. After solving an LP at each stage, we need to determine whether or not a node that just reached

its energy binding constraint belongs to the minimum node set for this stage. Note that $\mathbf{v}$ and $\hat{\mathbf{b}} = \mathcal{B}^{-1}\mathbf{b}$ can be readily obtained when we solve the primal LP problem. To determine whether a node, say i belongs to the minimum node set, we examine v_i. If $v_i < 0$, then node i belongs to the minimum node set and the complexity is $O(1)$. On the other hand, if $v_i = 0$, we need to further examine whether $\hat{\varepsilon}_i > 0$ or not. Based (2.12), the computation for $\hat{\varepsilon}_i$ is $O(N)$. So at each stage, the complexity in PA for each node is $O(N)$. The total complexity of PA at each stage for the node set is thus $|\hat{\mathcal{N}}_l| \cdot O(N)$ or $O(N \cdot N) = O(N^2)$ for the non-degenerate case, which is smaller than the complexity of an LP. Thus, the major complexity at each stage is solving an LP. As there are at most N stages, the overall complexity is at most of solving N LPs.

For the degenerate case, upon the completion of Step 2 in Algorithm 1, we denote $\mathcal{U}_l^{(0)} = \mathcal{U}_l$. Since we need to solve at most $|\mathcal{U}_l^{(0)} - \mathcal{N}_l| \leq N$ LPs. Hence, the complexity at each stage is at most solving N LPs. Since there are at most N stages, the overall complexity is at most of solving N^2 LPs.

We now analyze the complexity to find the routing solution. At each stage l, we first need to compute $f_{ik}^{(l)}$ and $f_{iB}^{(l)}$ with complexity $O(N^2)$. Since there are at most N stages, the overall complexity is $N \cdot O(N^2) = O(N^3)$, which is smaller than the complexity of an LP. Now combining the complexity of both parts in our approach, the overall complexity is at most of solving N LPs for the non-degenerate case and of solving N^2 LPs for the degenerate case. Both are polynomial.

Comparison of SLP-SV. We now compare the complexity of our approach with the SLP-SV approach in [15]. First of all, SLP-SV needs to keep track of each sub-flow along its route from the source node toward the base station. Such a flow-based (or more precisely, sub-flow based) approach usually makes the size of the LP coefficient matrix exponential, which leads to an exponential-time algorithm even with the most efficient LP technique [6].[3] Second, even if a link-based LP formulation such as ours is adopted in [15], the computational efficiency of SV-based approach would be still worse than SLP-PA. This is because that at each stage, the SV-based approach in [15] solves several additional LPs (up to $|\hat{\mathcal{N}}_l - \mathcal{N}_l|$) to determine $\mathcal{N}_l$, in contrast with

[3] Incidentally, the revised simplex method proposed in [15] is not as efficient as that in [6] and is itself exponential.

the simpler parametric analysis for the SLP-PA approach, which only involves $O(N^2)$ effort for the non-degenerate case. Even for the degenerate case, the number of additional LPs is up to $|\mathcal{U}_l^{(0)} - \mathcal{N}_l|$ ($\leq |\hat{\mathcal{N}}_l - \mathcal{N}_l|$). Consequently, for any problem, our approach is computationally more efficient than the SLP-SV approach in [15].

Finally, we discuss a hybrid link-flow approach mentioned in [15]. This approach requires a sub-flow accounting on each link and results in an order of magnitude more constraints than the link-based approach proposed in this chapter. Although this approach can solve the LMM node lifetime problem in polynomial-time (e.g., using interior point methods [6]), the overall complexity is still orders of magnitude higher than that for our proposed SLP-PA approach. Furthermore, there remains the additional burden associated with the SLP-SV approach for solving the additional LPs even using the hybrid link-flow based approach.

2.4 Extension to LMM Rate Allocation Problem

In this section, we show that our SLP-PA algorithm can be used to solve the *LMM rate allocation problem.*

2.4.1 The LMM Rate Allocation Problem

The LMM rate allocation problem considers the following scenario. For a network with a set $\mathcal{N}$ of AFNs, with a network lifetime requirement T and initial energy e_i for AFN $i \in \mathcal{N}$, how can we maximize the data rate for *all* AFNs in the network?

More formally, denote r_i the rate for each AFN $i \in \mathcal{N}$. Note that T is fixed here, while r_i's are the optimization variables, which are different from the LMM node lifetime problem that we studied in the last section. Denote $[g_1, g_2, \cdots, g_N]$ as the *sorted* sequence of the r_i values in nondecreasing order. Then LMM-optimal rate vector can be defined as follows.

Definition 2. (LMM-optimal Rate Allocation) A sorted rate vector $[g_1, g_2, \cdots, g_N]$ with $g_1 \leq$

$g_2 \leq \cdots \leq g_N$ is LMM-optimal if and only if for any other sorted rate vector $[\hat{g}_1, \hat{g}_2, \cdots, \hat{g}_N]$ with $\hat{g}_1 \leq \hat{g}_2 \leq \cdots \leq \hat{g}_N$, there exists a $k \leq N$, such that $g_i = \hat{g}_i$ for $1 \leq i \leq k-1$ and $g_k > \hat{g}_k$.

2.4.2 Solution

It should be clear that, under the LMM-optimal rate allocation objective, we must *maximize* each rate level that a set of nodes use up their energy while *minimizing* the number of nodes that drain up their energy at each rate level. We now show that the SLP-PA algorithm developed for the LMM node lifetime problem can be directly applied to solve the LMM rate allocation problem.

Suppose that $[g_1, g_2, \cdots, g_N]$ with $g_1 \leq g_2 \leq \cdots \leq g_N$ is LMM-optimal. To keep track of *distinct* rate levels in this vector, we remove all repetitive elements in the vector and rewrite it as $[\lambda_1, \lambda_2, \cdots, \lambda_L]$ such that $\lambda_1 < \lambda_2 < \cdots < \lambda_L$, where $\lambda_1 = g_1, \lambda_L = g_N$, and $L \leq N$. Corresponding to these rate levels, denote $\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_L$ as the sets of nodes that drain up their energy at rate levels $\lambda_1, \lambda_2, \cdots, \lambda_L$, respectively. Then $\bigcup_{l=1}^{L} \mathcal{N}_l = \mathcal{N}$. The problem is to find the LMM-optimal values of $\lambda_1, \lambda_2, \cdots, \lambda_L$ and the corresponding sets $\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_L$.

Similar to the LMM node lifetime problem, the LMM rate allocation problem can be formulated as an iterative optimization problem as follows. Denote $\lambda_0 = 0$, $\mathcal{N}_0 = \emptyset$, and $\delta_l = \delta_l - \delta_{l-1}$. Starting from $l = 1$, we solve the following LP iteratively.

$$
\begin{aligned}
\underline{\textbf{LMM-Rate:}}\ \text{Max} \quad & \delta_l & & \\
\text{s.t.} \quad & V_{iB} + \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} V_{ik} - \sum_{m\in\mathcal{N}}^{m\neq i} V_{mi} - \delta_l T = \lambda_{l-1} T & & (i \notin \bigcup_{j=1}^{l-1} \mathcal{N}_j) \\
& V_{iB} + \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} V_{ik} - \sum_{m\in\mathcal{N}}^{m\neq i} V_{mi} = \lambda_h T & & (i \in \mathcal{N}_h, 1 \leq h < l) \\
& \textstyle\sum_{m\in\mathcal{N}}^{m\neq i} \beta_{\text{rec}} V_{mi} + \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{ik} V_{ik} + \beta_{iB} V_{iB} \leq e_i & & (i \notin \bigcup_{j=1}^{l-1} \mathcal{N}_j) \\
& \textstyle\sum_{m\in\mathcal{N}}^{m\neq i} \beta_{\text{rec}} V_{mi} + \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{ik} V_{ik} + \beta_{iB} V_{iB} = e_i & & (i \in \mathcal{N}_h, 1 \leq h < l) \\
& V_{ik}, V_{iB}, \delta_l \geq 0 & & (i, k \in \mathcal{N}, k \neq i)
\end{aligned}
\tag{2.20}
$$

Comparing the above LMM-Rate problem to the LMM-Lifetime problem that we studied in

Section 2.3.1, we find that they are exactly of the same form. The only differences are that under the LMM-Rate problem, the network lifetime T is a constant and the rates r_i are variables (subject to optimization), while under the LMM-Lifetime problem, the t_i are variables (subject to optimization) and the node rates r_i are given, $i \in \mathcal{N}$. Since the mathematical formulation for the two problems are identical, we can apply the SLP-PA algorithm to solve the LMM rate allocation problem as well.

2.5 Duality Theorem

In this section, we present an elegant and powerful result showing that there is a duality relationship between the LMM node lifetime problem and the LMM rate allocation problem. As a result, it is only necessary to solve only one of the two problems and the results for the other can be obtained via simple algebraic calculations.

To start with, we denote $\mathcal{P}_L$ the LMM node lifetime problem where N AFNs have the same local bit rate R (constant). Denote t_i the LMM node lifetime for node $i \in \mathcal{N}$ under $\mathcal{P}_L$. Similarly, we denote $\mathcal{P}_R$ the LMM rate allocation problem where all nodes have a common given lifetime requirement T (constant). Denote r_i the LMM-optimal rate allocation for node $i \in \mathcal{N}$ under $\mathcal{P}_R$. Then the following theorem shows how the solution to one problem can be used to obtain the solution to the other.

Theorem 1. (Duality) For a given local bit rate R for all nodes under problem $\mathcal{P}_L$ and a given node lifetime requirement T for all nodes under problem $\mathcal{P}_R$, we have the following relationship between the solutions to these two problems.

(i) Suppose that we have solved problem $\mathcal{P}_L$ and obtained the LMM-optimal node lifetime t_i for each node $i \in \mathcal{N}$. Then under $\mathcal{P}_R$, the LMM rate allocation r_i for node i is

$$r_i = \frac{t_i R}{T} \,. \tag{2.21}$$

(ii) Suppose that we have solved problem $\mathcal{P}_R$ and obtained the LMM-optimal rate allocation r_i for

Table 2.2: Duality relationship between LMM node lifetime problem $\mathcal{P}_L$ and LMM rate allocation problem $\mathcal{P}_R$.

$\mathcal{P}_L$	$\mathcal{P}_R$
t_i (optimization variable)	$t_i = T$ (constant)
$r_i = R$ (constant)	r_i (optimization variable)
Total bit volume at AFN i: $r_i \cdot T = t_i \cdot R$	

each node $i \in \mathcal{N}$. Then under $\mathcal{P}_L$, the LMM node lifetime t_i for node i is

$$t_i = \frac{r_i T}{R} \ .$$

Table 2.2 shows the duality relationship between solutions to problems $\mathcal{P}_L$ and $\mathcal{P}_R$.

Proof. We prove (i) and (ii) in Theorem 1 separately.

(i) We organize our proof into two parts. First, we show that r_i's are feasible in terms of flow balance and energy constraints on each node $i \in \mathcal{N}$. Then we show that it is indeed the LMM-optimal rate allocation.

Feasibility. Since we have obtained the solution to problem $\mathcal{P}_L$, we have one feasible flow routing solution for sending bit streams R from each AFN $i \in \mathcal{N}$ to the base station for time $[0, t_i]$. Under problem $\mathcal{P}_L$, the bit volumes (V_{ij} and V_{iB} values) must meet the following equalities under the LMM-optimal node lifetime:

$$V_{iB} + \sum_{k \in \mathcal{N}}^{k \neq i} V_{ik} - \sum_{m \in \mathcal{N}}^{m \neq i} V_{mi} = t_i R \ ,$$

$$\sum_{m \in \mathcal{N}}^{m \neq i} \beta_{\text{rec}} V_{mi} + \sum_{k \in \mathcal{N}}^{k \neq i} \beta_{ik} V_{ik} + \beta_{iB} V_{iB} = e_i \ .$$

Now replacing $t_i R$ by $r_i T$, we see that the *same* bit volume solution yields a feasible bit volume solution to the rate allocation problem $\mathcal{P}_R$.

Optimality. To prove that r_i's obtained via (2.21) are indeed LMM-optimal for problem $\mathcal{P}_R$, we sort t_i's in non-decreasing order and denote it as $[\tau_1, \tau_2, \cdots, \tau_N]$. We also introduce a node index $I = [i_1, i_2, \cdots, i_N]$ for $[\tau_1, \tau_2, \cdots, \tau_N]$. For example, $i_3 = 7$ means that τ_3 actually corresponds to the lifetime of AFN 7, i.e., $\tau_3 = t_7$.

Since r_i is proportional to t_i through the relationship ($r_i = \frac{R}{T} \cdot t_i$), listing r_i's according to $I = [i_1, i_2, \cdots, i_N]$ will yield a *sorted* (in *non-decreasing* order) rate list, denoted as $[g_1, g_2, \cdots, g_N]$. We now prove that $[g_1, g_2, \cdots, g_N]$ is indeed LMM-optimal for problem $\mathcal{P}_R$.

Our proof is based on contradiction. Suppose that $[g_1, g_2, \cdots, g_N]$ is not LMM-optimal for problem $\mathcal{P}_R$. Assume that the LMM-optimal rate vector to problem $\mathcal{P}_R$ is $[\hat{g}_1, \hat{g}_2, \cdots, \hat{g}_N]$ (sorted in non-decreasing order) with the corresponding node index being $\hat{I} = [\hat{i}_1, \hat{i}_2, \cdots, \hat{i}_N]$. Then, by Definition 1, there exists a $k \leq N$ such that $\hat{g}_j = g_j$ for $1 \leq j \leq k-1$ and $\hat{g}_k > g_k$.

We now claim that $\hat{\tau}_i$ obtained via $\hat{\tau}_i = \frac{\hat{g}_i T}{R}$, $i = 1, 2, \cdots, N$, is also a feasible solution to problem $\mathcal{P}_L$. The proof to this claim follows identically as above. Using this result, we can obtain a corresponding feasible solution $[\hat{\tau}_1, \hat{\tau}_2, \cdots, \hat{\tau}_N]$ with $\hat{\tau}_i = \frac{\hat{g}_i T}{R}$ and the node index $\hat{I}$ for problem $\mathcal{P}_L$. Hence we have $\hat{\tau}_j = \frac{\hat{g}_j T}{R} = \frac{g_j T}{R} = \tau_j$ for $1 \leq j \leq k-1$ but $\hat{\tau}_k = \frac{\hat{g}_k T}{R} > \frac{g_k T}{R} = \tau_k$. That is, $[\tau_1, \tau_2, \cdots, \tau_N]$ is not LMM-optimal and this leads to a contradiction.

(ii) The proof for this part follows the same token as the above proof for (i) and is thus omitted here. □

2.6 Numerical Investigation

In this section, we use numerical results to illustrate the solution of LMM node lifetime problem and compare our SLP-PA to some other approaches. In particular, we will compare SLP-PA with the naive approach (see Section 2.2) that uses a serial LP "blindly" to solve the LMM node lifetime problem. We call this naive approach SLP-ER (serial LP with energy reservation). As discussed in Section 2.2, the naive SLP-ER approach requires energy reservation at each stage and cannot give

Table 2.3: Locations (in meters) for each AFN in a 10-node network.

Node Index	Location	Node Index	Location
1	(400, -320)	6	(-500, 100)
2	(300, 440)	7	(-400, 0)
3	(-300, -420)	8	(420, 120)
4	(320, -100)	9	(200, 140)
5	(340, -120)	10	(220, -340)

the correct LMM-optimal solution. We also compare our SLP-PA approach with the *Minimum-Power Routing* (MPR) approach that has been considered in the literature (e.g. [30, 43, 59, 66, 67, 76, 92, 96]) and is used to achieve energy efficiency. Under the MPR approach, an AFN always chooses the path that consumes the minimum amount of power toward the base station.

We also use numerical results to illustrate the duality between the LMM node lifetime problem and the LMM rate allocation problem.

2.6.1 Network Configurations and Parameter Settings

We consider two network topologies. The first network consists of 10 AFNs while the second network consists of 20 AFNs. Under each network, the base station B is located at the origin $(0,\ 0)$ (in meters). The locations for these 10 and 20 AFNs are generated at random and are shown in Tables 2.3 and 2.4, respectively.

2.6.2 SLP-PA Algorithm to the LMM Node Lifetime Problem

10-AFN network. We assume that the initial energy at each AFN is 50 kJ and local data generated by each AFN is 0.2 kb/s. We set $\beta_1 = 50$ nJ/b, $\beta_2 = 0.0013$ pJ/b/m^4, $\beta_{\text{rec}} = 50$ nJ/b, and path loss index $\alpha = 4$ [46].

Table 2.4: Locations (in meters) for each AFN in a 20-node network.

Node Index	Location	Node Index	Location
1	(200, 130)	11	(110, -230)
2	(-400, -380)	12	(-210, 0)
3	(-100, 420)	13	(210, 320)
4	(0, 430)	14	(300, -480)
5	(-410, 440)	15	(-420, -420)
6	(-200, 230)	16	(120, -240)
7	(400, -490)	17	(220, -440)
8	(410, -300)	18	(-220, 240)
9	(100, 310)	19	(-500, -110)
10	(100, 140)	20	(0, 330)

Table 2.5: Node lifetime (in days) under the three approaches for the 10-node network.

Sorted	SLP-PA		SLP-ER		MPR	
Index	τ_i	AFN	τ_i	AFN	τ_i	AFN
1	45.71	3	45.71	1	28.91	7
2	45.71	6	45.71	2	46.09	3
3	45.71	7	45.71	3	61.63	6
4	146.08	1	45.71	5	87.75	9
5	146.08	2	45.71	6	92.77	4
6	146.08	4	45.71	7	118.79	5
7	146.08	5	45.71	10	142.96	8
8	146.08	8	303.70	4	150.29	2
9	146.08	9	303.70	8	157.62	10
10	146.08	10	303.70	9	182.55	1

Table 2.5 gives each AFN's lifetime (days) under each approach.[4] The "sorted index" column represents the node index, in which the AFNs are sorted by their node lifetimes in nondecreasing order. Clearly, the node lifetime vector under SLP-PA dominates that under the SLP-ER and MPR approaches with respect to the LMM-optimal node lifetime definition (Definition 1). For example, comparing the node lifetime vector under SLP-PA and SLP-ER, we find that $\tau_1^{\text{SLP-PA}} = \tau_1^{\text{SLP-ER}}$, $\tau_2^{\text{SLP-PA}} = \tau_2^{\text{SLP-ER}}$, $\tau_3^{\text{SLP-PA}} = \tau_3^{\text{SLP-ER}}$, and $\tau_4^{\text{SLP-PA}} > \tau_4^{\text{SLP-ER}}$. Similarly, comparing the node lifetime vector under SLP-PA and MPR, we have $\tau_1^{\text{SLP-PA}} > \tau_1^{\text{MPR}}$. In general, τ_1^{MPR} (28.91 days) is the smallest among the three approaches (45.71 days under both SLP-PA and SLP-ER) since minimum power routing does not guarantee a good performance with respect to node lifetime performance. Although SLP-ER and SLP-PA have the same node lifetime (45.71 days) at the first stage, SLP-PA gives a smaller AFN set ($|\mathcal{N}_1|^{\text{SLP-PA}} = 3$) at this lifetime level than SLP-ER ($|\mathcal{N}_1|^{\text{SLP-ER}} = 7$), which shows that the naive SLP-ER approach cannot offer the correct solution to the LMM node lifetime problem.

Another way to visualize the LMM node lifetime performance in Table 2.5 is to plot the total number of remaining "alive" AFNs over time, which is given in Fig. 2.2. Viewing Fig. 2.2(a), in the beginning, all 10 AFNs are alive. As time goes on, one AFN under MPR drains its energy (at 28.91 days) and the remaining alive AFNs drop to 9. Under both the SLP-PA and SLP-ER approaches, the first lifetime level is 45.71 days, during which 3 AFNs drain their energy under SLP-PA while 7 AFNs drain energy under SLP-ER. Among the three approaches, only the SLP-PA provides a node lifetime solution that meets the LMM-optimal definition (Definition 1).

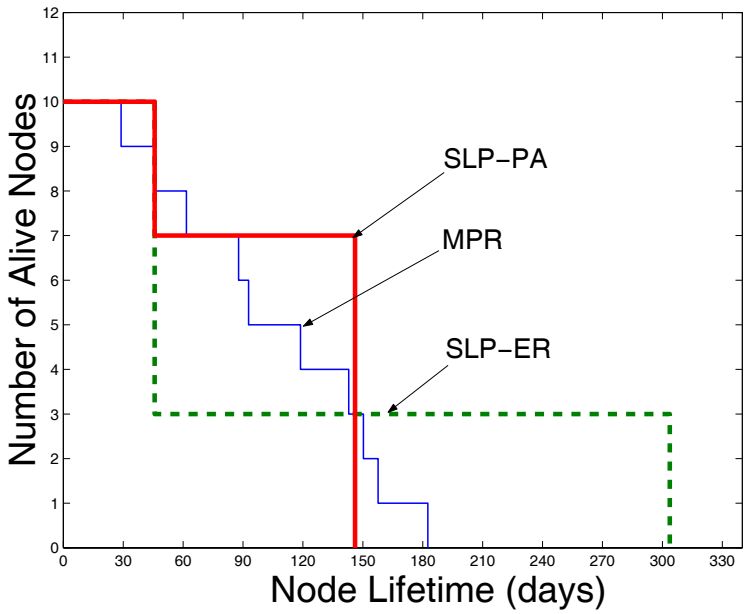
20-AFN network. For the 20-AFN network (Table 2.4), we assume that the initial energy at each AFN is 50 kJ and that the local data generated by each AFN is 0.5 kb/s. Table 2.6 shows the sorted node lifetime performance under the three approaches. Plots for the node lifetime curve are given in Fig. 2.2(b). Again, we have similar observations as those for the 10-AFN network.

Flow routing solution. We now show how to use Algorithm 2 to calculate a flow routing solution that achieves the LMM-optimal node lifetime vector for the 10-AFN network. Under the SLP-PA approach, we have $\mu_1 = 45.71$ days with $\mathcal{N}_1 = \{3,\ 6,\ 7\}$ and $\mu_2 = 146.08$ days with

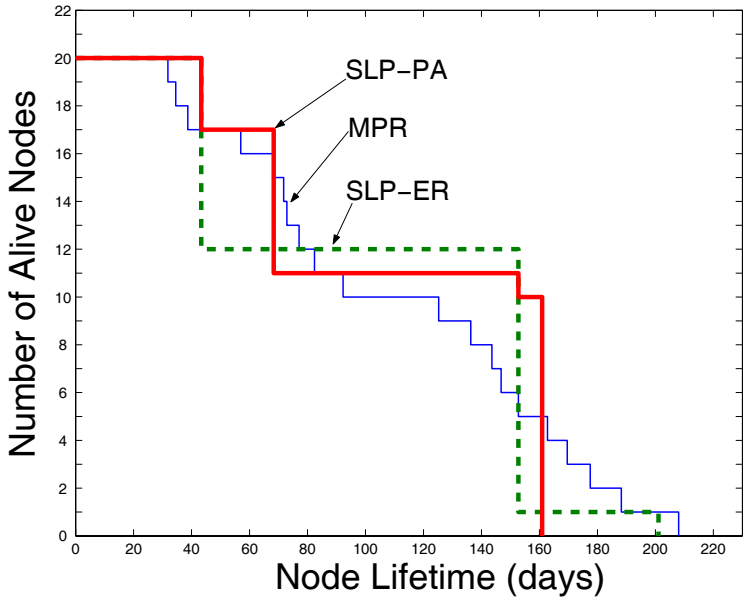
[4]The results for the SLP-SV are not shown since they are the same as those under SLP-PA. The difference is in the computational complexity.

Table 2.6: Node lifetime (in days) under the three approaches for the 20-node network.

Sorted	SLP-PA		SLP-ER		MPR	
Index	τ_i	AFN	τ_i	AFN	τ_i	AFN
1	43.35	2	43.35	2	31.85	19
2	43.35	15	43.35	7	34.54	11
3	43.35	19	43.35	8	38.72	2
4	68.32	7	43.35	14	56.99	15
5	68.32	8	43.35	15	67.98	16
6	68.32	11	43.35	16	71.79	8
7	68.32	14	43.35	17	72.88	17
8	68.32	16	43.35	19	77.08	14
9	68.32	17	152.72	1	82.40	7
10	152.72	5	152.72	3	92.27	10
11	160.91	1	152.72	4	125.25	6
12	160.91	3	152.72	5	136.33	1
13	160.91	4	152.72	6	143.59	12
14	160.91	6	152.72	9	146.77	9
15	160.91	9	152.72	10	152.72	5
16	160.91	10	152.72	12	162.77	20
17	160.91	12	152.72	13	169.59	18
18	160.91	13	152.72	18	177.54	13
19	160.91	18	152.72	20	188.26	4
20	160.91	20	201.09	11	208.04	3



(a) 10-node network.



(b) 20-node network.

Figure 2.2: Node lifetime curves under the three approaches.

$\mathcal{N}_2 = \{1,\ 2,\ 4,\ 5,\ 8,\ 9,\ 10\}$. Also, we obtain the following bit volumes (all in 10^4 kb) among the nodes from the last LMM-Lifetime solution:

$V_{1,5} = 320.0419,\ \ V_{1,B} = 46.7550;$

$V_{2,9} = 233.8006,\ \ V_{2,B} = 18.6306;$

$V_{3,7} = 48.6548,\ \ V_{3,B} = 30.3317;$

$V_{4,B} = 303.3560;$

$V_{5,4} = 50.9249,\ \ V_{5,8} = 390.6881,\ \ V_{5,B} = 130.8601;$

$V_{6,7} = 22.2673,\ \ V_{6,B} = 56.7191;$

$V_{7,B} = 149.9086;$

$V_{8,9} = 576.2578,\ \ V_{8,B} = 66.8615;$

$V_{9,B} = 1062.4895;$

$V_{10,1} = 114.3658,\ \ V_{10,B} = 138.0654.$

We now find the flow routing solution for each interval, i.e., $[0,\ \mu_1]$ and $(\mu_1,\ \mu_2]$, respectively. For time interval $[0, \mu_1]$, we get the following.

- Nodes 2, 3, 6, and 10 do not receive any data. Using Algorithm 2, node 2 sends 0.185 kb/s to node 9 and 0.015 kb/s to the base station B. Similarly, node 3 sends 0.123 kb/s to node 7 and 0.077 kb/s to the base station B; node 6 sends 0.057 kb/s to node 7 and 0.143 kb/s to the base station B; node 10 sends 0.091 kb/s to node 1 and 0.109 kb/s to the base station B.

- Now, since the in-coming flow to nodes 1 and 7 are defined, we can calculate their out-going flow rates. Using Algorithm 2, node 1 sends 0.254 kb/s to node 5 and 0.037 kb/s to the base station B; node 7 sends 0.380 kb/s to the base station B.

- Next, we consider node 5. After calculation, we find that node 5 should send 0.040 kb/s to node 4, 0.310 kb/s to node 8, and 0.104 kb/s to the base station B.

- Following this, we consider nodes 4 and 8. We find that node 4 sends 0.240 kb/s to the base station B; node 8 sends 0.457 kb/s to node 9 and 0.053 kb/s to the base station B.

- Finally, we consider node 9. Using Algorithm 2, we find that node 9 sends 0.842 kb/s to the base station B.

In summary, during $[0,\ \mu_1] = [0, 45.71]$, we have the following flow rates (all in kb/s):

$f_{1,5} = 0.254, \quad f_{1,B} = 0.037;$

$f_{2,9} = 0.185, \quad f_{2,B} = 0.015;$

$f_{3,7} = 0.123, \quad f_{3,B} = 0.077;$

$f_{4,B} = 0.240;$

$f_{5,4} = 0.040, \quad f_{5,8} = 0.310, \quad f_{5,B} = 0.104;$

$f_{6,7} = 0.057, \quad f_{6,B} = 0.143;$

$f_{7,B} = 0.380;$

$f_{8,9} = 0.457, \quad f_{8,B} = 0.053;$

$f_{9,B} = 0.842;$

$f_{10,1} = 0.091, \quad f_{10,B} = 0.109.$

Likewise, applying Algorithm 2 during $(\mu_1,\ \mu_2] = (45.71, 146.08]$, we obtain the following flow routing rates (all in kb/s):

$f_{1,5} = 0.254, \quad f_{1,B} = 0.037;$

$f_{2,9} = 0.185, \quad f_{2,B} = 0.015;$

$f_{4,B} = 0.240;$

$f_{5,4} = 0.040, \quad f_{5,8} = 0.310, \quad f_{5,B} = 0.104;$

$f_{8,9} = 0.457, \quad f_{8,B} = 0.053;$

$f_{9,B} = 0.842;$

$f_{10,1} = 0.091, \quad f_{10,B} = 0.109.$

It is easy to verify that above flow routing solution will indeed obtain the LMM-optimal node lifetime vector.

The flow routing solution that achieves the LMM-optimal node lifetime vector for the 20-AFN network can be obtained in a similar manner (by using Algorithm 2). This is omitted for the sake of brevity.

Table 2.7: Numerical results verifying the duality relationship $T \cdot r_i = R \cdot t_i$ between the LMM rate allocation problem $\mathcal{P}_R$ and the LMM node lifetime problem $\mathcal{P}_L$ for the 10-node network.

AFN	$\mathcal{P}_L$ ($R = 0.2$ Kb/s)		$\mathcal{P}_R$ ($T = 100$ days)	
	t_i	$R \cdot t_i$	r_i	$T \cdot r_i$
1	146.08	29.22	0.2922	29.22
2	146.08	29.22	0.2922	29.22
3	45.71	9.14	0.0914	9.14
4	146.08	29.22	0.2922	29.22
5	146.08	29.22	0.2922	29.22
6	45.71	9.14	0.0914	9.14
7	45.71	9.14	0.0914	9.14
8	146.08	29.22	0.2922	29.22
9	146.08	29.22	0.2922	29.22
10	146.08	29.22	0.2922	29.22

2.6.3 Duality Results

We now use numerical results to verify the duality relationship between the LMM node lifetime problem $\mathcal{P}_L$ and the LMM rate allocation problem $\mathcal{P}_R$ (see Section 2.5). Again, we use the 10-AFN and 20-AFN network configurations in Tables 2.3 and 2.4, respectively. For both networks, we assume that the initial energy at each AFN is 50 KJ. Under $\mathcal{P}_R$, the network lifetime requirement is $T = 100$ days. For the 10-AFN network, the local bit rate for all AFNs is $R = 0.2$ Kb/s under $\mathcal{P}_L$. For the 20-AFN network, the local bit rate for all AFNs is $R = 0.5$ Kb/s under $\mathcal{P}_L$.

To verify the duality relationship in Theorem 1, we perform the following calculations. First, we solve the LMM node lifetime problem $\mathcal{P}_L$ and the LMM rate allocation problem $\mathcal{P}_R$ *independently* with the above initial conditions using the SLP-PA algorithm. Consequently, we obtain the LMM-optimal node lifetime (t_i for each AFN i) under $\mathcal{P}_L$ and the LMM-optimal rate allocation (r_i for each AFN i) under $\mathcal{P}_R$. Then we compute $R \cdot t_i$ and $T \cdot r_i$ separately for each AFN i and examine if they are equal to each other.

Table 2.8: Numerical results verifying the duality relationship $T \cdot r_i = R \cdot t_i$ between the LMM rate allocation problem $\mathcal{P}_R$ and the LMM node lifetime problem $\mathcal{P}_L$ for the 20-node network.

AFN	$\mathcal{P}_L$ ($R = 0.5$ Kb/s)		$\mathcal{P}_R$ ($T = 100$ days)	
	t_i	$R \cdot t_i$	r_i	$T \cdot r_i$
1	160.91	80.46	0.8046	80.46
2	43.35	21.68	0.2168	21.68
3	160.91	80.46	0.8046	80.46
4	160.91	80.46	0.8046	80.46
5	152.72	76.36	0.7636	76.36
6	160.91	80.46	0.8046	80.46
7	68.32	34.16	0.3416	34.16
8	68.32	34.16	0.3416	34.16
9	160.91	80.46	0.8046	80.46
10	160.91	80.46	0.8046	80.46
11	68.32	34.16	0.3416	34.16
12	160.91	80.46	0.8046	80.46
13	160.91	80.46	0.8046	80.46
14	68.32	34.16	0.3416	34.16
15	43.35	21.68	0.2168	21.68
16	68.32	34.16	0.3416	34.16
17	68.32	34.16	0.3416	34.16
18	160.91	80.46	0.8046	80.46
19	43.35	21.68	0.2168	21.68
20	160.91	80.46	0.8046	80.46

The results for the LMM-optimal node lifetime (t_i, $i = 1, 2, \cdots, 10$) and the LMM-optimal rate allocation (r_i, $i = 1, 2, \cdots, 10$) for the 10-AFN network are shown in Table 2.7. We find that $R \cdot t_i$ and $T \cdot r_i$ are exactly equal for all AFNs, precisely as we would expect under Theorem 1. Similarly, the results for the 20-AFN network are shown in Table 2.8.

2.7 Related Work

Due to energy constraints in wireless sensor networks, there has been active research on exploring the performance limits of such networks. These performance limits include, among others, *network lifetime* and *network capacity*. Network lifetime refers to the maximum time that the nodes in the network remain alive before one or more nodes deplete their energy, while network capacity typically refers to the maximum amount of bit volume that can be successfully delivered to the base station ("sink node") by all the nodes in the network.

The network lifetime problem and network capacity problem have so far been studied disjointly in the literature. For example, in many other efforts (e.g., [11, 13, 18, 55, 115]), the focus was on how to maximize the time until the first node drains up its energy. While, in [56], the problem of how to maximize network capacity via routing was studied.

In this chapter, we study the important overarching problem that considers both network lifetime and network capacity. The notion of network lifetime for wireless sensor networks has been studied in [10, 11, 13, 17–19, 55, 115]. The notion of network lifetime discussed in these work focuses on the time until the first node fails without further consideration of the remaining nodes in the network. As wireless sensor networks will typically remain useful even if some nodes run out of energy, it is essential to further investigate how to maximize the lifetime for all the remaining nodes in the network, which is the focus of the LMM node lifetime problem. Under the LMM node lifetime problem, we studied how to maximize the lifetime for *all* nodes when the local bit rate for each node is given. The closest work related to the LMM node lifetime problem is that in [15], which has been discussed in detail in this chapter.

Under the LMM rate allocation problem, we studied how to maximize rate allocations for *all* the nodes in the network under a given network lifetime requirement. The LMM rate allocation criterion effectively mitigates the unfairness issue when the objective is to maximize the total bit volume generated by the network. Although the LMM rate allocation is somewhat similar to the classical max-min strategy [9], there is a fundamental difference between the two. In particular, the LMM rate allocation problem implicitly embeds (or couples) a flow routing problem within rate allocation, while under the classical max-min rate allocation, there is no routing problem involved since the routes for all flows are given. Due to this coupling of flow routing and rate allocation, a solution approach (i.e., SLP-PA) to the LMM rate allocation problem is much more challenging than that for the classical max-min. In [94], Srinivasan et al. applied game theory and Nash equilibrium among the nodes to forward packets such that the total throughput (capacity) can achieve an optimal operating point subject to a common lifetime requirement on all nodes. However, the fairness issue in information collection was not considered.

2.8 Conclusions

In this chapter, we considered the problem of how to maximize the lifetime for all the nodes in a wireless sensor network. We formally defined this optimization problem as the lexicographic max-min (LMM) node lifetime problem and investigated approaches to solve it. The main contributions in this problem are two-fold. First, we developed a polynomial-time algorithm to obtain the LMM-optimal node lifetime vector, which improves upon the computational complexity associated with a state-of-the-art algorithm. Second, we discussed how to obtain the routing solution among the AFNs such that the LMM-optimal node lifetime vector can be achieved. We show that the SLP-PA can also be employed to address the LMM rate allocation problem. More important, we discovered a simple and elegant duality relationship between the LMM node lifetime problem and the LMM rate allocation problem, which enables us to develop solutions and insights on both problems by solving one of the two problems. The results in this chapter help lay the essential algorithmic foundation for studying network lifetime problems in energy-constrained wireless sensor networks.

Chapter 3

Base Station Placement for Wireless Sensor Networks

3.1 Introduction

An important characteristic for wireless sensor networks is the so-called network lifetime performance, which is highly dependent upon the physical topology of the network. This is because energy expenditure at a node to transmit data to another node not only depends on the data bit rate, but also on the physical distance between these two nodes. Consequently, it is important to understand the impact of location related issues on network lifetime performance and to optimize topology during network deployment stage.

This chapter considers the important problem of base station placement for a given sensor network such that network lifetime can be maximized. Specifically, we consider the following problem. Given a sensor network with each node i producing data at a rate of r_i, where should we place the base station in this sensor network such that all the data can be forwarded to the base station (via multi-hop and multi-path if necessary) such that the network lifetime is maximized?

In Section 3.5, we give a comprehensive review of related work on network lifetime and node

placement problems and contrast their differences with this work. The most important work on this problem is done by Efrat et al. [32], which represents the state-of-the-art result on this problem. However, the computational complexity associated with the algorithm in [32] is higher than the one to be presented in this chapter for most cases.

The main idea in our approximation algorithm is to exploit a clever way to discretize cost parameters associated with energy consumption with tight upper and lower bounds. As a result, we can divide the continuous search space into a finite number of subareas. By further exploiting the cost property of each subarea, we conceive a novel idea to represent each subarea with a so-called "fictitious cost point" (FCP), which is an N-tuple cost vector with each component representing the upper bound of cost to a sensor node in the network. Based on these ideas, we have successfully reduced an infinite search space for base station location into a finite "points" upon which we can apply a linear program (LP) to find the corresponding achievable network lifetime and data routing solution. By comparing the achievable network lifetime among all the FCPs, we show that the largest one is $(1 - \varepsilon)$-optimal. We also prove that locating the base station at any point in the subarea corresponding to the best FCP is $(1 - \varepsilon)$-optimal. We analyze the complexity of our approximation algorithm and show that it is practically faster than the algorithm proposed in [32] for most cases, which was the best known result on this problem. As a result, the algorithm presented in this chapter represents the best known result to the base station placement problem.

The remainder of the chapter is organized as follows. Section 3.2 presents the network model used in this study and describes the base station placement problem. Section 3.3 presents the main result of this chapter, which is a $(1 - \varepsilon)$-approximation algorithm for the base station placement problem. In Section 3.4, we present some numerical results illustrating the efficacy of the proposed algorithm. Section 3.5 reviews related work and Section 3.6 concludes this chapter.

Table 3.1: Notation in Chapter 3.

p_m	Fictitious cost point for subarea m
p_{opt}	The best base station location
p^*	The best fictitious cost point
p_ε	A point in the subarea corresponding to p^*
T_m	Maximum achievable network lifetime by p_m
T_{opt}	Optimal network lifetime achieved by p_{opt}
T^*	$= \max\{T_m : m = 1, 2, \cdots, M\}$
T_ε	$(1-\varepsilon)$-optimal network lifetime achieved by p_ε
ψ_{opt}	The best routing solution for p_{opt}
ψ^*	The routing solution for p^* obtained via a linear program (LP)
ψ_ε	Routing solution for p_ε obtained via an LP

3.2 Network Model and Problem Description

3.2.1 Network Model

We consider a sensor network consisting a set $\mathcal{N}$ of sensor nodes deployed over a two-dimensional area. The location of each sensor node is fixed and the initial energy on sensor node i is denoted as e_i. Each sensor node i generates data at a rate r_i. We assume there is one base station that needs to be deployed to collect sensing data.

Power consumption model has been discussed in Section 2.2.2. It is easy to observe from (2.1), (2.2), and (2.3) that the *location for the base station* and *data routing* in the network both have a profound impact on energy consumption among the nodes. Table 3.1 lists all notation that are used only in this chapter.

3.2.2 Problem Description

The focus of this chapter is to investigate how to optimally place a base station to collect data in a sensor network so that the network lifetime can be maximized. The network lifetime is defined as the time until any sensor node uses up its energy. To achieve optimality, the data generated by each sensor node is allowed to be routed to the base station via multi-hop or multi-path. Also, power control at a node is allowed, as modeled in (2.1) and (2.2).

Assume that base station B is located at a point p. Denote (x_B, y_B) the location of base station B and T the network lifetime. A feasible routing solution achieving this network lifetime T should satisfy both flow balance in (2.4) and energy constraints in (2.5) at each sensor node. By (2.2), we have

$$\beta_{iB}(p) = \beta_1 + \beta_2 \left[\sqrt{(x_B - x_i)^2 + (y_B - y_i)^2}\right]^{\alpha} ,$$

which is a non-linear function of base station location (x_B, y_B).

Our objective is maximizing the network lifetime T under the flow balance and energy constraints, i.e.,

$$\begin{aligned}
\text{Max} \quad & T \\
\text{s.t.} \quad & \textstyle\sum_{j\in\mathcal{N}}^{j\neq i} f_{ij} + f_{iB} - \sum_{k\in\mathcal{N}}^{k\neq i} f_{ki} = r_i && (i \in \mathcal{N}) && (3.1)\\
& \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki} T + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij} T + \beta_{iB}(p) f_{iB} T \leq e_i && (i \in \mathcal{N}) && (3.2)\\
& \beta_{iB}(p) - \beta_2 \left[\sqrt{(x_B - x_i)^2 + (y_B - y_i)^2}\right]^{\alpha} = \beta_1 && (i \in \mathcal{N}) \\
& (x_B, y_B) \in \mathcal{A}, T, f_{ij}, f_{iB} \geq 0 && (i, j \in \mathcal{N}, i \neq j) ,
\end{aligned}$$

where $\mathcal{A}$ is an area of all possible base station locations. This optimization problem is in the form of *non-linear program*.

3.3 An Approximation Algorithm

3.3.1 Our Approach

As discussed earlier, network lifetime depends on both base station location and data routing. To start with, we show that for a *given* base station location, we can find the corresponding maximum achievable network lifetime and optimal routing via a single linear program (LP) as follows.

The objective function is network lifetime T and the constraints are given in (3.1) and (3.2). Multiply both sides of (3.1) by T and denote $V_{ij} = f_{ij}T$ and $V_{iB} = f_{iB}T$, where V_{ij} (or V_{iB}) can be interpreted as the total data volume from sensor node i to sensor node j (or base station B). We have

$$
\begin{array}{lll}
\text{Max} & T & \\
\text{s.t.} & \sum_{k\in\mathcal{N}}^{k\neq i} V_{ki} + r_i T - \sum_{j\in\mathcal{N}}^{j\neq i} V_{ij} - V_{iB} = 0 & (i \in \mathcal{N}) \\
& \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} V_{ki} + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} V_{ij} + \beta_{iB}(p) V_{iB} \leq e_i & (i \in \mathcal{N}) \\
& T, V_{ij}, V_{iB} \geq 0 & (i, j \in \mathcal{N}, i \neq j)
\end{array}
$$

It should be note that for a given base station location, $\beta_{iB}(p)$'s are constants. Once we solve the above LP, we can obtain optimal routing solution for f_{ij} and f_{iB} by $f_{ij} = \frac{V_{ij}}{T}$ and $f_{iB} = \frac{V_{iB}}{T}$.

Although for a given base station location, we can find the corresponding maximum achievable network lifetime via a single LP, it is not possible to examine all points (infinite) in the two-dimensional plane and select the point with the maximum network lifetime among all the points. Our approach to this search problem is to narrow down the search space to a *finite* set of points, among which there exist at least one point with $(1-\varepsilon)$-optimal network lifetime.

As a first step, we show that it is only necessary to consider points inside the so-called *smallest enclosing disk* (SED) [106],[1] which is a unique disk with the smallest radius that contains all the

[1] In fact, we can consider points in an even smaller area, i.e., the convex hull of all sensor nodes. However, using convex hull cannot reduce the order of complexity of our algorithm. On the other hand, the use of SED can simplify the discussion.

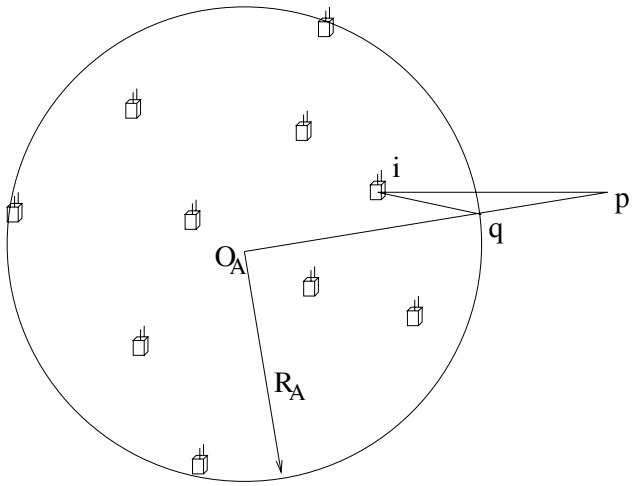


Figure 3.1: A schematic diagram showing that optimal base station location must be within SED.

N sensor nodes in the network and can be found in $O(N)$ time [25]. This is formally stated in the following lemma.

Lemma 3. *To maximize network lifetime, the base station location must be within the smallest enclosing disk $\mathcal{A}$ that covers all sensor nodes in the network.*

Proof. The proof is based on contradiction. That is, if the base station location is not in SED, then its network lifetime can be further increased and thus this location cannot be maximum.

Assume that the optimal base station location is at point p, which is outside SED $\mathcal{A}$. Denote $O_{\mathcal{A}}$ the center of SED (see Fig. 3.1). Let q be the intersecting point between the line segment $[p, O_{\mathcal{A}}]$ and the circle of SED. Then for any sensor i (all in $\mathcal{A}$), we have $d_{iq} < d_{ip}$. Consequently, $\beta_{iq} < \beta_{ip}$. As a result, we can save transmission energy on every sensor node $i \in \mathcal{N}$ by relocating p to q and thereby increase network lifetime. This completes the proof. □

Now we have narrowed down the search space for base station B from a two-dimensional plane to SED $\mathcal{A}$. However, the number of points in $\mathcal{A}$ remains infinite. It is tempting to divide $\mathcal{A}$ into

small subareas (e.g., a grid-like structure), $\mathcal{A}_1, \mathcal{A}_2, \cdots$, up to say $\mathcal{A}_M$, i.e.,

$$\mathcal{A} = \bigcup_{m=1}^{M} \mathcal{A}_m \, .$$

When each subarea is sufficiently small (i.e., M is sufficiently large), we can use some point $q_m \in \mathcal{A}_m$ to represent $\mathcal{A}_m$, $m = 1, 2, \cdots, M$. By applying an LP on each of the M points, we can find the best location and obtain a good solution for base station placement. However, such approach is *heuristic* at best and does not provide any *theoretical guarantee* on performance.

The key to provide a theoretical guarantee on performance is to divide SED $\mathcal{A}$ in such a way that certain bounds can be imposed on any point in a subarea. If this is possible, then we can exploit such properties and develop an approximation algorithm that yields *provably* $(1-\varepsilon)$-optimal network lifetime performance. In the following section, we show a novel technique to divide SED $\mathcal{A}$ into subareas where each subarea can be represented by a point with a set of tight bounds. Consequently, a $(1-\varepsilon)$-optimal approximation algorithm can be developed.

3.3.2 Subarea Division and Fictitious Cost Points

Subarea Division

The proposed subarea division (with guaranteed performance bounds) hinges upon discretization of the cost parameters. A close look at the energy constraint in (3.2) suggests that the location of the base station is *embedded* in the cost parameter $\beta_{iB}(p)$'s. In other words, if we can discretize these cost parameters, we may also discretize the location for the base station.

Since the search space is narrowed down to the SED $\mathcal{A}$, we can limit the range for the distance between a sensor node i to the possible location for the base station. Denote $O_{\mathcal{A}}$ and $R_{\mathcal{A}}$ the origin and radius of the SED $\mathcal{A}$. For each sensor node $i \in \mathcal{N}$, denote $d_{i,O_{\mathcal{A}}}$ the distance from sensor node i to the origin of disk $\mathcal{A}$ (see node 4 in Fig. 3.2 as an example). Denote $d_{iB}^{\min}$ and $d_{iB}^{\max}$ the minimum and maximum distances between sensor node i and possible location for the base station

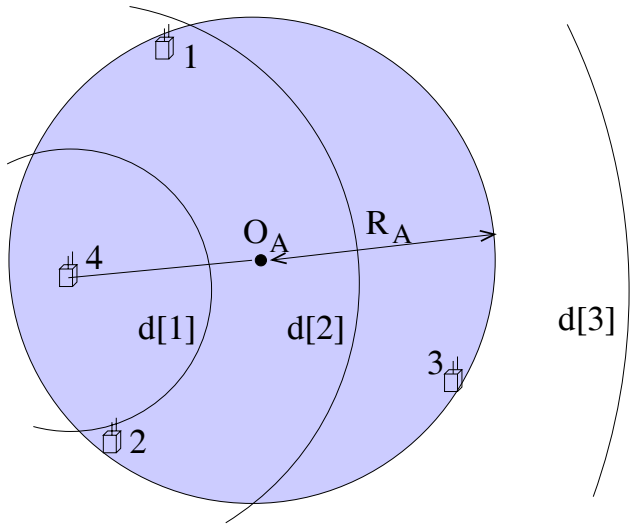


Figure 3.2: A sequence of circles with increasing costs at centering node 4.

B, respectively. Then we have

$$d_{iB}^{\min} = 0 \quad \text{and} \quad d_{iB}^{\max} = d_{i,O_{\mathcal{A}}} + R_{\mathcal{A}} \ .$$

Corresponding to $d_{iB}^{\min}$ and $d_{iB}^{\max}$, denote $\beta_{iB}^{\min}$ and $\beta_{iB}^{\max}$ the minimum and maximum cost between sensor node i and base station B, respectively. Then by (2.2), we have

$$\beta_{iB}^{\min} = \beta_1 \tag{3.3}$$

$$\beta_{iB}^{\max} = \beta_1 + \beta_2 (d_{iB}^{\max})^{\alpha} = \beta_1 + \beta_2 (d_{i,O_{\mathcal{A}}} + R_{\mathcal{A}})^{\alpha} \ . \tag{3.4}$$

Given the range of $d_{iB} \in [d_{iB}^{\min}, d_{iB}^{\max}] = [0, d_{i,O_{\mathcal{A}}} + R_{\mathcal{A}}]$ for each sensor node i, we now show how to divide disk $\mathcal{A}$ into a finite number of subareas with the distance of each subarea to sensor node i meeting some tight bounds. Specifically, from a sensor node i, we draw a sequence of circles centered at this sensor node, each with increasing radius $d[1], d[2], \cdots, d[H_i]$ corresponding to costs $\beta[1], \beta[2], \cdots, \beta[H_i]$ that are defined as follows.

$$\beta[h] = \beta_{iB}^{\min}(1+\varepsilon)^h = \beta_1(1+\varepsilon)^h \quad (1 \le h \le H_i) \tag{3.5}$$

The geometric series $\beta[h]$ (with a factor of $(1+\varepsilon)$) is carefully chosen and will offer tight performance bounds for any point in a subarea (more on this later). The number of required circles

H_i can be determined by having the last circle in the sequence (with radius $d[H_i]$) to completely contain disk $\mathcal{A}$, i.e. $d[H_i] \geq d_{iB}^{\max}$, or equivalently, $\beta[H_i] \geq \beta_{iB}^{\max}$. This will set the limit H_i as

$$H_i = \left\lceil \frac{\ln(\beta_{iB}^{\max}/\beta_{iB}^{\min})}{\ln(1+\varepsilon)} \right\rceil = \left\lceil \frac{\ln(1+\frac{\beta_2}{\beta_1}(d_{i,O_{\mathcal{A}}}+R_{\mathcal{A}})^{\alpha})}{\ln(1+\varepsilon)} \right\rceil \ . \tag{3.6}$$

For example, for node 4 in Fig. 3.2, we have $H_4 = 3$, i.e., $d[3]$ is the circle centered at node 4 that will completely contain the disk. As a result, with sensor node i as center, we have a total of H_i circles, each with cost $\beta[h]$, $h = 1, 2, \cdots, H_i$.

The above partitioning of SED $\mathcal{A}$ is in respect to a specific node i. We now perform the above process for all sensor nodes. These intersecting circles will cut disk $\mathcal{A}$ into a finite number of *non-uniform* subareas, with the boundaries of each subarea being either an arc (with a center at some sensor node i and some cost $\beta[h]$, $1 \leq h < H_i$) or an arc from SED $\mathcal{A}$. As an example, the SED $\mathcal{A}$ in Fig. 3.3 is now divided into 28 subareas.

Now we claim that under this subarea partitioning technique, for any point in a given subarea, its cost to each sensor node in the network can be *tightly* bounded. This is because with respect to each sensor node i, a subarea $\mathcal{A}_m$ must be enclosed within some ring centered at at sensor node i. Denote the index of this ring (w.r.t. sensor node i) as $h_i(\mathcal{A}_m)$. So when the base station B is at any point $p \in \mathcal{A}_m$, we have

$$\beta[h_i(\mathcal{A}_m) - 1] \leq \beta_{iB}(p) \leq \beta[h_i(\mathcal{A}_m)] \ , \tag{3.7}$$

where we define $\beta[0] = \beta_{iB}^{\min} = \beta_1$. Since $\frac{\beta[h_i(\mathcal{A}_m)]}{\beta[h_i(\mathcal{A}_m)-1]} = 1+\varepsilon$ by (3.5), we have a very tight upper and lower bounds for $\beta_{iB}(p)$ (now we see the benefit of our discretization technique for cost and distance).

Fictitious Cost Point

We now introduce a novel concept called *fictitious cost point* (FCP). It will be used to represent upper bound of cost for any point in a subarea $\mathcal{A}_m$, $m = 1, 2, \cdots, M$.

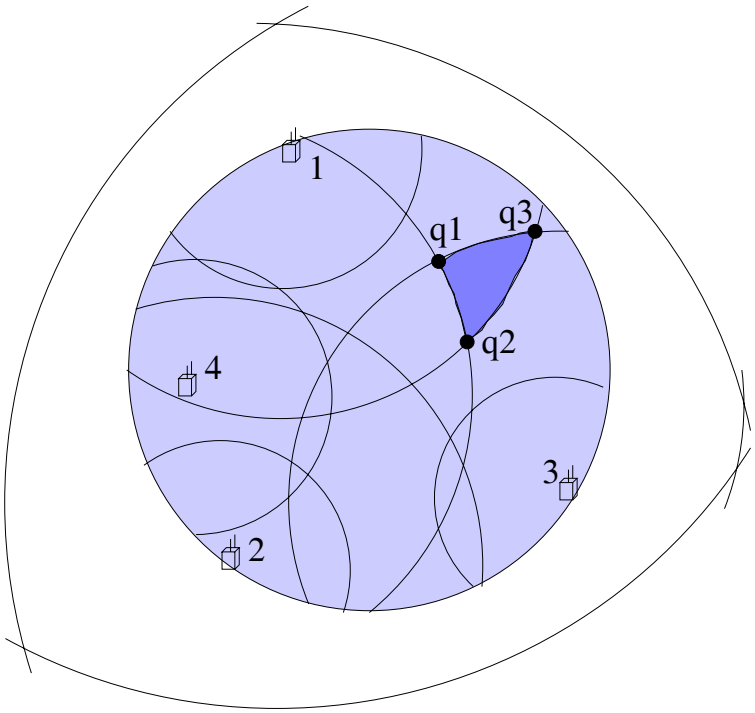


Figure 3.3: An example of subareas within disk $\mathcal{A}$ that are obtained by intersecting arcs from different circles.

Definition 3. *Denote the fictitious cost point for subarea $\mathcal{A}_m$ ($m = 1, 2, \cdots, M$) as p_m, which is represented by an N-tuple vector $[\beta_{1B}(p_m), \beta_{2B}(p_m), \cdots, \beta_{NB}(p_m)]$, with its i-th element ($i \in \mathcal{N}$) being upper cost bound for any point in subarea $\mathcal{A}_m$ to sensor node i, i.e.,*

$$\beta_{iB}(p_m) = \beta[h_i(\mathcal{A}_m)] \ , \tag{3.8}$$

where $h_i(\mathcal{A}_m)$ is determined by (3.7).

As an example, the FCP for subarea with corner points (q_1, q_2, q_3) in Fig. 3.3 can be represented by 4-tuple cost vector $[\beta_{1B}(p_m), \beta_{2B}(p_m), \beta_{3B}(p_m), \beta_{4B}(p_m)] = [\beta[2], \beta[3], \beta[2], \beta[3]]$, where the first component $\beta[2]$ represents an upper bound of cost for any point in this subarea to sensor node 1, the second component $\beta[3]$ represents an upper bound of cost (which is loose here) for any point in this subarea to sensor node 2, and so forth.

We emphasize that the reason we use the word "fictitious" is that an FCP p_m may not be

mapped to a *physical* point within the corresponding subarea $\mathcal{A}_m$. This happens when there does not exist a physical point in subarea $\mathcal{A}_m$ that has its costs to all the N sensor nodes *equal* (one by one) to the respective N-tuple cost vector embodied by p_m *simultaneously.* As an example, any point within the dark subarea bounded by corner points q_1, q_2, and q_3 cannot have its costs to the four sensor nodes in the network equal to the respective element in $[\beta[2], \beta[3], \beta[2], \beta[3]]$ simultaneously, where $[\beta[2], \beta[3], \beta[2], \beta[3]]$ is the cost vector of the FCP for this subarea.

Using "fictitious" points to represent subareas and to construct a finite search space is a key step in design our low complexity approximation algorithm. As a contrast, in [32], the authors use physical points to construct a finite search space. Thus, they cannot discretize cost directly. Instead, they consider how to discretize transmission energy, flow rate, and network lifetime such that cost can be discretized. The number of discretized costs is the product of the numbers of discretized transmission energies, flow rates, and network lifetimes. We will show that the complexity associated with the algorithm in [32] is higher than ours for most cases.

The following important property for FCP p_m will be used in the proof of $(1-\varepsilon)$-optimality of the approximation algorithm.

Property 1. *Denote p_m the FCP for subarea $\mathcal{A}_m$, $m = 1, 2, \cdots, M$. For any point $p \in \mathcal{A}_m$, we have*

$$\beta_{iB}(p_m) \leq (1+\varepsilon)\beta_{iB}(p) \ .$$

Proof. By definition of FCP p_m in (3.8), we have

$$\frac{\beta_{iB}(p_m)}{\beta_{iB}(p)} = \frac{\beta[h_i(\mathcal{A}_m)]}{\beta_{iB}(p)} = \frac{(1+\varepsilon)\cdot\beta[h_i(\mathcal{A}_m)-1]}{\beta_{iB}(p)} \leq \frac{(1+\varepsilon)\cdot\beta_{iB}(p)}{\beta_{iB}(p)} = 1+\varepsilon \ ,$$

where the inequality follows from (3.7). This completes the proof. □

3.3.3 Summary of Algorithm and Example

By discretizing the cost parameters and the corresponding distances, we have partitioned the search space (SED $\mathcal{A}$) into a finite number (M) subareas. By introducing the concept of FCP, we can

represent each subarea with a point. As a result, we can now readily apply the LP approach discussed in Section 3.3.1 to examine each point and choose the point that offers the maximum network lifetime. The complete approximation algorithm is outlined below. The correctness proof of $(1-\varepsilon)$-optimality is given in Section 3.3.4.

Algorithm 3. (A $(1-\varepsilon)$-Approximation Algorithm)

1. Find the smallest enclosing disk $\mathcal{A}$ that covers all N nodes.
2. Within $\mathcal{A}$, compute the lower and upper cost bounds $\beta_{iB}^{\min}$ and $\beta_{iB}^{\max}$ for each node $i \in \mathcal{N}$ by (3.3) and (3.4).
3. For a given $\varepsilon > 0$, define a sequence of costs $\beta[1], \beta[2], \cdots, \beta[H_i]$ by (3.5), where H_i can by calculated by (3.6).
4. For each node i, draw a sequence of $H_i - 1$ circles corresponding to cost $\beta[h]$ centered at node i, $1 \leq h < H_i$. The intersection of these circles within disk $\mathcal{A}$ will divide $\mathcal{A}$ into M subareas $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_M$.
5. For each subarea $\mathcal{A}_m$, $1 \leq m \leq M$, define an FCP p_m, which embodies an N-tuple cost vector $[\beta_{1B}(p_m), \beta_{2B}(p_m), \cdots, \beta_{NB}(p_m)]$, where $\beta_{iB}(p_m)$ is defined in (3.8).
6. For each of these M FCPs, apply the LP in Section 3.3.1 and obtain the maximum achievable network lifetime.
7. Choose the FCP p^* that offers the maximum network lifetime among these M FCPs. The base station can be placed at any point p_ε within the subarea corresponding to p^*.
8. For point p_ε, apply the LP in Section 3.3.1 and obtain a routing solution ψ_ε with $(1-\varepsilon)$-optimal network lifetime T_ε.

Remark. Note that the above algorithm can be easily extended to accommodate any additional area constraint on base station placement. That is, we can define $\mathcal{A}$ in Step 1 to be the intersection between SED and the allowed area for base station placement.

Table 3.2: Each node's location, data rate, and initial energy of the example sensor network.

Location	Data rate	Initial energy
$(0.1, 0.5)$	0.8	390
$(1.1, 0.7)$	1.0	400
$(0.4, 0.1)$	0.5	130

Example. We use a small 3-node network to illustrate the steps of the approximation algorithm. The location, data rate, and initial energy for each sensor are shown in Table 3.2, where the units of distance, rate, and energy are all normalized. Also, we set $\alpha = 2$, $\beta_{\text{rec}} = 1$, $\beta_1 = 1$, and $\beta_2 = 0.5$ under the normalized units. For illustration, we set the error bound $\varepsilon = 0.2$.[2]

Step 1. We identify SED $\mathcal{A}$ with origin $O_{\mathcal{A}} = (0.61, 0.57)$ and radius $R_{\mathcal{A}} = 0.51$ (see Fig. 3.4).

Step 2. We first have $d_{i,O_{\mathcal{A}}} = R_{\mathcal{A}} = 0.51$ for each node i, $1 \le i \le 3$. We then find the lower and upper bounds of β_{iB} for each node i, $1 \le i \le 3$, as follows.

$$\begin{aligned} \beta_{iB}^{\min} &= \beta_1 = 1 \\ \beta_{iB}^{\max} &= \beta_1 + \beta_2 (d_{i,O_{\mathcal{A}}} + R_{\mathcal{A}})^{\alpha} = 1 + 0.5 \cdot (0.51 + 0.51)^2 = 1.52 \end{aligned}$$

Step 3. For each node i, $1 \le i \le 3$, we find

$$H_i = \left\lceil \frac{\ln(1 + \frac{\beta_2}{\beta_1}(d_{i,O_{\mathcal{A}}} + R_{\mathcal{A}})^{\alpha})}{\ln(1+\varepsilon)} \right\rceil = \left\lceil \frac{\ln(1 + \frac{0.5}{1}(0.51 + 0.51)^2)}{\ln(1+0.2)} \right\rceil = 3$$

and

$$\begin{aligned} \beta[1] &= \beta_1(1+\varepsilon) = 1 \cdot (1 + 0.2) = 1.20 \ , \\ \beta[2] &= \beta_1(1+\varepsilon)^2 = 1 \cdot (1 + 0.2)^2 = 1.44 \ , \\ \beta[3] &= \beta_1(1+\varepsilon)^3 = 1 \cdot (1 + 0.2)^3 = 1.73 \ . \end{aligned}$$

Step 4. We draw circles centered at each node i with cost $\beta[h]$, $1 \le h < H_i = 3$, to divide the whole disk $\mathcal{A}$ into 16 subareas $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_{16}$.

[2]This ε is used here for illustration only. In our numerical results in Section 3.4, we use $\varepsilon = 0.05$ for all computations.

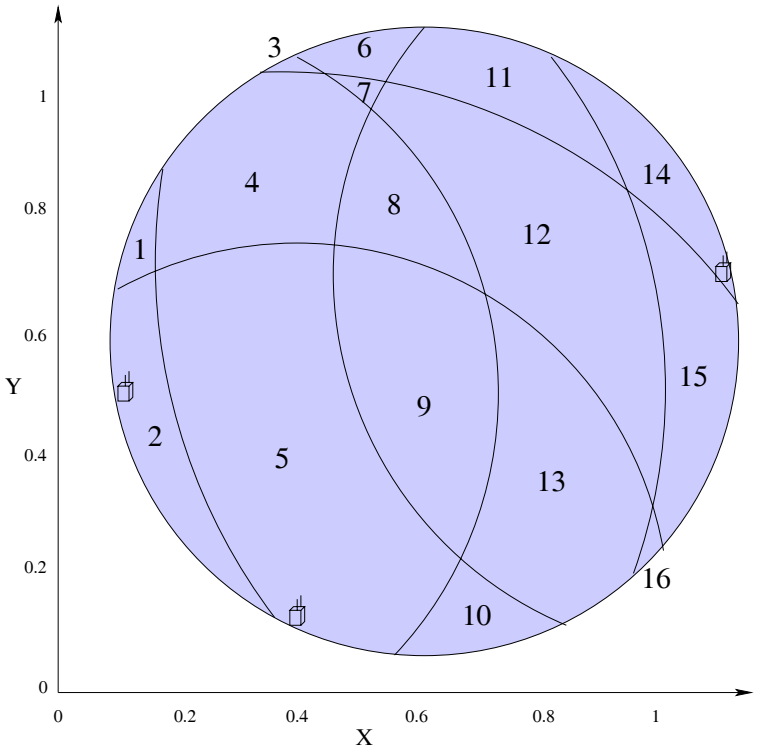


Figure 3.4: The SED is divided into 16 subareas in the example sensor network.

Step 5. We define an FCP p_m for each subarea $\mathcal{A}_m$, $1 \leq m \leq 16$. For example, for FCP p_1, we define the 3-tuple cost vector as $[\beta_{1B}(p_1),\ \beta_{2B}(p_1), \beta_{3B}(p_1)] = [\beta[1], \beta[3], \beta[2]] = [1.20, 1.73, 1.44]$.

Step 6. We apply LP in Section 3.3.1 on each of these 16 FCPs and obtain the corresponding maximum achievable network lifetime.

Step 7. Since FCP $p^* = p_9$ achieves the largest network lifetime 226.47 among all 16 FCP, we can place the base station at any point in subarea $\mathcal{A}_9$, e.g. $p_\varepsilon = (0.6, 0.6)$.

Step 8. We apply LP in Section 3.3.1 on p_ε and obtain a $(1-\varepsilon)$-optimal network lifetime $T_\varepsilon = 227.07$. This completes the algorithm.

3.3.4 Proof and Complexity Analysis

In this section, we give a formal proof that the solution obtained by Algorithm 3 is $(1-\varepsilon)$-optimal and analyze its complexity.

Denote p_{opt} the optimal location for base station placement and T_{opt} and ψ_{opt} the corresponding maximum network lifetime and data routing solution, all of which are unknown. Denote p^* the best FCP among the M FCPs, based on their maximum achievable network lifetime performance. Denote T^* and ψ^* the corresponding network lifetime and data routing solution, i.e., $T^* = \max\{T_m : m = 1, 2, \cdots, M\}$.

Based on Step 7 in Algorithm 3, we choose a physical point p_ε in the subarea corresponding to p^* for base station placement. For point p_ε, denote the maximum achievable network lifetime as T_ε and corresponding routing solution as ψ_ε.

Our roadmap for the proof is as follows. In Theorem 2, we will prove that T^* for the best FCP p^* is at least $(1-\varepsilon)$ of the optimum, i.e., $T^* \geq (1-\varepsilon)T_{\text{opt}}$. Subsequently, in Theorem 3, we will show that for the physical point p_ε, the corresponding network lifetime T_ε is also at least $(1-\varepsilon)$ of the optimum, i.e., $T_\varepsilon \geq (1-\varepsilon)T_{\text{opt}}$.

Theorem 2. *For T^* and T_{opt} as defined, we have $T^* \geq (1-\varepsilon)T_{opt}$.*

To prove Theorem 2, we first present the following lemma, which is a general case for the theorem.

Lemma 4. *For any given base station location p, the corresponding optimal routing solution φ, and the network lifetime T, denote $\mathcal{A}_m$ the subarea that contains p. Then for the corresponding FCP p_m, the achievable network lifetime T_m is at least $(1-\varepsilon)$ of T, i.e., $T_m \geq (1-\varepsilon)\cdot T$.*

Proof. Instead of considering the optimal routing solution for FCP p_m, we use the same routing φ on p_m, which is clearly sub-optimal. That is, denoting $\hat{T}_m$ the network lifetime for FCP p_m under φ, we have $T_m \geq \hat{T}_m$.

Now we show $\hat{T}_m \geq (1-\varepsilon)\cdot T$. At time $(1-\varepsilon)\cdot T$, the total consumed energy on a node $i \in \mathcal{N}$ under φ is

$$\begin{aligned} & \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki}\cdot(1-\varepsilon)T + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij}\cdot(1-\varepsilon)T + \beta_{iB}(p_m) f_{iB}\cdot(1-\varepsilon)T \\ < & \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki} T + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij} T + (1-\varepsilon)\beta_{iB}(p) f_{iB}\cdot(1-\varepsilon)T \\ < & \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki} T + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij} T + \beta_{iB}(p) f_{iB} T \leq e_i \end{aligned}$$

The first inequality holds via Property 1. The last inequality holds by the energy constraint in routing solution φ for point p. Thus, the network lifetime $\hat{T}_m$ for FCP p_m under routing solution φ is at least $(1-\varepsilon)\cdot T$. As a result, we have $T_m \geq \hat{T}_m \geq (1-\varepsilon)\cdot T$. This completes the proof. □

With this Lemma 4, we are ready to prove Theorem 2.

Proof. Consider the special case of Lemma 4 that the given base station location p is the optimal location p_{opt}, with corresponding optimal data routing solution ψ_{opt}, and maximum network lifetime T_{opt}. Follow the same token in Lemma 4, we can find a corresponding subarea $\mathcal{A}_m$ that contains point p_{opt} and corresponding FCP p_m. As a result, for FCP p_m, we have $T_m \geq (1-\varepsilon)T_{\text{opt}}$. Thus, for the best FCP p^* among all FCPs, we have $T^* \geq T_m \geq (1-\varepsilon)T_{\text{opt}}$. This completes the proof. □

Theorem 2 guarantees that the best network lifetime among the M FCPs is at least $(1-\varepsilon)$ of T_{opt}. Now consider a point p_ε in the subarea represented by the best FCP p^*. We have the following theorem.

Theorem 3. *For T_ε and T_{opt} as defined, we have $T_\varepsilon \geq (1-\varepsilon)T_{opt}$.*

Proof. Denote $\hat{T}_\varepsilon$ the network lifetime for point p_ε under the same routing solution ψ^* for point p^*. Since ψ^* is a sub-optimal routing for p_ε, we have $T_\varepsilon \geq \hat{T}_\varepsilon$. Thus, to show $T_\varepsilon \geq (1-\varepsilon)T_{\text{opt}}$, we only need to show $\hat{T}_\varepsilon \geq T^* \geq (1-\varepsilon)T_{\text{opt}}$, where the second inequality follows from Theorem 2.

For each sensor node $i \in \mathcal{N}$, the total consumed energy under ψ^* by time T^* is

$$\begin{aligned} & \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki} T^* + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij} T^* + \beta_{iB}(p) f_{iB} T^* \\ \leq & \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki} T^* + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij} T^* + \beta_{iB}(p_m) f_{iB} T^* \leq e_i \ . \end{aligned}$$

The first inequality holds by (3.7) and (3.8). The second inequality holds by the energy constraint on p^* under routing solution ψ^*. Thus, the network lifetime $\hat{T}_\varepsilon$ for location p_ε under ψ^* is at least T^*. As a result, the maximum achievable network lifetime T_ε for location p_ε is at least $\hat{T}_\varepsilon \geq T^* \geq (1-\varepsilon) \cdot T_{\text{opt}}$. This completes the proof. □

The complexity of Algorithm 3 can be measured by the number of LPs that need to be solved, which is equal to the number of subareas M. Note that the boundaries of each subarea being either an arc centered at some sensor node i (with some cost $\beta[h]$, $1 \leq h < H_i$, where H_i is defined in (3.6)) or an arc of disk $\mathcal{A}$. Since there are $H_i - 1$ circles radiating from each sensor node i and one circle for disk $\mathcal{A}$, the number of circles is $K = 1 + \sum_{i\in\mathcal{N}}(H_i - 1)$. The maximum number of subareas that can be obtained by K circles is upper bounded by [25]

$$M \leq K^2 - K + 2 = O(K^2) = O\left(\left(\sum_{i\in\mathcal{N}} H_i\right)^2\right) = O\left(\left(\frac{N}{\varepsilon}\right)^2\right) \ . \tag{3.9}$$

As for comparison, the complexity of the approximation algorithm proposed in [32] can also be measured by the number of LPs that need to be solved, which is

$$\left\lfloor \frac{4}{\varepsilon} \right\rfloor \left\lceil \frac{\alpha \ln 2}{\ln(1+\varepsilon/8)} \right\rceil \left\lceil \frac{8\alpha\pi}{\varepsilon} \right\rceil \cdot \sum_{i\in\mathcal{N}} \left\lceil \frac{\ln\left(8N^3 \sum_{j\in\mathcal{N}} r_j/(\varepsilon r_i)\right)}{\ln(1+\varepsilon/8)} \right\rceil = O\left(\frac{N \ln(N/\varepsilon)}{\varepsilon^4}\right) \ . \tag{3.10}$$

We now compare (3.9) and (3.10). For a $(1-\varepsilon)$-approximation algorithm, it is important to have a low complexity for small error bound ε, such that we can obtain a solution that is very close to the optimum. For a given network (N is fixed), our algorithm has a much lower complexity than the algorithm in [32], i.e., $O(\varepsilon^{-2})$ is better than $O(\varepsilon^{-4}\ln(\varepsilon^{-1}))$ for $0 < \varepsilon \ll 1$. It is also important

Table 3.3: Each node's location, data rate and initial energy for a small 10-node network.

Location	Data rate	Initial energy	Location	Data rate	Initial energy
$(0.81, 0.86)$	0.7	390	$(0.25, 0.71)$	0.4	400
$(0.47, 0.44)$	1.0	440	$(0.28, 0.03)$	0.6	330
$(0.25, 0.36)$	0.2	440	$(0.48, 0.22)$	0.1	300
$(0.53, 0.16)$	0.8	410	$(0.66, 0.52)$	0.2	210
$(0.91, 0.86)$	0.1	320	$(0.44, 0.21)$	0.9	330

that a $(1-\varepsilon)$-approximation algorithm should have a low complexity on N, such that we can obtain a solution for a network with a large number of sensors. It seems that under a given ε, the algorithm in [32] has a better performance on N, i.e., $O(N \ln N)$ is better than $O(N^2)$. However, we should note that the $O()$ results do not show the constant coefficients for $N \ln N$ and N^2. Note that the constant coefficient for the algorithm in [32] is about $1.3 \cdot 10^4 \alpha^2$ (with $2 \leq \alpha \leq 4$), which is a very large number. Due to this large constant coefficient, the algorithm in [32] has a much higher computational complexity than our algorithm for most cases. For example, for all the four networks (with up to 100 nodes) in Section 3.4, numerical comparison show that the complexity of the algorithm in [32] is at least 10^6 times of the complexity of our algorithm.

3.4 Numerical Results

In this section, we apply the approximation algorithm on various network topology and use numerical results to demonstrate its efficacy. The units of distance, rate, and energy are all normalized appropriately. The normalized parameters in energy consumption model are $\beta_{\text{rec}} = \beta_1 = \beta_2 = 1$ and we set path loss index $\alpha = 2$.

We consider four randomly generated networks consisting of 10, 20, 50, and 100 nodes deployed over a 1×1 square. In all cases, the targeted accuracy for approximation algorithm is 0.95-optimal, i.e., $\varepsilon = 0.05$.

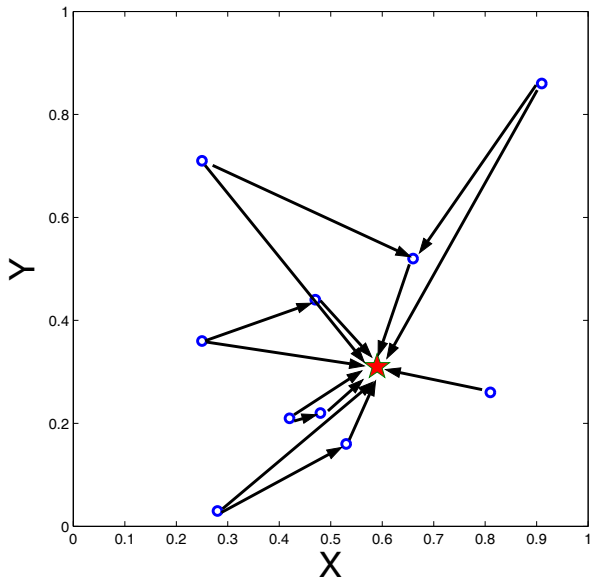


Figure 3.5: A schematic showing the routing solution for the small 10-node network.

The network setting (location, data rate, and initial energy for each node) for the 10-node network is given in Table 3.3. By applying Algorithm 3, we find the best FCP with cost vector $(1.05, 1.28, 1.05, 1.22, 1.16, 1.05, 1.05, 1.05, 1.41, 1.05)$ and the maximum achievable network lifetime $T^* = 357.49$, which is at least 95% of the optimum. By placing the base station at a point in the corresponding subarea, e.g., at point $(0.59, 0.31)$, the network lifetime is $T_\varepsilon = 359.17 > T^*$. Thus, the network lifetime is also at least 95% of the optimum. The routing solution is shown in Fig. 3.5, where a circle represents a sensor node and a star represents the location of the base station $(0.59, 0.31)$.

The network setting for a 20-node network (with location, data rate, and initial energy for each of the 20 sensor nodes) is given in Table 3.4. By applying Algorithm 3, we find the best FCP with a cost vector $(1.55, 1.05, 1.16, 1.05, 1.41, 1.22, 1.41, 1.22, 1.41, 1.28, 1.63, 1.05, 1.16, 1.98, 1.71, 1.16, 1.28, 1.80, 1.05, 1.05)$ and the maximum achievable network lifetime $T^* = 82.86$. Subsequently, we place the base station at a point in the corresponding subarea, say at point $(0.31, 0.79)$. The corresponding network lifetime is $T_\varepsilon = 82.91 > T^*$, which is also at least 95% of the optimum. The routing solution is shown in Fig. 3.6.

Table 3.4: Each node's location, data rate and initial energy for a 20-node network.

Location	Data rate	Initial energy	Location	Data rate	Initial energy
$(0.98, 0.49)$	0.4	180	$(0.44, 0.67)$	0.8	320
$(0.57, 0.52)$	0.1	340	$(0.13, 0.19)$	0.6	430
$(0.74, 0.73)$	0.1	350	$(0.24, 0.19)$	0.7	310
$(0.49, 0.38)$	0.9	410	$(0.63, 0.33)$	0.7	500
$(0.76, 0.63)$	0.6	270	$(0.92, 0.33)$	0.5	180
$(0.09, 0.84)$	0.7	60	$(0.65, 0.62)$	0.1	100
$(0.92, 0.05)$	0.1	310	$(1.00, 0.33)$	0.6	280
$(0.63, 1.00)$	0.2	210	$(0.11, 0.36)$	0.3	70
$(0.89, 0.12)$	0.7	420	$(0.52, 0.86)$	0.3	270
$(0.24, 0.91)$	0.9	160	$(0.40, 0.67)$	1.0	180

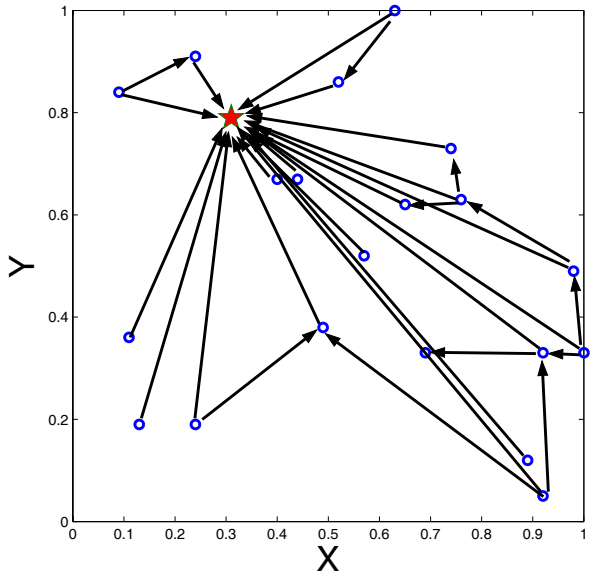
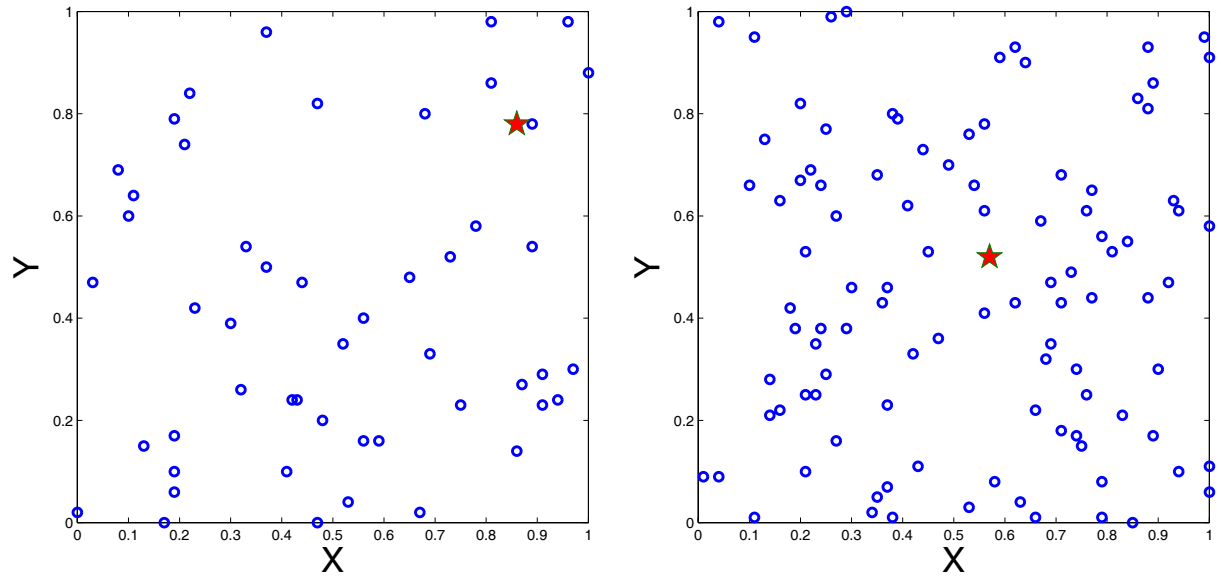


Figure 3.6: A schematic showing the routing solution for the 20-node network.



(a) A 50-node network.

(b) A 100-node network.

Figure 3.7: Network topology for 50-node and 100-node networks.

For the 50-node and 100-node networks, the positions of the nodes and location for the base station are shown in Fig. 3.7(a) and (b), respectively. The data rate and initial energy for each node are randomly generated between $[0.1, 1]$ and $[50, 500]$, respectively. For the 50-node network, we obtain a $(1-\varepsilon)$-optimal solution with $T_\varepsilon = 135.17$ when the base station is placed at $(0.51, 0.68)$; for the 100-node network, the $(1-\varepsilon)$-optimal solution is $T_\varepsilon = 61.73$ when the base station is placed at $(0.57, 0.52)$.

We now compare the complexity of our algorithm with the approximation algorithm proposed in [32] for the 10, 20, 50, and 100-node network considered in this section. Corresponding to each network topology, based on (3.9) and (3.10), we find that the complexity of the approximation algorithm in [32] is $3.7 \cdot 10^7$, $1.5 \cdot 10^7$, $5.2 \cdot 10^6$, and $3.2 \cdot 10^6$ times of the complexity of our approximation algorithm.

3.5 Related Work

Due to energy constraint, network lifetime for a wireless sensor network is limited. As a result, there is a flourish of research activities in this area in recent years. Many of these efforts (e.g., [11, 15, 55, 111]) studied lifetime problem under given network topology and without explicit consideration on the impact of node placement on network performance.

Among the body research on node placement, researchers have studied sensor node placement [29, 103, 107, 114], relay node placement [49, 109], and base station placement [14, 32, 73]. The main focus of sensor node placement has been on coverage in order to have either better geographical coverage of the area or better connectivity in the network. Relay node placement deals with how to place special auxiliary nodes within a sensor network so that network performance (e.g., connectivity, lifetime) can be improved. Related work in relay node placement (e.g., [49, 109]) have been limited to heuristic algorithm instead of providing performance guarantee.

Related work on base station placement include [14, 32, 73]. In [14], Bogdanov et al. studied how to place base station so that the network flow is proportionally maximized subject to link capacity. The authors show that although it is possible to find optimal solutions for special network topology (e.g., grid), the base station placement problem for an arbitrary network is NP-complete. The authors also pointed out that an approximation algorithm with any guarantee was not known at the time of their paper and subsequently proposed two heuristic algorithms. In [73], Pan et al. studied base station placement problem to maximize network lifetime. The optimal location is only determined for the simple case where only single-hop routing is allowed. The more difficult problem involving multi-hop routing was not addressed.

In [32], Efrat et al. proposed the first $(1 - \varepsilon)$-optimal approximation algorithm to the base station placement problem. However, since they constructed a finite search space of physical points, the computational complexity of their algorithm is higher than the one proposed in this chapter for most cases, as illustrated in Section 3.3.4.

3.6 Conclusions

In this chapter, we investigated the base station placement problem for a multi-hop sensor network. The main result is an approximation algorithm that can guarantee $(1-\varepsilon)$-optimal network lifetime performance for base station placement problem with any desired error bound $\varepsilon > 0$. The proposed $(1-\varepsilon)$-approximation algorithm was based on several novel techniques such as discretization of cost parameters (and distances), division of search space into finite number of subareas, and representation of subareas with fictitious cost points (with nice bounding properties on costs). We gave a proof that the proposed approximation algorithm is $(1-\varepsilon)$-optimal. The proposed approximation algorithm offers significant complexity reduction when compared to a state-of-the-art algorithm and represents the best known result to the described base station placement problem.

Chapter 4

Mobile Base Station for Wireless Sensor Networks

4.1 Introduction

The benefits of using mobile base station to prolong sensor network lifetime have been well recognized [62, 110]. Since base station is the sink node for data generated by all the sensor nodes in the network, this approach aims to alleviate the traffic aggregation burden from a *fix* set of sensor nodes near the base station to other sensor nodes in the network, and it is possible to extend the network lifetime significantly.

Indeed, although the practical feasibility of using a mobile base station is still considered far-fetched a few years ago, such capabilities are nowadays reality, thanks to recent breakthrough in unmanned autonomous vehicle (UAV) competition by DARPA's Grand Challenge program [24] and advances in customized robotics for sensors [91]. It has now become entirely plausible to consider that an unmanned vehicle be employed to serve as a mobile base station for sensor data collection.

Although the potential benefit of using mobile base station to prolong sensor network lifetime

is significant, the theoretical difficulty of this problem is enormous. There are two components that are tightly coupled in this problem. First, the location of the base station is now a function of time, i.e., at different time instances, we have a different physical network topology, with the sink node being at different positions. Second, the traffic (or flow) routing behavior may change with both time as well as the location of the base station. As a result, an optimization problem with the objective of maximizing the network lifetime needs to consider both base station location (time-dependent) and flow routing. Due to these difficulties, existing solutions to this problem remain heuristic at best (e.g., [62, 110]) and cannot provide *provably* optimal solution to network lifetime performance.

To fill in this theoretical gap, this chapter offers an in-depth study on the network lifetime performance limit when mobile base station is employed. We formulate an optimization problem with base station movement and flow routing as part of the constraints. As a first step, we show that as far as network lifetime objective is concerned, flow routing only needs to be dependent on the base station location, regardless of *when* the base station is present at this location. Further, the specific time instances for the base station to visit a location is not important, as long as the total sojourn time for the base station to be present at this location is fixed. With this finding, we show how to make a novel transformation from a time-dependent problem formulation to location (space)-dependent problem formulation.

As a second step, we show that when base station is only allowed to be present at a finite set of pre-determined points (called *constrained mobile base station* (C-MB) problem), we can find the optimal sojourn time for the base station to stay on each of these points (as well as the corresponding flow routing solution) such that the overall network lifetime (i.e., sum of the sojourn times) is maximized via a single linear program (LP).

Building upon these results, the main result in this chapter (Section 4.5) shows that for the *un-constrained mobile base station* (U-MB) problem, i.e., the base station can be present at any point in the two dimensional plane, we can develop a *provably* $(1-\varepsilon)$-optimal algorithm that provides a solution with network lifetime guaranteed to be at least $(1-\varepsilon)$ of the maximum network lifetime

(albeit it is unknown), where ε can be made arbitrarily small depending on required precision. The main idea in this approximation algorithm is to exploit a clever way of dividing the search space into subareas, with each of its link cost having some nice properties that are related to $(1-\varepsilon)$-optimal. A novel idea in the design of $(1-\varepsilon)$-optimal algorithm is to represent each subarea with so-called "fictitious cost point," which is an N-tuple cost vector with each component representing an upper bound of cost to the respective node. As a result, we can apply the LP approach developed for the C-MB problem on the fictitious cost points and develop provably $(1-\varepsilon)$-optimal solution.

The remainder of this chapter is organized as follows. In Section 4.2, we describe the network model and introduce the mobile base station problem. In Section 4.3, we present a novel transformation that enables to transform a time-dependent problem to a space-dependent problem. In Section 4.4, we develop optimal solution for the constrained mobile base station (C-MB) problem. Section 4.5 presents the main result of this chapter, which is an algorithm for the unconstrained mobile base station (U-MB) problem with provably $(1-\varepsilon)$-optimal network lifetime. Section 4.6 reviews related work and Section 4.7 concludes this chapter.

4.2 Network Model and Problem Formulation

4.2.1 Network Model

We consider a set $\mathcal{N}$ of nodes deployed over a two-dimensional area, with the location of each sensor node $i \in \mathcal{N}$ being at a point (x_i, y_i). We assume each node generates data at a rate of r_i. There is a base station B for the sensor network and it serves as the sink node for all data. Data generated by each sensor node can be transmitted via single or multi-hop toward the base station.

Each node $i \in \mathcal{N}$ is initially provisioned with an amount of energy e_i. The base station is not constrained with energy. In this study, network lifetime is defined as the first time instance when any of the sensor nodes runs out of energy. Power consumption model has been discussed in Section 2.2.2. From (2.1), (2.2), and (2.3), it is easy to understand that the location of the base

Table 4.1: Notation in Chapter 4.

$f_{ij}(p)$	Flow rate from node i to node j when base station B is at point p
$r_{ij}(t)$	Flow rate from node i to node j at time t
$T^*_{\text{C-MB}}$	The maximum network lifetime achieved by $\psi^*_{\text{C-MB}}$
$T_{\text{U-MB}}$	$(1-\varepsilon)$-optimal network lifetime achieved by $\psi_{\text{U-MB}}$
$W(p)$	Sojourn time for the base station to be present at point p
$W(A_m)$	Sojourn time for the base station to be present in subarea A_m
$(x,y)(t)$	Location of base station B at time t
(x_i, y_i)	Location of sensor node i
$\beta_{iB}(t)$	Cost for transmitting data from sensor i to base station B at time t
$\beta_{iB}(p)$	Cost for transmitting data from sensor node i to base station B when B is at point p
$\psi^*_{\text{C-MB}}$	An optimal solution to the C-MB problem
$\psi_{\text{U-MB}}$	$(1-\varepsilon)$-optimal solution to the U-MB problem

station and the corresponding flow routing among the nodes will determine energy consumption behavior at each node and thus the network lifetime. Table 4.1 lists all notation that are used only in this chapter.

4.2.2 Problem Description

The focus of this chapter is to investigate how to optimally move a mobile base station to collect real time data in a sensor network so that the network lifetime can be maximized. Note that the network lifetime problem has attracted great interest even for static (fixed) base station problem (e.g., [15, 17, 18, 87]).

As a first step, we consider the case when the base station is only allowed to be present at a set of pre-determined M positions, denoted as $p_1, p_2, \cdots, p_M$. We call this problem as constrained mobile base station (C-MB) problem. Results on C-MB problem will help us devise solution to the general problem where the base station is allowed to roam anywhere on the two-dimensional plane. We term the latter problem unconstrained mobile base station (U-MB) problem.

Denote $(x, y)(t)$ the position of base station B at time t and T the network lifetime (which is an optimization variable). Then a feasible flow routing solution realizing this network lifetime T must satisfy both flow balance and energy constraints at each sensor node. These constraints can be formally stated as follows. Denote $r_{ij}(t)$ and $r_{iB}(t)$ the data rates from node i to node j and base station B at time t, respectively. Then the flow balance for each sensor node $i \in \mathcal{N}$ at any time $t \in [0, T]$ is

$$\sum_{k \in \mathcal{N}}^{k \neq i} r_{ki}(t) + r_i = \sum_{j \in \mathcal{N}}^{j \neq i} r_{ij}(t) + r_{iB}(t) \ ,$$

i.e., the sum of total incoming flow rates plus self-generated data rate is equal to the sum of total outgoing flow rates at time $t \in [0, T]$. The energy constraint for each sensor node $i \in \mathcal{N}$ is

$$\int_0^T \left[\sum_{k \in \mathcal{N}}^{k \neq i} \beta_{\text{rec}} r_{ki}(t) + \sum_{j \in \mathcal{N}}^{j \neq i} \beta_{ij} r_{ij}(t) + \beta_{iB}(t) r_{iB}(t) \right] dt \leq e_i \ ,$$

i.e., total consumed energy due to reception and transmission over time T cannot exceed its initial energy e_i. By (2.2), we have

$$\beta_{iB}(t) = \beta_1 + \beta_2 \left[\sqrt{(x(t) - x_i)^2 + (y(t) - y_i)^2} \right]^{\alpha} \ .$$

Denote $\mathcal{A}$ the search space for the base station, which can be narrowed down to the smallest enclosing disk for all nodes in the network (see Lemma 3). The general U-MB problem can be formulated as follows.

$$\begin{array}{lll}
\textbf{Max} & T & \\
\textbf{s.t.} & \sum_{k \in \mathcal{N}}^{k \neq i} r_{ki}(t) + r_i = \sum_{j \in \mathcal{N}}^{j \neq i} r_{ij}(t) + r_{iB}(t) & (i \in \mathcal{N}, 0 \leq t \leq T) \\
& \int_0^T \left[\sum_{k \in \mathcal{N}}^{k \neq i} \beta_{\text{rec}} r_{ki}(t) + \sum_{j \in \mathcal{N}}^{j \neq i} \beta_{ij} r_{ij}(t) + \beta_{iB}(t) r_{iB}(t) \right] dt \leq e_i & (i \in \mathcal{N}) \\
& \beta_{iB}(t) = \beta_1 + \beta_2 \left[\sqrt{(x(t) - x_i)^2 + (y(t) - y_i)^2} \right]^{\alpha} & (i \in \mathcal{N}, 0 \leq t \leq T) \\
& (x, y)(t) \in \mathcal{A} & (0 \leq t \leq T) \\
& T, r_{ij}(t), r_{iB}(t) \geq 0 & (i, j \in \mathcal{N}, i \neq j, 0 \leq t \leq T)
\end{array}$$

In the above formulation, the base station location (i.e., $(x, y)(t)$ for $0 \leq t \leq T$) and the corresponding flow routing (i.e., $r_{ij}(t)$ and $r_{iB}(t)$ for $0 \leq t \leq T$) form a joint optimization space

for the objective T. This formulation is in the form of *non-polynomial program*. Since even a simpler non-linear programming problem is NP-hard in general [40], we conclude that the above formulation is also NP-hard in general.

4.3 From Time Domain to Space Domain

The difficulty of the formulated problem in last section resides in that base station location $(x, y)(t)$ and flow routing $r_{ij}(t)$ and $r_{iB}(t)$ are all functions of time. In this section, we show that as far as network lifetime performance is concerned, such dependency on time can be relaxed. Specifically, we will show (Theorem 4) that the flow routing only needs to be dependent on the location of the base station and independent of when the base station is present at this location. Further, as long as the total sojourn time for this location is the same, "when" the base station visits this location is not important.

To start with, we define the indicator function $1^+\{\text{event}\}$ as 1 if event is true and 0 otherwise. Denote $W(p)$ the cumulative time periods for the base station B to be present at location p under a solution φ, i.e.,

$$W(p) = \int_0^T 1^+\{(x, y)(t) = p\}dt \ . \tag{4.1}$$

We have the following theorem.

Theorem 4. *Denote T^* the maximum network lifetime achieved by an optimal solution φ^* with a base station moving path $(x, y)^*(t)$ and a flow routing $r^*_{ij}(t)$ and $r^*_{iB}(t)$. There exists an equivalent solution $\bar{\varphi}^*$ with $W^*(p)$ and time-independent flow rates $f^*_{ij}(p)$ and $f^*_{iB}(p)$ that yields the same maximum network lifetime. Under $f^*_{ij}(p)$ and $f^*_{iB}(p)$, as long as $W^*(p)$ remain the same, the network lifetime T^* will remain unchanged regardless of the specific time instances when the base station visits each point p.*

The proof of Theorem 4 is based on the Lemmas 5 and 6.

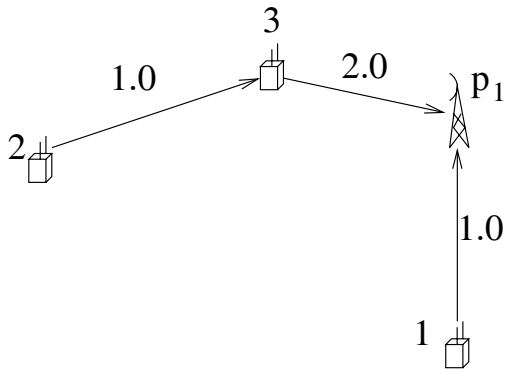
Lemma 5. *Given a feasible solution φ with a specific base station moving path $(x, y)(t)$, a time-varying flow routing $r_{ij}(t)$ and $r_{iB}(t)$, and a corresponding network lifetime T, we can find a solution $\bar{\varphi}$ with location-dependent flow rates $f_{ij}(p)$ and $f_{iB}(p)$ for each point p in path $(x, y)(t)$ that yields the same network lifetime T by having the base station following the same moving path and setting the flow routing on any point p in this path as*

$$f_{ij}(p) = \frac{\int_0^T r_{ij}(t) \cdot 1^+\{(x, y)(t) = p\}dt}{W(p)},$$
$$f_{iB}(p) = \frac{\int_0^T r_{iB}(t) \cdot 1^+\{(x, y)(t) = p\}dt}{W(p)}.$$

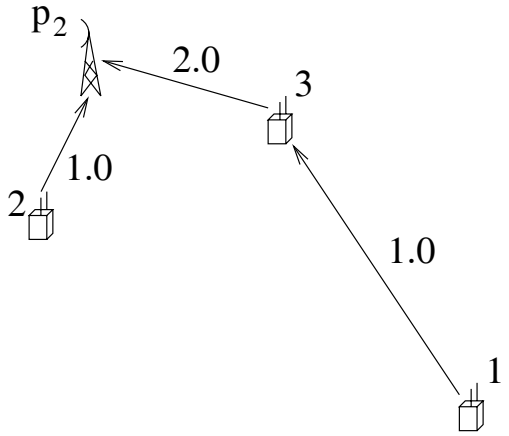
In essence, Lemma 5 focus on flow routing component and shifts the flow routing's dependency from time domain to space domain. We use the following simple example to illustrate this idea.

Example 1. *For illustration purpose, we consider a simple 3-node sensor network shown in Fig. 4.1. Each node generates data at a bit rate of 1. We assume all units are normalized appropriately. The network lifetime is 130 under a solution φ, where the base station stays at p_1 during periods $[0, 50]$ and $[100, 130]$, and stays at p_2 during $[50, 100]$. The flow routing for each period is given in Fig. 4.1(a)–(d), with bit rate of each flow marked on the respective flow arrow. Note that the flow routing for periods $[0, 50]$ and $[100, 130]$ are different, despite that the base station is located at p_1 during both periods. Further, during $[50, 100]$, the flow routing changes between periods $[50, 90]$ and $[90, 100]$ while the base station is at p_2.*

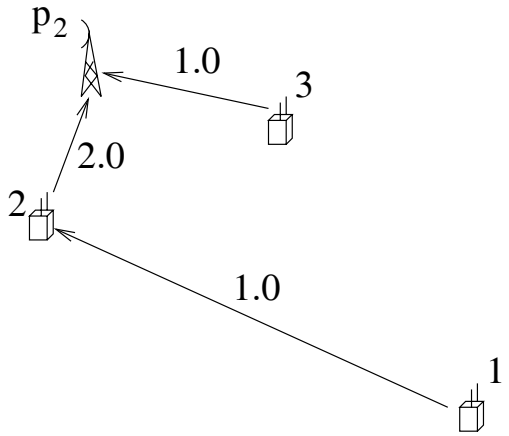
Under $\bar{\varphi}$, we show that as far as achieving the same network lifetime, flow routing only needs to be dependent on the base station's location. For example, under φ, the total time that the base station stays at p_1 is 80 $(= 50 + 30)$. During the time periods when the base station is at p_1, $f_{3B}(t) = 2.0$ when $0 \le t \le 50$ and $f_{3B}(t) = 1.0$ when $100 \le t \le 130$. Thus, the average rate on link $3 \to B$ when the base station stays at p_1 is $(2.0 * 50 + 1.0 * 30)/80 = 1.625$. We can set $f_{3B}(p_1) = 1.625$ under $\bar{\varphi}$. Similarly, we obtain the average rate $f_{23}(p_1) = 0.625$ and $f_{2B}(p_1) = 0.375$, respectively. This lead to a new flow routing shown in Fig. 4.2(a) for periods $[0, 50]$ and $[100, 130]$. Following the same token, we can convert the flow routing in Fig. 4.1(b) and (c) to that in Fig. 4.2(b), i.e., $f_{12}(p_2) = 0.2$, $f_{13}(p_2) = 0.8$, $f_{2B}(p_2) = 1.2$, and $f_{3B}(p_2) = 1.8$.



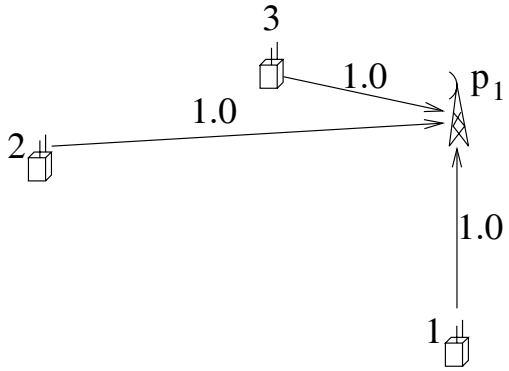
(a) Flow routing during $[0, 50]$ when the base station is at p_1.



(b) Flow routing during $[50, 90]$ when the base station is at p_2.

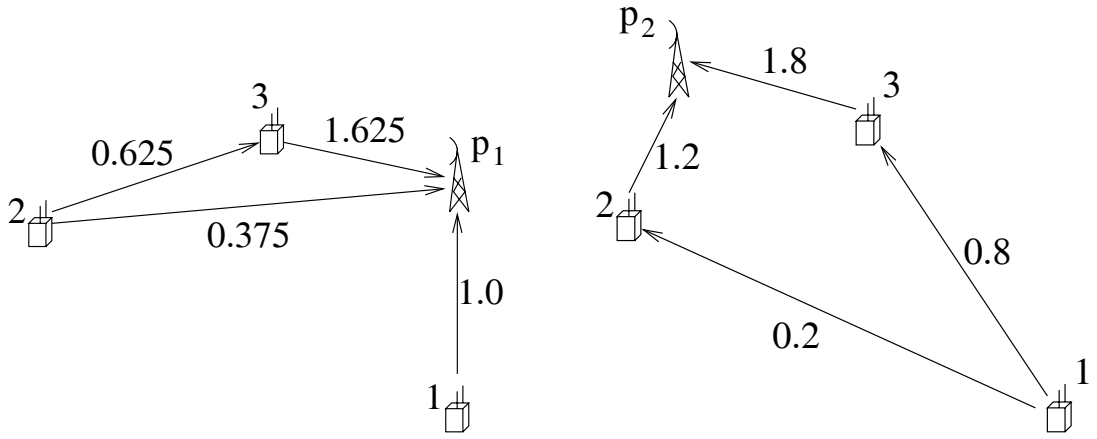


(c) Flow routing during $[90, 100]$ when the base station is at p_2.



(d) Flow routing during $[100, 130]$ when the base station is at p_1.

Figure 4.1: A simple example illustrating time-varying flow routing under φ in Lemma 5.



(a) Flow routing during $[0, 50]$ and $[100, 130]$ when the base station is at p_1.

(b) Flow routing during $[50, 100]$ when the base station is at p_2.

Figure 4.2: A space-dependent flow routing under $\bar{\varphi}$ in Lemma 5.

It is easy to verify that flow balance and total data volume transmitted on each link under $\bar{\varphi}$ remain the same as those under φ. Based on (2.1) and (2.3), it is easy to verify that the network lifetime under $\bar{\varphi}$ also remains the same as that under φ (i.e., 130). As a result, we have obtained a solution $\bar{\varphi}$ with flow routing only dependent on base station's location.

The following proof of Lemma 5 follows the same spirit in the above example.

Proof. Denote $\mathcal{P}$ the base station moving path $(x, y)(t)$ for $0 \leq t \leq T$ in solution φ. We let base station follow the same path in solution $\bar{\varphi}$. For the data routing in $\bar{\varphi}$, we can define $f_{ij}(p)$ and $f_{iB}(p)$ for each point $p \in \mathcal{P}$ as follows.

$$f_{ij}(p) = \frac{\int_0^T r_{ij}(t) \cdot 1^+\{(x, y)(t) = p\}dt}{W(p)} \tag{4.2}$$

$$f_{iB}(p) = \frac{\int_0^T r_{iB}(t) \cdot 1^+\{(x, y)(t) = p\}dt}{W(p)} \tag{4.3}$$

To show the data routing scheme with $f_{ij}(p)$ and $f_{iB}(p)$ is feasible and $\bar{\varphi}$ has the same network lifetime T, we need to prove that flow balance holds at any point $p \in \mathcal{P}$ and energy consumption is the same as that in solution φ at time T.

For flow balance when base station is at any point $p \in \mathcal{P}$, we have

$$\begin{aligned}
\sum_{k\in\mathcal{N}}^{k\neq i} f_{ki}(p) + r_i &= \sum_{k\in\mathcal{N}}^{k\neq i} \frac{\int_0^T r_{ki}(t)\cdot 1^+\{(x,y)(t)=p\}dt}{W(p)} + \frac{\int_0^T r_i \cdot 1^+\{(x,y)(t)=p\}dt}{W(p)} \\
&= \frac{\int_0^T \left[\sum_{k\in\mathcal{N}}^{k\neq i} r_{ki}(t) + r_i\right]\cdot 1^+\{(x,y)(t)=p\}dt}{W(p)} \\
&= \frac{\int_0^T \left[\sum_{j\in\mathcal{N}}^{j\neq i} r_{ij}(t) + r_{iB}(t)\right]\cdot 1^+\{(x,y)(t)=p\}dt}{W(p)} \\
&= \sum_{j\in\mathcal{N}}^{j\neq i} \frac{\int_0^T r_{ij}(t)\cdot 1^+\{(x,y)(t)=p\}dt}{W(p)} + \frac{\int_0^T r_{iB}(t)\cdot 1^+\{(x,y)(t)=p\}dt}{W(p)} \\
&= \sum_{j\in\mathcal{N}}^{j\neq i} f_{ij}(p) + f_{iB}(p)
\end{aligned}$$

The first equality holds by (4.2). The third equality holds by the flow balance in solution φ. The last equality holds by (4.2) and (4.3).

For energy consumption at time T, we first have

$$\begin{aligned}
&\int_{\mathcal{P}}\int_0^T \beta_{iB}(p) f_{iB}(p)\cdot 1^+\{(x,y)(t)=p\}dtdp \\
=\ &\int_{\mathcal{P}} \beta_{iB}(p) f_{iB}(p)\cdot \int_0^T 1^+\{(x,y)(t)=p\}dtdp \\
=\ &\int_{\mathcal{P}} \beta_{iB}(p) \frac{\int_0^T r_{iB}(\tau)\cdot 1^+\{(x,y)(\tau)=p\}d\tau}{W(p)} \cdot \int_0^T 1^+\{(x,y)(t)=p\}dtdp \\
=\ &\int_{\mathcal{P}} \beta_{iB}(p) \int_0^T r_{iB}(\tau)\cdot 1^+\{(x,y)(\tau)=p\}d\tau dp \\
=\ &\int_{\mathcal{P}}\int_0^T \beta_{iB}(p) r_{iB}(t)\cdot 1^+\{(x,y)(t)=p\}dtdp \\
=\ &\int_{\mathcal{P}}\int_0^T \beta_{iB}(t) r_{iB}(t)\cdot 1^+\{(x,y)(t)=p\}dtdp \\
=\ &\int_0^T \beta_{iB}(t) r_{iB}(t) dt
\end{aligned} \tag{4.4}$$

for any node $i \in \mathcal{N}$. The second equality holds by (4.3). The third equality holds by (4.1). Similarly, we have

$$\int_{\mathcal{P}}\int_0^T f_{ij}(p)\cdot 1^+\{(x,y)(t)=p\}dtdp = \int_0^T r_{ij}(t)dt \tag{4.5}$$

for any $i, j \in \mathcal{N}, i \neq j$. Thus, we have

$$\begin{aligned}
& \int_{\mathcal{P}} \int_0^T \left[\sum_{k \in \mathcal{N}}^{k \neq i} \beta_{\text{rec}} f_{ki}(p) + \sum_{j \in \mathcal{N}}^{j \neq i} \beta_{ij} f_{ij}(p) + \beta_{iB}(p) \cdot f_{iB}(p) \right] \cdot 1^+\{(x,y)(t) = p\} dt dp \\
= & \sum_{k \in \mathcal{N}}^{k \neq i} \beta_{\text{rec}} \int_{\mathcal{P}} \int_0^T f_{ki}(p) 1^+\{(x,y)(t) = p\} dt dp + \sum_{j \in \mathcal{N}}^{j \neq i} \beta_{ij} \int_{\mathcal{P}} \int_0^T f_{ij}(p) 1^+\{(x,y)(t) = p\} dt dp \\
& + \int_{\mathcal{P}} \int_0^T \beta_{iB}(p) f_{iB}(p) 1^+\{(x,y)(t) = p\} dt dp \\
= & \sum_{k \in \mathcal{N}}^{k \neq i} \int_0^T \beta_{\text{rec}} r_{ki}(t) dt + \sum_{j \in \mathcal{N}}^{j \neq i} \int_0^T \beta_{ij} r_{ij}(t) dt + \int_0^T \beta_{iB}(t) f_{iB}(t) dt \leq e_i
\end{aligned}$$

The second equality holds by (4.4) and (4.5). The last inequality holds due to the energy constraint under φ. Thus, the data routing with $f_{ij}(p)$ and $f_{iB}(p)$ is feasible and has the same network lifetime T. This completes the proof. □

The following lemma further extends Lemma 5 and says that the specific time instances for the base station to visit a particular point p is not important. As long as $W(p)$ remain the same, the network lifetime remains unchanged. For example, in the transformed solution $\bar{\varphi}$ in Fig. 4.2, as long as the base station stays at point p_1 for a time period of 80 $(= 50 + 30)$ and point p_2 for a time period of 50, the exact time instances when the base station is present at this location is not important in terms of achieving the same network lifetime.

Lemma 6. *Given a feasible solution φ with a base station moving path $(x, y)(t)$, a flow routing $r_{ij}(t)$ and $r_{iB}(t)$, and a network lifetime T, we can compute $W(p)$ by (4.1) and define base station location-dependent flow routing $f_{ij}(p)$ and $f_{iB}(p)$ as in Lemma 5. Under $f_{ij}(p)$ and $f_{iB}(p)$, as long as $W(p)$ remain the same, the network lifetime T will remain unchanged regardless of the ordering and frequency of the base station's presence at each point p.*

Lemma 6 can be easily proved by analyzing the energy consumption behavior at each node over time T. We omit its proof here. Combining Lemmas 5 and 6 and considering the special case that φ is optimal, we have Theorem 4.

Based on Theorem 4, we conclude that as far as network lifetime is concerned, it is sufficient for us to obtain $W(p)$, $f_{ij}(p)$, and $f_{iB}(p)$ values when the base station is at each point p. The specific realization for $(x, y)(t)$ is not important and such realizations are certainly not unique. As a result, we can transform the optimization space from time-dependent functions $x(t, y)(t), r_{ij}(t)$, and $r_{iB}(t)$ to space-dependent functions $W(p), f_{ij}(p)$, and $f_{iB}(p)$. Subsequently, U-MB problem formulation given in Section 4.2.2 can be reformulated as follows.

$$
\begin{aligned}
\textbf{Max} \quad & T \\
\textbf{s.t.} \quad & \int_{\mathcal{A}} W(p)dp = T \\
& \textstyle\sum_{k\in\mathcal{N}}^{k\neq i} f_{ki}(p) + r_i = \sum_{j\in\mathcal{N}}^{j\neq i} f_{ij}(p) + f_{iB}(p) && (i \in \mathcal{N}, p \in \mathcal{A}) \\
& \textstyle\int_{\mathcal{A}} \left[\sum_{k\in\mathcal{N}}^{k\neq i}\beta_{\text{rec}} f_{ki}(p) + \sum_{j\in\mathcal{N}}^{j\neq i}\beta_{ij} f_{ij}(p) + \beta_{iB}(p) f_{iB}(p)\right] W(p)dp \leq e_i && (i \in \mathcal{N}) \qquad (4.6) \\
& T, W(p), f_{ij}(p), f_{iB}(p) \geq 0 && (i, j \in \mathcal{N}, i \neq j, p \in \mathcal{A})
\end{aligned}
$$

Note that integration $\int_0^T (\cdot)dt$ (with respect to time) in the original problem formulation (Section 4.2.2) has been changed to $\int_{\mathcal{A}}(\cdot)dp$ (with respect to space) in the new formulation. This novel transformation will enable us to develop provably approximation algorithm in the space domain, which we will elaborate in Section 4.5.

4.4 Optimal Solution to the C-MB Problem

In this section, we show that C-MB problem can be formulated as an LP problem, which can be solved in polynomial-time [6]. This result will be useful when we solve the U-MB problem in Section 4.5.

Recall that in the C-MB problem, the location of base station is limited to a given set of predetermined locations p_m, $1 \leq m \leq M$. Based on the results in Section 4.3, we need to find optimal sojourn time $W(p_m)$ and the corresponding flow routing $f_{ij}(p_m)$ and $f_{iB}(p_m)$ when the base station is at each p_m to maximize the network lifetime.[1]

[1] When $W(p_m) = 0$ for some p_m, it means that base station does not visit p_m in this solution.

When the base station is at point p_m, $1 \leq m \leq M$, the flow balance for node $i \in \mathcal{N}$ is

$$\sum_{k \in \mathcal{N}}^{k \neq i} f_{ki}(p_m) + r_i = \sum_{j \in \mathcal{N}}^{j \neq i} f_{ij}(p_m) + f_{iB}(p_m) \ . \tag{4.7}$$

The energy constraint for node $i \in \mathcal{N}$, at the end of network lifetime T is

$$\sum_{1 \leq m \leq M} \left[\sum_{k \in \mathcal{N}}^{k \neq i} \beta_{\text{rec}} f_{ki}(p_m) + \sum_{j \in \mathcal{N}}^{j \neq i} \beta_{ij} f_{ij}(p_m) + \beta_{iB}(p_m) f_{iB}(p_m) \right] W(p_m) \leq e_i \ . \tag{4.8}$$

Note that for each i and p_m, $\beta_{iB}(p_m)$ is a constant.

We can formulate C-MB problem as an LP by letting $V_{ij}(p_m) = f_{ij}(p_m) \cdot W(p_m)$ and $V_{iB}(p_m) = f_{iB}(p_m) \cdot W(p_m)$, where $V_{ij}(p_m)$ (or $V_{iB}(p_m)$) can be interpreted as the total data volume from sensor node i to sensor node j (or base station B) when the base station is at p_m. We have

$$\begin{aligned}
&\underline{\textbf{LP(C-MB)}} \quad \textbf{Max} && T \\
&\qquad\qquad \textbf{s.t.} && \sum_{1 \leq m \leq M} W(p_m) - T = 0 \\
&\sum_{k \in \mathcal{N}}^{k \neq i} V_{ki}(p_m) + r_i \cdot W(p_m) - \sum_{j \in \mathcal{N}}^{j \neq i} V_{ij}(p_m) - V_{iB}(p_m) = 0 && (i \in \mathcal{N}, 1 \leq m \leq M) \quad (4.9) \\
&\sum_{1 \leq m \leq M} \left[\sum_{k \in \mathcal{N}}^{k \neq i} \beta_{\text{rec}} V_{ki}(p_m) + \sum_{j \in \mathcal{N}}^{j \neq i} \beta_{ij} V_{ij}(p_m) + \beta_{iB}(p_m) V_{iB}(p_m) \right] \leq e_i && (i \in \mathcal{N}) \quad (4.10) \\
&\qquad\qquad T, W(p_m), V_{ij}(p_m), V_{iB}(p_m) \geq 0 && (i, j \in \mathcal{N}, i \neq j, 1 \leq m \leq M) \ ,
\end{aligned}$$

where (4.9) and (4.10) follow from (4.7) and (4.8), respectively. Once we solve the above LP, we can obtain $f_{ij}(p_m)$ and $f_{iB}(p_m)$ by $f_{ij}(p_m) = \frac{V_{ij}(p_m)}{W(p_m)}$ and $f_{iB}(p_m) = \frac{V_{iB}(p_m)}{W(p_m)}$. We summarize the result in this section with the following proposition.

Proposition 1. *C-MB problem can be solved via a single LP in polynomial-time.*

The solution to the above LP problem yields the sojourn times for the base station at each location p_m, $1 \leq m \leq M$, and the optimal flow routing when the base station is present at p_m. So far we assume that after base station B completes its stay at p_i, it can move to p_j, $j \neq i$, in zero time. In practice, such travel will take some time. We assume that such travel time is negligible

comparing with the time scale of network lifetime (typically several months). It can be shown that if buffering is available at sensor node, the maximum network lifetime can still be achieved. In this case, each node needs to slightly delay its transmission until the base station arrives at its new location.

4.5 An Approximation Algorithm to the U-MB Problem

Building upon the results in the previous sections, we are now ready to present the main result of this chapter.

4.5.1 Subareas and Fictitious Cost Points

Under the U-MB problem, the base station can move to any point in the two-dimensional plane. Clearly, in order to maximize network lifetime, we would expect that the location for the base station can be narrowed down to the *smallest enclosing disk* (SED) [106]. The proof is similar to that for Lemma 3.

Although we have narrowed down the search space for base station B from an infinite two-dimensional plane to SED $\mathcal{A}$, there still exist infinite number of points in $\mathcal{A}$. To obtain a finite search space, we follow the same approach in Section 3.3.2 to divide SED into subareas and to represent each subarea by a fictitious cost point (FCP).

Suppose we have the set of M non-uniform subareas that are represented by M FCPs. We now readily apply the LP approach discussed in Section 4.4 to formulate an optimization problem on these M FCPs. In next section, we will show that an optimal solution to the C-MB problem on FCPs can be used to construct a $(1-\varepsilon)$-optimal solution to the U-MB problem.

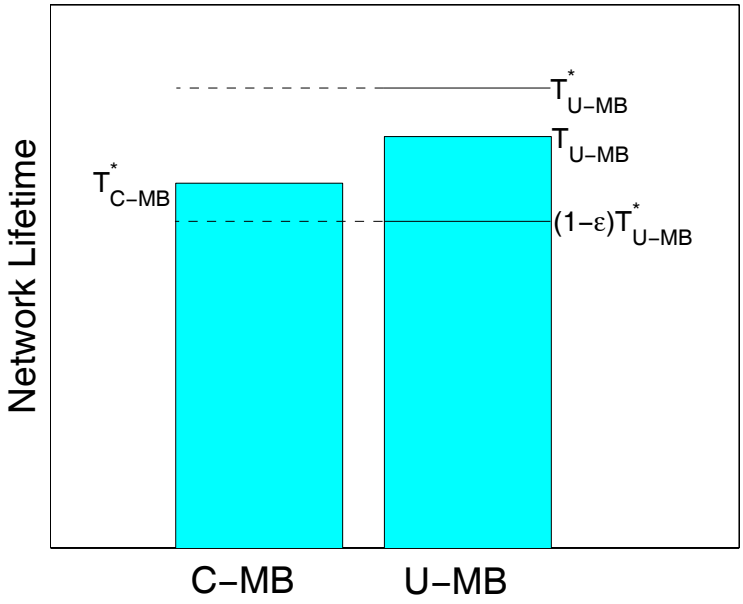


Figure 4.3: Comparison of network lifetimes under different solutions used to construct $(1-\varepsilon)$-optimal solution.

4.5.2 Design of Approximation Algorithm

Denote $\psi^*_{\text{U-MB}}$ an optimal solution to the U-MB problem and $T^*_{\text{U-MB}}$ the corresponding maximum network lifetime, both of which are unknown. Our objective is to find a solution to the U-MB problem that has provably $(1-\varepsilon)$-optimal network lifetime. Denote $\psi^*_{\text{C-MB}}$ an optimal solution to the C-MB problem obtained by applying an LP on the M FCPs and $T^*_{\text{C-MB}}$ the corresponding network lifetime.

Our roadmap for theoretical results is as follows. In Theorem 5, we will prove that $T^*_{\text{C-MB}} \geq (1-\varepsilon)T^*_{\text{U-MB}}$ (see Fig. 4.3). Since the optimal solution $\psi^*_{\text{C-MB}}$ corresponding to $T^*_{\text{C-MB}}$ are based on the M FCPs instead of physical points, in Theorem 6 we will further show how to construct a solution $\psi_{\text{U-MB}}$ to U-MB problem based on $\psi^*_{\text{C-MB}}$ and prove that the corresponding network lifetime is $(1-\varepsilon)$-optimal, i.e., $T_{\text{U-MB}} \geq (1-\varepsilon)T^*_{\text{U-MB}}$ (see Fig. 4.3).

Theorem 5. *For $T^*_{C\text{-}MB}$ and $T^*_{U\text{-}MB}$ as defined, we have $T^*_{C\text{-}MB} \geq (1-\varepsilon)\cdot T^*_{U\text{-}MB}$.*

To prove that Theorem 5 is true, we need the following lemma, where subareas $\mathcal{A}_m$ and FCPs p_m, $1 \leq m \leq M$ are defined the same as in Section 3.3.2 for a given $\varepsilon > 0$. Denote $W(\mathcal{A}_m)$ the cumulative time period when the base station B is present within subarea $\mathcal{A}_m$ for the U-MB problem. We have

$$W(\mathcal{A}_m) = \int_{\mathcal{A}_m} W(p)dp \,. \tag{4.11}$$

Lemma 7. *Suppose we have a given solution $\varphi_{\text{U-MB}}$ to the U-MB problem with $W(p)$, $f_{ij}(p)$, $f_{iB}(p)$, and a network lifetime $T_{\text{U-MB}}$. For a given $\varepsilon > 0$, we can always construct a solution $\varphi_{\text{C-MB}}$ to C-MB problem on these FCPs such that network lifetime $T_{\text{C-MB}} \geq (1-\varepsilon) \cdot T_{\text{U-MB}}$ by having the base station spend $W(p_m)$ amount of time on FCP p_m, where*

$$W(p_m) = (1-\varepsilon) \cdot W(\mathcal{A}_m) \tag{4.12}$$

and setting the flow routing on p_m as

$$f_{ij}(p_m) = \frac{\int_{\mathcal{A}_m} f_{ij}(p)W(p)dp}{W(\mathcal{A}_m)} \,, \tag{4.13}$$

$$f_{iB}(p_m) = \frac{\int_{\mathcal{A}_m} f_{iB}(p)W(p)dp}{W(\mathcal{A}_m)} \,. \tag{4.14}$$

Proof. We will show solution $\varphi_{\text{C-MB}}$ is feasible, i.e., flow balance holds at each point, and the network lifetime of $\varphi_{\text{C-MB}}$ is at least $(1-\varepsilon) \cdot T_{\text{U-MB}}$.

For flow balance when the base station location is p_m, we have

$$\begin{aligned}
\sum_{k\in\mathcal{N}}^{k\neq i} f_{ki}(p_m) + r_i &= \sum_{k\in\mathcal{N}}^{k\neq i} \frac{\int_{\mathcal{A}_m} f_{ki}(p)W(p)dp}{W(\mathcal{A}_m)} + \frac{\int_{\mathcal{A}_m} r_i W(p)dp}{W(\mathcal{A}_m)} \\
&= \frac{\int_{\mathcal{A}_m} \left[\sum_{k\in\mathcal{N}}^{k\neq i} f_{ki}(p) + r_i\right] W(p)dp}{W(\mathcal{A}_m)} \\
&= \frac{\int_{\mathcal{A}_m} \left[\sum_{j\in\mathcal{N}}^{j\neq i} f_{ij}(p) + f_{iB}(p)\right] W(p)dp}{W(\mathcal{A}_m)} \\
&= \sum_{j\in\mathcal{N}}^{j\neq i} \frac{\int_{\mathcal{A}_m} f_{ij}(p)dp}{W(\mathcal{A}_m)} + \frac{\int_{\mathcal{A}_m} f_{iB}(p)dp}{W(\mathcal{A}_m)} \\
&= \sum_{j\in\mathcal{N}}^{j\neq i} f_{ij}(p_m) + f_{iB}(p_m) \,.
\end{aligned}$$

The first equality holds by (4.13). The third equality holds by the flow balance in solution $\varphi_{\text{U-MB}}$. The last equality holds by (4.13) and (4.14). Thus, solution $\varphi_{\text{C-MB}}$ is feasible.

For the total consumed energy on node i at time $\sum_{1\leq m\leq M} W(p_m) = \sum_{1\leq m\leq M}(1-\varepsilon)\cdot W(\mathcal{A}_m) = (1-\varepsilon)\cdot T_{\text{U-MB}}$, we first have

$$\begin{aligned}
\sum_{1\leq m\leq M} \beta_{iB}(p_m) f_{iB}(p_m) W(p_m) &= \sum_{1\leq m\leq M} \beta_{iB}(p_m) \frac{\int_{\mathcal{A}_m} f_{iB}(p)dp}{W(\mathcal{A}_m)} W(p_m) \\
&= (1-\varepsilon) \sum_{1\leq m\leq M} \beta_{iB}(p_m) \int_{\mathcal{A}_m} f_{iB}(p)dp \\
&\leq (1-\varepsilon) \sum_{1\leq m\leq M} (1+\varepsilon) \int_{\mathcal{A}_m} \beta_{iB}(p) f_{iB}(p)dp \\
&< \sum_{1\leq m\leq M} \int_{\mathcal{A}_m} \beta_{iB}(p) f_{iB}(p)dp \\
&= \int_{\mathcal{A}} \beta_{iB}(p) f_{iB}(p)dp \qquad (4.15)
\end{aligned}$$

for any node $i \in \mathcal{N}$. The first equality holds by (4.14). The second equality holds by (4.12). The third inequality holds by $\beta_{iB}(p_m) \leq (1+\varepsilon)\beta_{iB}(p)$ in Property 1. Similarly, we have

$$\sum_{1\leq m\leq M} f_{ij}(p_m) W(p_m) < \int_{\mathcal{A}} f_{ij}(p)dp \qquad (4.16)$$

for $i, j \in \mathcal{N}$ and $i \neq j$. Thus, we have

$$\begin{aligned}
&\sum_{1\leq m\leq M} \left[\sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki}(p_m) W(p_m) + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij}(p_m) W(p_m) + \beta_{iB}(p_m) f_{iB}(p_m) W(p_m) \right] \\
= &\sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} \sum_{1\leq m\leq M} f_{ki}(p_m) W(p_m) + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} \sum_{1\leq m\leq M} f_{ij}(p_m) W(p_m) \\
&+ \sum_{1\leq m\leq M} \beta_{iB}(p_m) f_{iB}(p_m) W(p_m) \\
< &\sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} \int_{\mathcal{A}} f_{ki}(p)dp + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} \int_{\mathcal{A}} f_{ij}(p)dp + \int_{\mathcal{A}} \beta_{iB}(p) f_{iB}(p)dp \\
= &\int_{\mathcal{A}} \left[\sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki}(p) + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij}(p) + \beta_{iB}(p) f_{iB}(p) \right] dp \leq e_i
\end{aligned}$$

The first equality holds by (4.15) and (4.16). The last inequality holds by the energy constraint in solution $\varphi_{\text{U-MB}}$. Thus, the network lifetime of solution $\bar{\varphi}_{\text{C-MB}}$ is at least $(1-\varepsilon)\cdot T_{\text{U-MB}}$. This completes the proof. □

Lemma 7 is a powerful result. It states that for any given solution $\varphi_{\text{U-MB}}$ to the U-MB problem, we can find a solution $\varphi_{\text{C-MB}}$ for the set of FCPs (corresponding to a given ε), such that the network lifetime $T_{\text{C-MB}}$ is at least $(1-\varepsilon)$ of $T_{\text{U-MB}}$.

Now we are ready to prove Theorem 5.

Proof. Considering the special case of Lemma 7 that the given solution to U-MB problem is an optimal solution with network lifetime $T^*_{\text{U-MB}}$, we can transform it into a solution to C-MB problem on FCPs with network lifetime at least $(1-\varepsilon)T^*_{\text{U-MB}}$, i.e., there is a solution to C-MB problem on FCPs with network lifetime of at least $(1-\varepsilon)T^*_{\text{U-MB}}$. Thus, the solution $\psi^*_{\text{C-MB}}$ to C-MB problem on FCPs must have a network lifetime $T^*_{\text{C-MB}} \geq (1-\varepsilon)T^*_{\text{U-MB}}$. □

Theorem 5 guarantees that the network lifetime obtained from the LP solution on the M FCPs is at least $(1-\varepsilon)$ of $T^*_{\text{U-MB}}$. As discussed in Section 3.3.2, the N-tuple vector embodied by an FCP only represents the upper cost bound for any point in the corresponding subarea to the N nodes in the network and an FCP may not be mapped to a physical point, which is required in the final solution. In the following theorem, we show that if we have an optimal solution to the C-MB problem based on FCPs, we can construct a solution with each point being physically realizable. Further, the network lifetime for this constructed solution is greater than or equal to the maximum network lifetime for the C-MB problem, i.e., $T_{\text{U-MB}} \geq T^*_{\text{C-MB}}$. As a result, this new solution is $(1-\varepsilon)$-optimal.

Theorem 6. *For a given $\varepsilon > 0$, define subareas $\mathcal{A}_m$ and FCPs p_m, $1 \leq m \leq M$, as discussed in Section 3.3.2. Obtain an optimal solution $\psi^*_{\text{C-MB}}$ on these M FCPs with $W^*(p_m)$, $f^*_{ij}(p_m)$, $f^*_{iB}(p_m)$, and corresponding network lifetime $T^*_{\text{C-MB}}$ as discussed in Section 4.4. A $(1-\varepsilon)$-optimal solution*

$\psi_{\text{U-MB}}$ to U-MB problem can be constructed by having the base station stay in $\mathcal{A}_m$ for

$$W(\mathcal{A}_m) = W^*(p_m) \tag{4.17}$$

amount of time with a corresponding flow routing for any point $p \in \mathcal{A}_m$ as

$$f_{ij}(p) = f^*_{ij}(p_m)\ , \tag{4.18}$$

$$f_{iB}(p) = f^*_{iB}(p_m)\ . \tag{4.19}$$

In Theorem 6, note that in the constructed solution to U-MB problem, when the base station is at *any* point $p \in \mathcal{A}_m$, the flow routing is the same. In other words, the flow routing only depends on the subarea instead of specific point within this subarea.

Proof. In this proof, we will show that $\psi_{\text{U-MB}}$ is feasible, i.e., flow balance holds at any point, and the network lifetime of $\psi_{\text{U-MB}}$ is at least $T^*_{\text{C-MB}}$. Once we proved this, since $T^*_{\text{C-MB}} \geq (1-\varepsilon)T^*_{\text{U-MB}}$ (see Theorem 5), we know that $T_{\text{U-MB}}$ is at least $(1-\varepsilon)T^*_{\text{U-MB}}$, i.e., $\psi_{\text{U-MB}}$ is $(1-\varepsilon)$-optimal.

For flow balance when the base station location p is in any subarea $\mathcal{A}_m$, we have

$$\sum_{k\in\mathcal{N}}^{k\neq i} f_{ki}(p) + r_i = \sum_{k\in\mathcal{N}}^{k\neq i} f^*_{ki}(p_m) + r_i = \sum_{j\in\mathcal{N}}^{j\neq i} f^*_{ij}(p_m) + f^*_{iB}(p_m) = \sum_{j\in\mathcal{N}}^{j\neq i} f_{ij}(p) + f_{iB}(p)$$

The first equality holds by (4.18). The second equality holds by the flow balance in solution $\psi^*_{\text{C-MB}}$. The third equality holds by (4.18) and (4.19). Thus, solution $\psi_{\text{U-MB}}$ is feasible.

For the total consumed energy on node i at time $\sum_{1\leq m\leq M} W(\mathcal{A}_m) = \sum_{1\leq m\leq M} W^*(p_m) = T^*_{\text{C-MB}}$, we first have

$$\begin{aligned}
\int_{\mathcal{A}} \beta_{iB}(p) f_{ij}(p) W(p) dp &= \sum_{1\leq m\leq M} \int_{\mathcal{A}_m} \beta_{iB}(p) f_{ij}(p) W(p) dp \\
&\leq \sum_{1\leq m\leq M} \int_{\mathcal{A}_m} \beta_{iB}(p_m) f^*_{ij}(p_m) W(p) dp \\
&= \sum_{1\leq m\leq M} \beta_{iB}(p_m) f^*_{ij}(p_m) W(\mathcal{A}_m) \\
&= \sum_{1\leq m\leq M} \beta_{iB}(p_m) f^*_{ij}(p_m) W^*(p_m)
\end{aligned} \tag{4.20}$$

for any node $i \in \mathcal{N}$. The second inequality holds by (4.19) and $\beta_{iB}(p) \leq \beta_{iB}(p_m)$ in Property 1. The third equality holds by (4.11). The last equality holds by (4.17). Similarly, we have

$$\int_{\mathcal{A}} f_{ij}(p)W(p)dp = \sum_{1\leq m\leq M} f^*_{ij}(p_m)W^*(p_m) \tag{4.21}$$

for $i, j \in \mathcal{N}$ and $i \neq j$. Thus, we have

$$\begin{aligned}
& \int_{\mathcal{A}} \left[\sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} f_{ki}(p)W(p) + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} f_{ij}(p)W(p) + \beta_{iB}(p) f_{iB}(p)W(p) \right] dp \\
= & \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} \int_{\mathcal{A}} f_{ki}(p)W(p)dp + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} \int_{\mathcal{A}} f_{ij}(p)W(p)dp + \int_{\mathcal{A}} \beta_{iB}(p) f_{iB}(p)W(p)dp \\
\leq & \sum_{k\in\mathcal{N}}^{k\neq i} \beta_{\text{rec}} \sum_{1\leq m\leq M} f^*_{ki}(p_m)W^*(p_m) + \sum_{j\in\mathcal{N}}^{j\neq i} \beta_{ij} \sum_{1\leq m\leq M} f^*_{ij}(p_m)W^*(p_m) \\
& + \sum_{1\leq m\leq M} \beta_{iB}(p_m) f^*_{iB}(p_m)W^*(p_m) \\
\leq & e_i \ .
\end{aligned}$$

The first equality holds by (4.20) and (4.21). The second equality holds by the energy constraint in solution $\psi^*_{\text{C-MB}}$. Thus, the network lifetime of solution $\psi_{\text{U-MB}}$ is at least $T^*_{\text{C-MB}} \geq (1-\varepsilon)T^*_{\text{U-MB}}$. This completes the proof. □

4.5.3 Summary of Algorithm and Example

The design of the $(1-\varepsilon)$-optimal algorithm are given in Sections 4.5.1 and 4.5.2. We now summarize it into the following algorithm.

Algorithm 4. (A $(1-\varepsilon)$-Optimal Algorithm)

1. Narrow down the search space for base station within smallest enclosing disk $\mathcal{A}$.
2. Within the smallest enclosing disk $\mathcal{A}$, compute the lower and upper cost bounds $\beta^{\min}_{iB}$ and $\beta^{\max}_{iB}$ for each node $i \in \mathcal{N}$ by (3.3) and (3.4).

Table 4.2: Sensor locations, data rate, and initial energy of the example sensor network

Node Index	Location	Data rate	Initial energy
1	$(0.1, 0.5)$	0.8	390
2	$(1.1, 0.7)$	1.0	400
3	$(0.4, 0.1)$	0.5	130

3. For a given $\varepsilon > 0$, define a sequence of costs $\beta[1]$, $\beta[2], \cdots, \beta[H_i]$ by (3.5), where H_i is defined by (3.6).

4. For each node $i \in \mathcal{N}$, draw a sequence of $(H_i - 1)$ circles corresponding to cost $\beta[h]$ centered at node i, $1 \leq h < H_i$. The intersection of these circles within disk $\mathcal{A}$ will divide $\mathcal{A}$ into M subareas $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_M$.

5. For each subarea $\mathcal{A}_m$, $1 \leq m \leq M$, define an FCP p_m, which is represented by N-tuple cost vector $[\beta_{1B}(p_m), \beta_{2B}(p_m), \cdots, \beta_{NB}(p_m)]$, where $\beta_{iB}(p_m)$ is defined in (3.8).

6. For the C-MB problem on these M FCPs, apply the LP formulation in Section 4.4 and obtain an optimal solution $\psi^*_{\text{C-MB}}$ with $W^*(p_m)$, $f^*_{ij}(p_m)$, and $f^*_{iB}(p_m)$.

7. Construct a $(1 - \varepsilon)$-optimal solution $\psi_{\text{U-MB}}$ to U-MB problem based on $\psi^*_{\text{C-MB}}$ using the procedure in Theorem 6.

In the above algorithm, Step 6 has the highest complexity (solving an LP). By (3.9), the total number of subareas M is upper bounded by $O((N/\varepsilon)^2)$. Thus, the LP in Step 6 has polynomial size and the complexity of the above algorithm is polynomial.

Example 2. *To illustrate the steps in Algorithm 4, we solve a small* 3*-sensor network problem as an example. The location, data rate, and initial energy for each sensor are shown in Table 4.2, where the units of distance, rate, and energy are all normalized. We use* $\alpha = 2$ *in this example and the network setting are* $\beta_1 = 1$, $\beta_2 = 0.5$ *and* $\beta_{rec} = 1$ *under normalized units. For illustration, we*

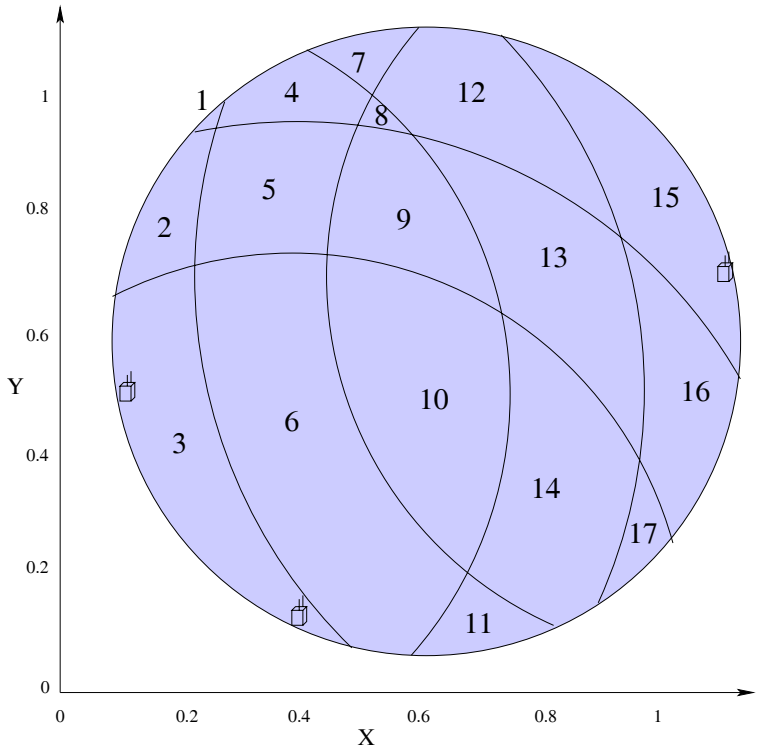


Figure 4.4: The subareas for the example sensor network.

set $\varepsilon = 0.2$.[2]

In Step 1, we identify SED $\mathcal{A}$ *with origin* $O_{\mathcal{A}} = (0.61, 0.57)$ *and radius* $R_{\mathcal{A}} = 0.51$ *(see Fig. 4.4).*

In Step 2, We have $D_{i,O_{\mathcal{A}}} = R_{\mathcal{A}} = 0.51$ *for each node* i, $1 \leq i \leq 3$. *We then find the lower and upper bounds of* β_{iB} *for each node* i *as follows.*

$$\begin{aligned} \beta_{iB}^{min} &= \beta_1 = 1 \\ \beta_{iB}^{max} &= \beta_1 + \beta_2 (D_{i,O_{\mathcal{A}}} + R_{\mathcal{A}})^{\alpha} = 1 + 0.5 \cdot (0.51 + 0.51)^2 = 1.52 \end{aligned}$$

In Step 3, for $\varepsilon = 0.2$, *we find*

$$H_i = \left\lceil \frac{\ln(1 + \frac{\beta_2}{\beta_1}(D_{i,O_{\mathcal{A}}} + R_{\mathcal{A}})^{\alpha})}{\ln(1+\varepsilon)} \right\rceil = \left\lceil \frac{\ln(1 + \frac{0.5}{1}(0.51 + 0.51)^2)}{\ln(1+0.2)} \right\rceil = 3$$

[2] In Section 4.5.4, we use $\varepsilon = 0.05$ for all numerical results.

for each node i, $1 \leq i \leq 3$, *and*

$$\begin{aligned}
\beta[1] &= \beta_1(1+\varepsilon) = 1 \cdot (1+0.2) = 1.20 \ , \\
\beta[2] &= \beta_1(1+\varepsilon)^2 = 1 \cdot (1+0.2)^2 = 1.44 \ , \\
\beta[3] &= \beta_1(1+\varepsilon)^3 = 1 \cdot (1+0.2)^3 = 1.73 \ .
\end{aligned}$$

In Step 4, we draw circles with centered at each node i, $1 \leq i \leq 3$, *and with cost* $\beta[h]$, $1 \leq h < H_i = 3$, *to divide the whole disk* $\mathcal{A}$ *into* 17 *subareas* $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_{17}$.

In Step 5, we define an FCP p_m *for each subarea* $\mathcal{A}_m$, $1 \leq m \leq 17$. *For example, for FCP* p_1, *we define the* 3*-tuple cost vector as* $[\beta_{1B}(p_m), \beta_{2B}(p_m), \beta_{3B}(p_m)] = [\beta[1], \beta[3], \beta[3]] = [1.20, 1.73, 1.73]$.

In Step 6, we obtain an optimal solution $\psi^*_{\text{C-MB}}$ *to C-MB problem on these* 17 *FCPs by the LP approach discussed in Section 4.4. We obtain the network lifetime* $T^*_{\text{C-MB}} = 190.37$, $W^*(p_3) = 157.00$, $W^*(p_6) = 33.37$, *and for all other* 15 *FCPs, we have* $W^*(p_m) = 0$ *(i.e., base station will not visit those areas). When the base station is at FCP* p_3, *the routing is* $f^*_{1B}(p_3) = 1.4$, $f^*_{2B}(p_3) = 1.0$, *and* $f^*_{31}(p_3) = 0.6$. *When the base station is at FCP* p_6, *the routing is* $f^*_{1B}(p_6) = 0.8$, $f^*_{2B}(p_6) = 1.0$, *and* $f^*_{3B}(p_6) = 0.6$.

In Step 7, we obtain a $(1-\varepsilon)$*-optimal solution* $\psi_{\text{U-MB}}$ *to U-MB problem as follows. Let the base station stay at any point in subarea* $\mathcal{A}_3$ *for* 157.00 *time and stay at any point in subarea* $\mathcal{A}_6$ *for* 33.37 *time. When the base station is at a point* p *in subarea* $\mathcal{A}_3$, *the routing is* $f_{1B}(p) = 1.4$, $f_{2B}(p) = 1.0$, *and* $f_{31}(p) = 0.6$. *When the base station is at a point* p *in subarea* $\mathcal{A}_6$, *the routing is* $f_{1B}(p) = 0.8$, $f_{2B}(p) = 1.0$, *and* $f_{3B}(p) = 0.6$. *The network lifetime for* $\psi_{\text{U-MB}}$ *is greater than or equal to* 190.37 *and is* $(1-\varepsilon)$*-optimal.*

4.5.4 Numerical Results

Now we apply the $(1-\varepsilon)$-optimal algorithm for larger sized networks and use numerical results to demonstrate the efficacy of the algorithm. We consider four randomly generated networks consist-

Table 4.3: Each node's location, data rate, and initial energy for a small 10-node network.

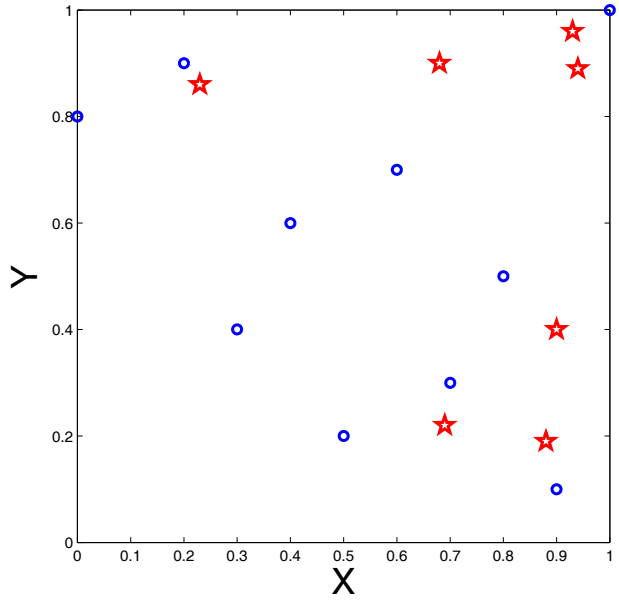
Location	Data rate	Initial energy	Location	Data rate	Initial energy
$(0.0, 0.8)$	0.8	150	$(0.6, 0.7)$	0.6	370
$(1.0, 1.0)$	1.0	200	$(0.4, 0.6)$	0.2	420
$(0.3, 0.4)$	0.6	130	$(0.2, 0.9)$	0.6	100
$(0.8, 0.5)$	0.8	460	$(0.9, 0.1)$	0.4	80
$(0.7, 0.3)$	0.3	170	$(0.5, 0.2)$	1.0	150

ing of 10, 20, 50, and 100 nodes deployed over a 1×1 square, respectively. The data rate and initial energy for each node are randomly generated between $[0.1, 1]$ and $[50, 500]$, respectively. The units of distance, rate, and energy are all normalized appropriately. The normalized parameters in energy consumption model are $\beta_1 = \beta_2 = \beta_{\text{rec}} = 1$. We assume the path loss index $\alpha = 2$.

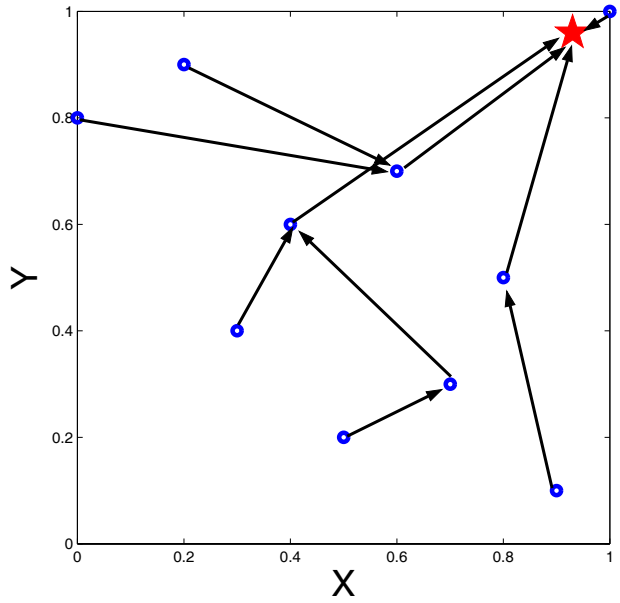
The required accuracy for approximation algorithm ε is set to $\varepsilon = 0.05$ for all numerical results. That is, we are pursuing a solution with a network lifetime that is at least 95% of the maximum network lifetime.

The network setting (location, data rate, and initial energy for each node) for the 10-node network is given in Table 4.3. By applying Algorithm 4, we obtain a $(1 - \varepsilon)$-optimal network lifetime 142.86, which is guaranteed to be at least 95% of the optimum. In Table 4.4, we have seven subareas that will be visited by the base station in the $(1 - \varepsilon)$-optimal solution. For illustration purpose, we use a point (x, y) within a subarea $\mathcal{A}_m$ to represent the approximate location of this subarea. For example, we use the point $(0.93, 0.96)$ to represent the subarea that contains this point. Table 4.4 lists the corresponding sojourn time for the base station to stay in each of these 7 subareas. The flow routing solution when the base station is in each of the 7 subareas is different as expected. For illustration, in Fig. 4.5(b), we show the flow routing solution when the base station is in subarea containing the point $(0.93, 0.96)$, where a circle represents a sensor node and a star represents a subarea for base station.

It is worth noting that for 95% of accuracy in optimality, there are only 7 subareas for the



(a) Locations to be visited by base station B in the $(1-\varepsilon)$-optimal solution



(b) Flow routing when the base station is at particular location.

Figure 4.5: Results for the small 10-node network.

Table 4.4: $(1-\varepsilon)$-optimal result for the small 10-node network with $\varepsilon = 0.05$.

$\mathcal{A}_m(x,y)$	$W(\mathcal{A}_m)$	$\mathcal{A}_m(x,y)$	$W(\mathcal{A}_m)$
$(0.93, 0.96)$	4.28	$(0.88, 0.19)$	3.54
$(0.68, 0.90)$	0.04	$(0.94, 0.89)$	50.36
$(0.69, 0.22)$	44.00	$(0.90, 0.40)$	27.15
$(0.23, 0.86)$	13.49		

Table 4.5: Each node's location, data rate, and initial energy for a 20-node network.

Location	Data rate	Initial energy	Location	Data rate	Initial energy
$(0.52, 0.02)$	0.6	480	$(0.29, 0.14)$	0.6	120
$(0.74, 0.76)$	0.3	310	$(0.05, 0.99)$	0.4	60
$(0.95, 0.03)$	0.8	150	$(0.84, 0.06)$	1.0	180
$(0.53, 0.63)$	0.6	220	$(0.99, 0.37)$	0.4	340
$(0.58, 1.00)$	0.4	230	$(0.73, 0.67)$	0.8	220
$(0.48, 0.84)$	0.7	160	$(0.53, 0.27)$	0.5	380
$(0.17, 0.83)$	0.1	380	$(0.57, 0.05)$	0.7	250
$(0.73, 0.39)$	0.1	500	$(0.88, 0.84)$	0.2	240
$(0.36, 0.98)$	0.1	430	$(0.26, 0.12)$	0.9	440
$(0.76, 0.02)$	0.7	500	$(0.71, 0.21)$	0.3	70

base station to visit. It turns out that for 20, 50, and 100 node networks, the number of subareas that needs to be visited by the base station is still very small (7 subareas for 20-node network, 8 subareas for 50-node network, and 12 subareas for 100-node network). This new observation is not obvious. But it is a good news as it hints that the base station does not need to be involved in frequent movement to achieve near-optimal solution.

The network setting for a 20-node network (with location, data rate, and initial energy for each of the 20 sensor nodes) is given in Table 4.5. By applying Algorithm 4, we obtain the $(1-\varepsilon)$-optimal network lifetime 144.23, which is guaranteed to be at least within 95% of the optimal.

Table 4.6: $(1-\varepsilon)$-optimal result for the 20-node network with $\varepsilon = 0.05$.

$\mathcal{A}_m(x,y)$	$W(\mathcal{A}_m)$	$\mathcal{A}_m(x,y)$	$W(\mathcal{A}_m)$
$(0.28, 0.33)$	2.86	$(0.27, 0.47)$	8.44
$(0.95, 0.86)$	9.62	$(0.80, 0.06)$	5.34
$(0.70, 0.11)$	108.05	$(0.88, 0.05)$	9.92

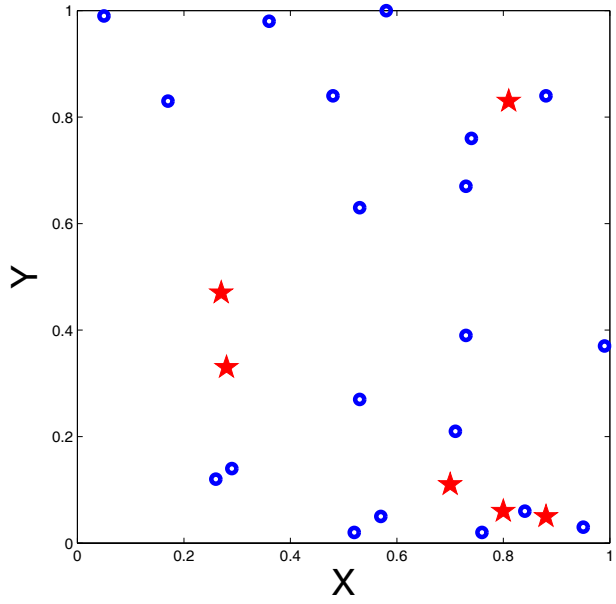


Figure 4.6: A 20-node network.

Table 4.7: $(1-\varepsilon)$-optimal result for the 50-node network with $\varepsilon = 0.05$.

$\mathcal{A}_m(x,y)$	$W(\mathcal{A}_m)$	$\mathcal{A}_m(x,y)$	$W(\mathcal{A}_m)$
$(0.13, 0.11)$	0.39	$(0.17, 0.64)$	1.47
$(0.32, 0.90)$	26.23	$(0.25, 0.61)$	27.72
$(0.49, 0.12)$	3.43	$(0.94, 0.10)$	9.66
$(0.34, 0.26)$	8.37	$(0.16, 0.30)$	45.03

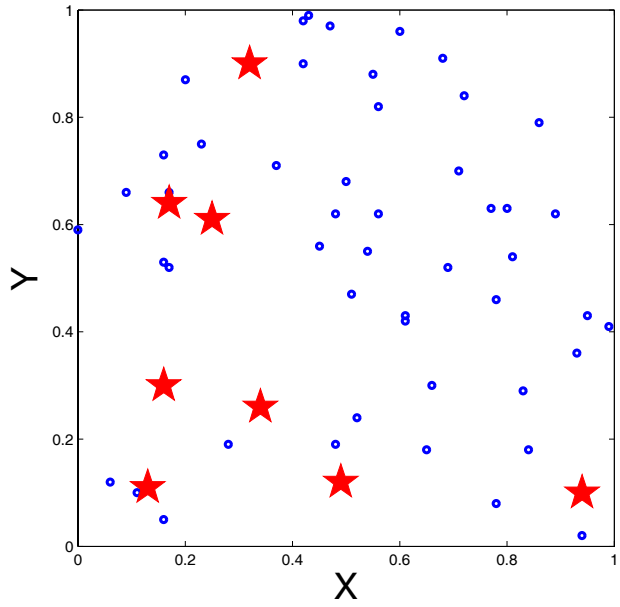


Figure 4.7: A 50-node network used in numerical investigation.

Again, we use a point within the chosen subarea $\mathcal{A}^*$ to represent the approximate location of this subarea. For this particular 20-node network setting, we have 6 subarea in the final solution, with the corresponding sojourn time in each subarea listed in Table 4.6.

For the 50-node network, the positions of the nodes are shown in Fig. 4.7, where a circle represents a sensor node and a star represents an optimal subarea for base station. The data rate and initial energy for each node are randomly generated between $[0.1, 1]$ and $[50, 500]$, respectively. By applying Algorithm 4, we obtain a $(1-\varepsilon)$-optimal network lifetime 122.30. Again, we use a point within a subarea $\mathcal{A}_m$ to represent the approximate location of this subarea. For this particular 50-node network setting, we have 8 subareas that the base station should visit in the final solution, with the corresponding sojourn time in each subarea listed in Table 4.7.

Finally, we consider a 100-node network shown in Fig. 4.8. The data rate and initial energy for each node are again randomly generated between $[0.1, 1]$ and $[50, 500]$, respectively. By applying Algorithm 4, we have a $(1-\varepsilon)$-optimal network lifetime 149.45. For this particular 100-node network setting, we have 12 subareas that the base station should visit in the solution, with the

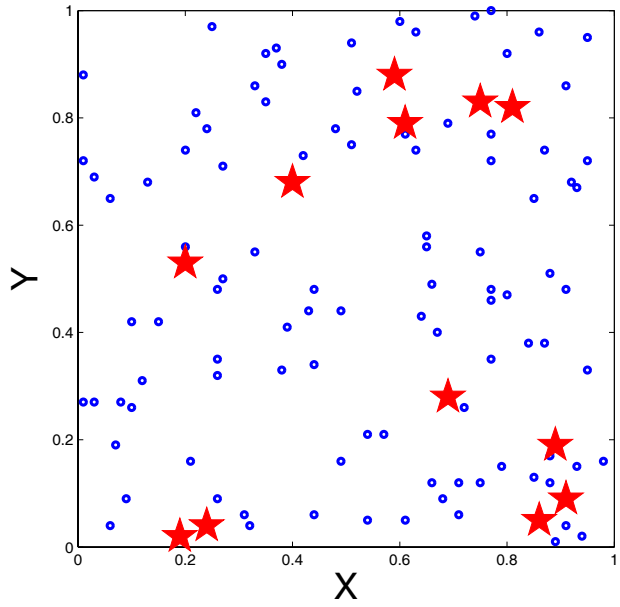


Figure 4.8: A 100-node network used in numerical investigation.

Table 4.8: $(1-\varepsilon)$-optimal result for the 100-node network with $\varepsilon = 0.05$.

$\mathcal{A}_m(x,y)$	$W(\mathcal{A}_m)$	$\mathcal{A}_m(x,y)$	$W(\mathcal{A}_m)$
$(0.75, 0.83)$	0.15	$(0.86, 0.05)$	0.38
$(0.69, 0.28)$	50.14	$(0.24, 0.04)$	2.10
$(0.59, 0.88)$	0.57	$(0.81, 0.82)$	0.09
$(0.20, 0.53)$	21.21	$(0.91, 0.09)$	0.05
$(0.40, 0.68)$	41.64	$(0.89, 0.19)$	3.16
$(0.61, 0.79)$	23.82	$(0.19, 0.02)$	6.14

corresponding sojourn time in each subarea listed in Table 4.8.

4.6 Related Work

Energy efficient routing has been an active area of research for sensor network in recent years (e.g., [86,92,96,105]). It is now well understood that energy efficient routing differs from lifetime-optimal routing as the former advocates the use of minimum energy-cost path, which may overload nodes along some common shared path, leading to poor performance in network lifetime.

Routing algorithms to maximize network lifetime has been an active area of research even for fixed base station location (see, e.g., [15, 17, 18, 87] and references therein). The focus is mainly devoted to how to split traffic flow along different routes and how to adjust power level at each node so that some optimal flow routing topology can be set up to maximize network lifetime. These early work have laid foundation on the importance of power control and flow routing topology on network lifetime performance.

There are some recent work on optimal base station placement [32, 73]. The focus on these efforts is on finding an optimal *fixed* position for the base station so that network lifetime can be maximized. However, as pointed out in [62,110], network lifetime can be substantially increased if the optimization space can be expanded to include movement of the base station during the course of sensor network operation.

Relevant work in the area of mobile base station for network lifetime problems include [5, 37, 62, 104, 110]. In [5, 37, 104], the locations of base station are constrained on a set of "pre-determined" locations. In [110], Younis et al. show that mobile base station can increase network lifetime. In [62], Luo and Hubaux propose to minimize the maximum load on a node among all the nodes in the network, which can be considered as an equivalent problem to maximize network lifetime. The results in [62, 110] are *heuristic*, and thus do not provide any theoretical bound on network lifetime performance.

Note that the mobile base station problem considered in this chapter differs fundamentally from the delay-tolerant network (DTN) (e.g., [53, 113]). DTN is assumed to experience frequent and long duration partition. The focus is to leverage storage at the intermediate nodes over long period of time and perform intermittent routing "over time" (i.e., delay tolerant) so as to achieve "eventual delivery". Network lifetime is not a major performance objective in the context of DTN.

4.7 Conclusions

The benefits of employing mobile base station to prolong sensor network lifetime are significant. However, due to the complexity of the problem (time-dependent network topology and traffic routing), provably optimal theoretical results have remained an open research area before this chapter. This chapter fills in this important gap by contributing a provably optimal algorithm regarding the movement of a mobile base station. The foundation of our result hinges upon a novel time-to-space transformation for problem formulation. Based on this transformation, we present an intermediate result which says that when the location of the base station are constrained to be on a set of pre-determined points, then the optimal solution can be obtained via a single LP. Building upon this intermediate result, our main result addressed the general mobile base station problem where the location of the base station is un-constrained. We developed a provably $(1-\varepsilon)$-optimal algorithm for the mobile base station problem such that the network lifetime is guaranteed to be at least $(1-\varepsilon)$ of the maximum network lifetime, where $\varepsilon > 0$ can be made arbitrarily small depending on required precision.

Chapter 5

Optimal Spectrum Sharing for CR Networks

5.1 Introduction

Recent studies sponsored by the FCC have shown that traditional fixed spectrum allocation policy is becoming inadequate in addressing today's rapidly evolving wireless communications. Studies show that many allocated spectrum blocks are not used in certain geographical areas and are idle most of the time. These frequency bands are called the spectrum "white space" or "hole" (a bandwidth is considered white space if it is wider than 1 MHz and remains unoccupied for at least 10 minutes). Measurements conducted by the Shared Spectrum Company [63] find that even in the most crowded area near downtown Washington, DC, where both government and commercial spectrum use is intensive, 62% of the spectrum remain white space. Another measurement, also conducted by the Shared Spectrum Company [64], shows that even during the 2004 Republican National Convention in New York City (perhaps the most heavily-congested area in the U.S. at that time), there was still significant white space available in the public sector spectrum. These studies have prompted the FCC to explore new innovative policies to encourage dynamic access to the under-utilized spectrum [35]. Wireless devices are allowed to sense and explore a wide range

of the frequency spectrum and identify currently unused spectrum blocks for data communication. This approach is also called *dynamic spectrum access* (DSA).

The enabling physical layer technology to realize DSA is *cognitive radio* (CR), which is a frequency-agile data communication device that has a rich control and monitoring (spectrum sensing) interface [45,69]. It capitalizes advances in signal processing and radio technology, as well as recent advancements in spectrum policy [83]. A frequency-agile radio module is capable of sensing the available bands [16,38,39,68,85,102], reconfiguring RF, and switching to newly-selected frequency bands. Thus, a CR can be programmed to tune and operate on specific frequency bands over a wide spectrum range [83]. An even more profound advance in CR technology is that *there is no requirement that selected frequencies/channels be contiguous:* the radio can send packets over non-contiguous frequency bands. From an application perspective, CR allows a single radio to provide a wide variety of functions, acting as a cell phone, broadcast receiver, GPS receiver, wireless data terminal, etc.

In this chapter, we focus on the *multi-hop* networking problem for a CR-based wireless network. For such a network, each node senses a set of spectrum bands that it can use. Due to the unequal size of spectrum bands, it is necessary to further divide each band into sub-bands (likely of unequal size) to schedule transmission and reception. There are many fundamental problems that can be posed for such a wireless network in the context of rates and capacity. In this chapter, we consider the following problem. Suppose there is a set of user sessions in the network that is characterized by a set of source-destination pairs each having a certain rate requirement. Then, how can we perform spectrum allocation, scheduling and interference avoidance, and multi-hop multi-path routing such that the required network-wide radio spectrum resource is minimized?

To formulate the problem mathematically, we characterize behaviors and constraints from multiple layers for a general multi-hop CR network. Special attention is given to modeling of spectrum sharing and unequal (non-uniform) sub-band division, scheduling and interference modeling, and multi-path routing. We formulate an optimization problem with the objective of minimizing the required network-wide radio spectrum resource for a set of source-destination pair rate require-

ments. Since such a problem formulation is a *mixed-integer non-linear program* (MINLP), which is NP-hard in general [40], we aim to derive a near-optimal solution.

We present a near-optimal algorithm for the formulated MINLP problem. First, we develop a lower bound for the objective by relaxing the integer variables and employing a linearization technique. This lower bound will be used as a measure for the quality of any solution. Then we present a novel *sequential fixing* (SF) solution procedure where the determination of integer variables is performed iteratively through a sequence of linear programs (LPs). Upon fixing all the integer variables, other variables in the optimization problem can be solved using an LP. Since the solution obtained by the proposed SF algorithm represents an upper bound for the objective, we compare it to the lower bound developed earlier. Simulations show that the results obtained by the SF algorithm are very close to the lower bound, thus suggesting that (1) the lower bound is very tight; and (2) the solution obtained by the SF algorithm is even closer to the optimum and thus is near-optimal. The significance of this theoretical work is to provide a performance benchmark which can be used to evaluate protocols and distributed algorithms for real implementation.

The remainder of this chapter is organized as follows. In Section 5.2, we characterize the behavior of CR networks from multiple layers and formulate them as mathematical constraints. We also elaborate on the optimal radio resource sharing problem and formulate it as an MINLP problem. In Section 5.3, we develop a lower bound for this MINLP problem by relaxing integer variables and using linearization. In Section 5.4, we describe the proposed SF algorithm. Section 5.5 presents simulation results and demonstrates the near-optimal performance of the SF algorithm. In Section 5.6, we review related work on CR and state-of-the-art on cross-layer optimization for MC-MR networks. Section 5.7 concludes this chapter.

5.2 CR Network Model and Problem Formulation

We consider an ad hoc network consisting of a set $\mathcal{N}$ of nodes. Among these nodes, there are a set $\mathcal{L}$ of uni-cast communication sessions. Denote $s(l)$ and $d(l)$ the source and destination nodes

Table 5.1: Notation in Chapter 5.

Symbol	Definition
$K^{(m)}$	Number of sub-bands in band m
M	The number of available bands in the network
Q	Transmission power spectral density at a transmitter
Q_T	The minimum threshold of power spectral density to decode a transmission at a receiver
Q_I	The maximum threshold of power spectral density for interference to be negligible at a receiver
R_T, R_I	Transmission range and interference range, respectively
$w^{(m,k)}$	The fraction of bandwidth for the k-th sub-band in band m
$W^{(m)}$	Bandwidth of band m
$x_{ij}^{(m,k)}$	Binary indicator to mark whether or not sub-band (m,k) is used by link $i \rightarrow j$.

of session $l \in \mathcal{L}$, and $r(l)$ the rate requirement (in b/s) of session l. Table 5.1 lists all notation that are used only in this chapter.

5.2.1 Modeling of Multi-layer Characteristics

Modeling of Spectrum Sharing and Sub-band Division. This mathematical modeling feature and constraints are unique to CR networks and do not exist in MC-MR networks. In a multi-hop CR network, the available spectrum bands at one node may be different from another node in the network. Given a set of available frequency bands at a node, the size (or bandwidth) of each band may differ drastically. For example, among the least-utilized spectrum bands found in [64], the bandwidth between $[1240, 1300]$ MHz (allocated to amateur radio) is 60 MHz, while bandwidth between $[1525, 1710]$ MHz (allocated to mobile satellites, GPS systems, and meteorological applications) is 185 MHz. Such large difference in bandwidths among the available bands suggests the need for further division of the larger bands into smaller sub-bands for more flexible and efficient frequency allocation. Since equal sub-band division of the available spectrum band is likely to yield sub-optimal performance, an unequal division is desirable.

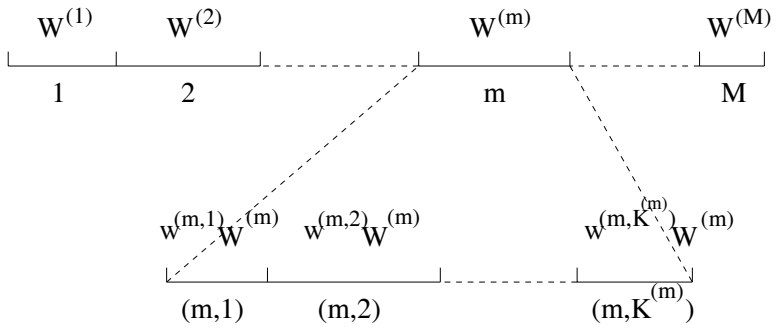


Figure 5.1: A schematic illustrating bands and sub-bands concepts in spectrum sharing.

More formally, we model the union of the available spectrum among all the nodes in the network as a set of M unequally sized bands (see Fig. 5.1). Denote $\mathcal{M}$ the set of these bands and $\mathcal{M}_i \subseteq \mathcal{M}$ the set of available bands (or white-space) at node $i \in \mathcal{N}$, which is likely to be different from that at another node, say $j \in \mathcal{N}$, i.e., possibly $\mathcal{M}_i \neq \mathcal{M}_j$. For example, at node i, $\mathcal{M}_i$ may consist of bands I, III, and V, while at node j, $\mathcal{M}_j$ may consist of bands I, IV, and VI. Denote $W^{(m)}$ the bandwidth of band $m \in \mathcal{M}$. For more flexible and efficient bandwidth allocation and to overcome the disparity in the bandwidth size among the spectrum bands, we assume that band m can be further divided into up to $K^{(m)}$ sub-bands, each of which may be of *unequal* bandwidth. Denote $w^{(m,k)}$ the fraction of bandwidth for the k-th sub-band in band m, which is part of our cross-layer optimization variables. Then we have

$$\sum_{k=1}^{K^{(m)}} w^{(m,k)} = 1 \ .$$

Note that some $w^{(m,k)}$'s can be 0 in the final optimization solution, in which case we will have fewer number of sub-bands than $K^{(m)}$.

As an example, Fig. 5.1 shows M bands in the network and for a specific band m, it displays a further division into $K^{(m)}$ sub-bands. Then the M bands in the network are effectively divided into $\sum_{m=1}^{M} K^{(m)}$ sub-bands, each of which may be of different size.

Transmission Range and Interference Range. We assume that the power spectral density from the transmitter of a CR node is Q. In this chapter, we assume that all nodes use the same power spectral density for transmission. The more complex issue of power control will be discussed in

Chapter 6. A widely-used model for power propagation gain is

$$g_{ij} = d_{ij}^{-\alpha} \ , \tag{5.1}$$

where α is the path loss index and d_{ij} is the distance between nodes i and j.[1] We assume a data transmission is successful only if the received power spectral density at a receiver exceeds a threshold Q_T. Likewise, we assume interference will become non-negligible when it produces a power spectral density over a threshold of Q_I at a receiver. Based on the threshold Q_T, the transmission range for a node is thus $R_T = (Q/Q_T)^{1/\alpha}$, which comes from $g_{ij} \cdot Q = Q_T$ and (5.1). Similarly, based on the interference threshold Q_I ($< Q_T$), the interference range for a node is $R_I = (Q/Q_I)^{1/\alpha}$. Since $Q_I < Q_T$, we have $R_I > R_T$. Both, the transmission range R_T and the interference range R_I, will be used in the modeling of the interference constraints.

Scheduling and Interference Constraints. Scheduling can be done either in time domain or frequency domain. In this chapter, we consider frequency domain sub-band assignment, i.e., how to assign sub-bands at a node for transmission and reception. A feasible scheduling of frequency bands must ensure that there is no interference *at the same node* and *among the nodes.*

Suppose that band m is available at both node i and node j, i.e., $m \in \mathcal{M}_i \bigcap \mathcal{M}_j$. To simplify the notation, let $\mathcal{M}_{ij} = \mathcal{M}_i \bigcap \mathcal{M}_j$. Denote

$$x_{ij}^{(m,k)} = \begin{cases} 1 & \text{If node } i \text{ transmits data to node } j \text{ on sub-band } (m,k), \\ 0 & \text{otherwise.} \end{cases}$$

For a node $i \in \mathcal{N}$ and a band $m \in \mathcal{M}_i$, denote $\mathcal{T}_i^m$ as the set of nodes that can use band m and are within the transmission range of node i, i.e.,

$$\mathcal{T}_i^m = \{j : d_{ij} \leq R_T, j \neq i, m \in \mathcal{M}_j\} \ .$$

Note that node i cannot transmit to multiple nodes on the same frequency sub-band. We therefore

[1] In this chapter, we consider a uniform gain model and assume the same gain model on all frequency bands. The case of a non-uniform gain model or a band-dependent gain behavior can be extended without much technical difficulty.

have

$$\sum_{q \in \mathcal{T}_i^m} x_{iq}^{(m,k)} \leq 1 \ . \tag{5.2}$$

Also, node i cannot use the same frequency sub-band for transmission and reception, due to "self-interference" at the physical layer. That is, if $x_{ij}^{(m,k)} = 1$, then for any $q \in \mathcal{T}_j^m$, $x_{jq}^{(m,k)}$ must be 0. In other words, we have

$$x_{ij}^{(m,k)} + \sum_{q \in \mathcal{T}_j^m} x_{jq}^{(m,k)} \leq 1 \ . \tag{5.3}$$

Note that in (5.3), we are referring to a specific node j to which node i is transmitting. If $x_{ij}^{(m,k)} = 1$, then $\sum_{q \in \mathcal{T}_j^m} x_{jq}^{(m,k)} = 0$, i.e., node j cannot use the same frequency sub-band (m, k) for transmission. On the other hand, if $x_{ij}^{(m,k)} = 0$, then $\sum_{q \in \mathcal{T}_j^m} x_{jq}^{(m,k)} \leq 1$, i.e., node j may use frequency sub-band (m, k) for transmission, but can only use it for one receiving node $q \in \mathcal{T}_j^m$ (same as in (5.2)).

In addition to the above constraints at the *same* node, there are also constraints on frequency sub-band use due to potential interference *among the nodes* in the network. In particular, for a frequency sub-band (m, k), if node i uses this sub-band for transmitting data to a node $j \in \mathcal{T}_i^m$, then any other node that can produce interference on node j should not use this sub-band.[2] To model this constraint, we denote by $\mathcal{P}_j^m$ the set of nodes that can produce interference at node j on band m, i.e.,

$$\mathcal{P}_j^m = \{p : d_{pj} \leq R_I, p \neq j, \mathcal{T}_p^m \neq \emptyset\} \ .$$

The physical meaning of $\mathcal{T}_p^m \neq \emptyset$ in the above definition is that node p may use band m for a valid transmission to a node in $\mathcal{T}_p^m$ and then may interfere node j. Then we have

$$x_{ij}^{(m,k)} + \sum_{q \in \mathcal{T}_p^m} x_{pq}^{(m,k)} \leq 1 \quad (p \in \mathcal{P}_j^m, p \neq i) \ . \tag{5.4}$$

In (5.4), if $x_{ij}^{(m,k)} = 1$, i.e., node i uses frequency sub-band (m, k) to transmit to node j, then any node p that can produce interference on node j should not transmit on this sub-band, i.e.,

[2] Note that the so-called "hidden terminal" problem is a special case under this constraint.

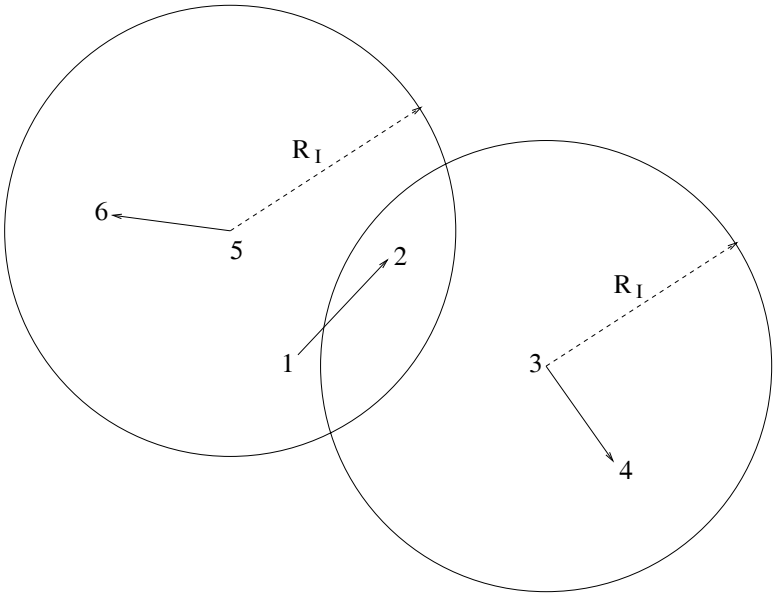


Figure 5.2: An example illustrating interference among links.

$\sum_{q\in\mathcal{T}_p^m} x_{pq}^{(m,k)} = 0$. On the other hand, if $x_{ij}^{(m,k)} = 0$, (5.4) degenerates into (5.2), i.e., node p may transmit on sub-band (m,k) to one node $q \in \mathcal{T}_p^m$, i.e., $\sum_{q\in\mathcal{T}_p^m} x_{pq}^{(m,k)} \leq 1$.

It is important to understand that in the interference constraint (5.4), if $x_{ij}^{(m,k)} = 0$, two nodes that can produce interference at node j but are far apart and outside each other's interference range can use the same sub-band (m,k) for transmission. We use an example to illustrate this point. In Fig. 5.2, suppose node 1 is transmitting to node 2 on sub-band (m,k), then any node that can produce interference at node 2 (i.e., node 3 or 5) cannot use the same sub-band for transmission. On the other hand, if node 1 is not using sub-band (m,k) to transmit to node 2, then node 3 may use this sub-band to transmit (to node 4) as stated in (5.4). Likewise, node 5 may also use this sub-band to transmit (to node 6) as stated in (5.4). That is, both nodes 3 and 5 may use the same sub-band for transmission.

We now use a compact form to include both (5.3) and (5.4). Denote

$$\mathcal{I}_j^m = \{p : d_{pj} \leq R_I, \mathcal{T}_p^m \neq \emptyset\}$$

which is equivalent to

$$\mathcal{I}_j^m = \begin{cases} \mathcal{P}_j^m \bigcup \{j\} & \text{If } \mathcal{T}_j^m \neq \emptyset \ , \\ \mathcal{P}_j^m & \text{otherwise} \ . \end{cases}$$

Thus, both (5.3) and (5.4) can be described by the following constraint.

$$x_{ij}^{(m,k)} + \sum_{q \in \mathcal{T}_p^m} x_{pq}^{(m,k)} \leq 1 \quad (p \in \mathcal{I}_j^m, p \neq i)$$

Routing. At the network level, a source node may need a number of relay nodes to route the data stream toward its destination node. Clearly, a route having a single path may be overly restrictive and is not able to take advantage of load balancing. A set of paths (or multi-path) is more flexible to route the traffic from a source node to its destination. Mathematically, this can be modeled as follows.

Denote $f_{ij}(l)$ the data rate on link $i \rightarrow j$ that is attributed to session l, where $i \in \mathcal{N}, j \in \bigcup_{m \in \mathcal{M}_i} \mathcal{T}_i^m$, and $l \in \mathcal{L}$. To simplify the notation, let $\mathcal{T}_i = \bigcup_{m \in \mathcal{M}_i} \mathcal{T}_i^m$. If node i is the source node of session l, i.e., $i = s(l)$, then

$$\sum_{j \in \mathcal{T}_i} f_{ij}(l) = r(l) \ . \tag{5.5}$$

If node i is an intermediate relay node for session l, i.e., $i \neq s(l)$ and $i \neq d(l)$, then

$$\sum_{j \in \mathcal{T}_i}^{j \neq s(l)} f_{ij}(l) = \sum_{p \in \mathcal{T}_i}^{p \neq d(l)} f_{pi}(l) \ . \tag{5.6}$$

If node i is the destination node of session l, i.e., $i = d(l)$, then

$$\sum_{p \in \mathcal{T}_i} f_{pi}(l) = r(l) \ . \tag{5.7}$$

It can be easily verified that if (5.5) and (5.6) are satisfied, then (5.7) must be satisfied. As a result, it is sufficient to list only (5.5) and (5.6) in the formulation.

In addition to the above flow balance equations at each node i for each session l, the aggregate flow rates on each radio link cannot exceed this link's capacity. To model this mathematically, we

need to first find the capacity on link (i, j) in sub-band (m, k). If node i sends data to node j on sub-band (m, k), i.e., $x_{ij}^{(m,k)} = 1$, then the capacity on link $i \rightarrow j$ in sub-band (m, k) is

$$c_{ij}^{(m,k)} = w^{(m,k)} W^{(m)} \log_2 \left(1 + \frac{g_{ij} Q}{\eta}\right) ,$$

where η is the ambient Gaussian noise density. Note that the denominator inside the log function contains only η. This is due to one of our interference constraints stated earlier, i.e., when node i is transmitting to node j on sub-band (m, k), then all the other neighbors of node j within its interference range are prohibited from using this sub-band. This interference constraint significantly helps to simplify the calculation of the link capacity $c_{ij}^{(m,k)}$. When $x_{ij}^{(m,k)} = 0$, we have $c_{ij}^{(m,k)} = 0$. Thus, $c_{ij}^{(m,k)}$ can be written in the following compact form.

$$c_{ij}^{(m,k)} = x_{ij}^{(m,k)} \cdot w^{(m,k)} W^{(m)} \log_2 \left(1 + \frac{g_{ij} Q}{\eta}\right) . \tag{5.8}$$

Now, returning to our earlier requirement that the aggregate data rates on each link $i \rightarrow j$ cannot exceed the link's capacity, we have,

$$\sum_{l \in \mathcal{L}}^{s(l) \neq j, d(l) \neq i} f_{ij}(l) \leq \sum_{m \in \mathcal{M}_{ij}} \sum_{k=1}^{K^{(m)}} c_{ij}^{(m,k)} = \sum_{m \in \mathcal{M}_{ij}} \sum_{k=1}^{K^{(m)}} x_{ij}^{(m,k)} \cdot w^{(m,k)} W^{(m)} \log_2 \left(1 + \frac{g_{ij} Q}{\eta}\right) .$$

5.2.2 Problem Formulation

For the multi-hop CR networks that we are investigating, various performance objectives can be used. In this chapter, we use the total required radio resource to support the user sessions as our performance objective. The radio resource can be measured in terms of the total bandwidth used by all nodes in the network, which is the simplified form of the so-called *space-bandwidth product* proposed in [61] when the transmission power spectral density is fixed. It is not hard to see that the solution procedure in this chapter can be applied when other performance objectives are used.

To re-cap, we are given a set of source-destination pairs (user sessions) in the network, each with a certain rate requirement. Each node in the network has a set of available frequency bands that it can use for communication. We want to find an optimal solution to divide the set of available frequency bands at each node, the scheduling of sub-bands for transmission and reception,

and multi-hop routing for each flow such that the total radio bandwidth used in the network is minimized (or we find that there is no feasible solution). Mathematically, we have the following optimization problem,

$$\begin{array}{lll}
\text{Min} & \sum_{i\in\mathcal{N}} \sum_{m\in\mathcal{M}_i} \sum_{j\in\mathcal{T}_i^m} \sum_{k=1}^{K^{(m)}} W^{(m)} x_{ij}^{(m,k)} w^{(m,k)} & \\
\text{s.t.} & \sum_{k=1}^{K^{(m)}} w^{(m,k)} = 1 & (m \in \mathcal{M}) \\
& \sum_{q\in\mathcal{T}_i^m} x_{iq}^{(m,k)} \leq 1 & (i\in\mathcal{N}, m\in\mathcal{M}_i, 1\leq k\leq K^{(m)}) \quad (5.9) \\
& x_{ij}^{(m,k)} + \sum_{q\in\mathcal{T}_p^m} x_{pq}^{(m,k)} \leq 1 & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m, \\
& & 1\leq k\leq K^{(m)}, p\in\mathcal{I}_j^m, p\neq i) \quad (5.10)
\end{array}$$

$$\sum_{l\in\mathcal{L}}^{s(l)\neq j, d(l)\neq i} f_{ij}(l) - \sum_{m\in\mathcal{M}_{ij}} \sum_{k=1}^{K^{(m)}} W^{(m)} \log_2 \left(1 + \frac{g_{ij}Q}{\eta}\right) x_{ij}^{(m,k)} w^{(m,k)} \leq 0 \quad (i \in \mathcal{N}, j \in \mathcal{T}_i)$$

$$\begin{array}{ll}
\sum_{j\in\mathcal{T}_i} f_{ij}(l) = r(l) & (l \in \mathcal{L}, i = s(l)) \\
\sum_{j\in\mathcal{T}_i}^{j\neq s(l)} f_{ij}(l) - \sum_{p\in\mathcal{T}_i}^{p\neq d(l)} f_{pi}(l) = 0 & (l \in \mathcal{L}, i \in \mathcal{N}, i \neq s(l), d(l)) \\
x_{ij}^{(m,k)} = 0 \text{ or } 1, w^{(m,k)} \geq 0 & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m, 1 \leq k \leq K^{(m)}) \\
f_{ij}(l) \geq 0 & (l \in \mathcal{L}, i \in \mathcal{N}, i \neq d(l), j \in \mathcal{T}_i, j \neq s(l)) \,,
\end{array}$$

where $W^{(m)}$, g_{ij}, Q, η, and $r(l)$ are all constants and $x_{ij}^{(m,k)}$, $w^{(m,k)}$, and $f_{ij}(l)$ are all optimization variables.

The above optimization problem is a *mixed-integer non-linear programming* (MINLP) problem, which is NP-hard in general [40]. Although existing software (e.g. BARON [4]) can solve very small-sized network instances (e.g., several nodes), the time complexity becomes prohibitively high for larger networks.

Our approach to solve this problem is as follows. In Section 5.3, we first explore a lower bound for the objective, which can be obtained by relaxing the integer variables and using a linearization technique. Using this lower bound as a performance benchmark, in Section 5.4, we develop a highly effective algorithm based on a novel *sequential fixing* (SF) procedure. Using simulation results, we show that the SF algorithm has a performance very close to the lower bound. Since

the optimal objective value lies between the lower bound and the solution obtained by the SF algorithm, the heuristic solution must be even closer to the true optimum.

5.3 A Lower Bound for the Objective Function

The complexity of the problem formulated in Section 5.2.2 arises from the binary $x_{ij}^{(m,k)}$ variables and the product of variables $x_{ij}^{(m,k)} w^{(m,k)}$. To pursue a lower bound for the objective, we first multiplying (5.9) and (5.10) by the corresponding $w^{(m,k)}$, so that $x_{ij}^{(m,k)}$ appears throughout as a product with $w^{(m,k)}$. We then relax the integrality (binary) requirement on $x_{ij}^{(m,k)}$ with $0 \leq x_{ij}^{(m,k)} \leq 1$ and replace $x_{ij}^{(m,k)} w^{(m,k)}$ with a single variable, say $s_{ij}^{(m,k)}$, i.e., $s_{ij}^{(m,k)} = x_{ij}^{(m,k)} w^{(m,k)} \leq w^{(m,k)}$. Such a relaxation leads to the following lower-bounding problem formulation.

$$
\begin{aligned}
\text{Min} \quad & \textstyle\sum_{i\in\mathcal{N}} \sum_{m\in\mathcal{M}_i} \sum_{j\in\mathcal{T}_i^m} \sum_{k=1}^{K^{(m)}} W^{(m)} s_{ij}^{(m,k)} \\
\text{s.t.} \quad & \textstyle\sum_{k=1}^{K^{(m)}} w^{(m,k)} = 1 && (m \in \mathcal{M}) \\
& \textstyle\sum_{q\in\mathcal{T}_i^m} s_{iq}^{(m,k)} - w^{(m,k)} \leq 0 && (i \in \mathcal{N}, m \in \mathcal{M}_i, 1 \leq k \leq K^{(m)}) \quad (5.11) \\
& s_{ij}^{(m,k)} + \textstyle\sum_{q\in\mathcal{T}_p^m} s_{pq}^{(m,k)} - w^{(m,k)} \leq 0 && (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m, \\
& && 1 \leq k \leq K^{(m)}, p \in \mathcal{I}_j^m, p \neq i) \quad (5.12) \\
& \textstyle\sum_{l\in\mathcal{L}}^{s(l)\neq j, d(l)\neq i} f_{ij}(l) - \sum_{m\in\mathcal{M}_{ij}} \sum_{k=1}^{K^{(m)}} W^{(m)} \log_2 \left(1 + \frac{g_{ij}Q}{\eta}\right) s_{ij}^{(m,k)} \leq 0 && (i \in \mathcal{N}, j \in \mathcal{T}_i) \\
& \textstyle\sum_{j\in\mathcal{T}_i} f_{ij}(l) = r(l) && (l \in \mathcal{L}, i = s(l)) \\
& \textstyle\sum_{j\in\mathcal{T}_i}^{j\neq s(l)} f_{ij}(l) - \sum_{p\in\mathcal{T}_i}^{p\neq d(l)} f_{pi}(l) = 0 && (l \in \mathcal{L}, i \in \mathcal{N}, i \neq s(l), d(l)) \\
& w^{(m,k)}, s_{ij}^{(m,k)} \geq 0 && (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m, 1 \leq k \leq K^{(m)}) \\
& f_{ij}(l) \geq 0 && (l \in \mathcal{L}, i \in \mathcal{N}, i \neq d(l), j \in \mathcal{T}_i, j \neq s(l))
\end{aligned}
$$

This new (relaxed) formulation is a standard linear program (LP), the solution of which can be obtained in polynomial-time. Due to the relaxation (and thus enlarged optimization space), the solution value to this LP problem yields a lower bound on the objective value of the original problem in Section 5.2.2. Note that there may not exist a feasible solution that achieves this lower bound.

Nevertheless, this lower bound offers a benchmark to measure the quality of a feasible solution, which we will develop in the next section. It turns out that this lower bound is extremely tight (see results in Section 5.4). This can be explained by the convex hull results presented by Sherali et al. [88].

5.4 A Near-Optimal Algorithm Based on Sequential Fixing

5.4.1 Basic Algorithm

We now take a closer look at the original MINLP problem formulation in Section 5.2.2. Observe that once the binary values for all x variables are determined, i.e., whether or not a node will indeed use a particular sub-band to send data to another node, then this MINLP reduces to an LP, which can be solved in polynomial-time. In other words, the key obstacle in solving this MINLP problem lies in the determination of the binary values for the x variables. To this end, we propose a two-step solution procedure: i) fix the binary values for x variables iteratively through a sequence of LPs; ii) once all the x variables are fixed, find a solution (to determine how to divide sub-bands and flow routing) corresponding to this set of x values. Such a two-step approach will yield a sub-optimal (upper bound) solution to the original MINLP problem. The quality of this algorithm can be assessed by how close its solution is to the lower bound we developed in the previous section.

As said, the key to the two-step approach resides in the determination of the binary values for all the x variables. Our main idea is to fix (set) the values of the x variables *sequentially* through solving a series of relaxed LP problems, with each iteration setting at least one binary value for some $x_{ij}^{(m,k)}$. Specifically, during the first iteration, we relax all binary variables $x_{ij}^{(m,k)}$ to $0 \leq x_{ij}^{(m,k)} \leq 1$ as in Section 5.3 to obtain an LP. Upon solving this LP, we have a solution with each $x_{ij}^{(m,k)} = s_{ij}^{(m,k)}/w^{(m,k)}$ being a value between 0 and 1. Among all the x values, suppose some $x_{ij}^{(m,k)}$ has the largest value. Then we fix (set) this particular $x_{ij}^{(m,k)}$ to 1. As a result of this fixing, by (5.9), we also need to fix $x_{iq}^{(m,k)} = 0$ for $q \in \mathcal{T}_i^m$ and $q \neq j$. By (5.10), we also fix $x_{pq}^{(m,k)}$ to 0

Sequential fixing (SF) Algorithm

1. Set up and solve the initial relaxed LP problem as shown in Section 5.3.
2. Suppose $x_{ij}^{(m,k)}$ is the largest among all the x values that remain to be fixed, fix this $x_{ij}^{(m,k)} = 1$. Also fix $x_{iq}^{(m,k)} = 0$ (for $q \in \mathcal{T}_i^m$ and $q \neq j$) and $x_{pq}^{(m,k)} = 0$ (for $p \in \mathcal{I}_j^m, p \neq i$, and $q \in \mathcal{T}_p^m$).
3. If all the x variables are fixed, go to Step 5.
4. Otherwise, reformulate and solve a new relaxed LP problem with newly fixed x variables and go to Step 2.
5. Formulate an LP problem based on all fixed x values. Obtain a solution to this LP problem.

Figure 5.3: Sequential fixing (SF) algorithm.

for $p \in \mathcal{I}_j^m, p \neq i$, and $q \in \mathcal{T}_p^m$. Technically, in an implementation, we can fix all the x variables that have a value of 1 and perform an additional fixing for the largest fractional variable as above.

Now, having fixed some x variables in the first iteration, we update the problem to obtain a new LP for the second iteration as follows. For those $x_{ij}^{(m,k)}$ variables that are already fixed at 1, since $s_{ij}^{(m,k)} = x_{ij}^{(m,k)} w^{(m,k)} = w^{(m,k)}$, we can replace the corresponding $s_{ij}^{(m,k)}$ by $w^{(m,k)}$. For those $x_{iq}^{(m,k)}$ and $x_{pq}^{(m,k)}$ that are fixed to 0, we can set $s_{iq}^{(m,k)} = 0$ and $s_{pq}^{(m,k)} = 0$. As a result, all the terms in the LP involving these s variables can be removed and the corresponding constraint in (5.11) and (5.12) can also be removed.

In the second iteration, we solve this new LP and then fix some additional x variables based on the same process (now the ordering of the x values is done only for the remaining un-fixed x variables). The iteration continues and eventually we obtain all x variables as either 0 and 1.

Upon fixing all the x values, the original MINLP reduces to an LP problem, which can be solved in polynomial-time. Unlike the solutions obtained in Section 5.3, the final solution obtained here is a *feasible* solution since all x values are binary. The complete sequential fixing (SF) algorithm is given in Fig. 5.3.

5.4.2 An Iteration-Speedup Technique

In the SF algorithm, we need to solve a sequence of LPs. The complexity of SF is polynomial. By exploiting the space and frequency dimensions involved in radio resource allocation, we may decrease the number of LPs by fixing more x variables during each iteration in Fig. 5.3. As a result, the complexity can be further decreased. From a space dimension viewpoint, a sub-band usage will only have an impact within the interference range and the same sub-band can be used by other links outside this range. Thus, for the same sub-band (m, k), we can then fix *multiple* links that have non-overlapping interference ranges within a single iteration of the sequential fixing algorithm. From the frequency dimension viewpoint, the transmission in one sub-band will not interfere with the transmission in a different sub-band. Thus, for the same link $i \rightarrow j$, we can then fix multiple sub-bands within a single iteration of the SF algorithm. Specifically, we can use a threshold $\alpha > 0.5$ in this fixing process and fix all the x variables that exceeds α to 1 in a single iteration. Note that in (5.9) and (5.10), it is required that at most one binary variable $x_{ij}^{(m,k)} = 1$ while in the relaxed problem, there is at most one fraction $s_{iq}^{(m,k)}/u^{(m,k)} > 0.5$. Thus, $\alpha > 0.5$ ensures that both the constraints (5.9) and (5.10) (interference constraints at each node and among the nodes) will hold during the SF procedure.[3] In the case that none of the x variables exceed α, we will fall back to the basic algorithm in Fig. 5.3 and simply choose the largest valued x variable.

5.5 Simulation Results

In this section, we present simulation results for our SF algorithm and compare them to the lower bounds obtained in Section 5.3. The units for distance, rate, and power density satisfy (5.1) and (5.8) and are all normalized with appropriate dimensions. We consider $|\mathcal{N}| = 20, 30$ or 40 nodes in a 50x50 area. Among these nodes, there are $|\mathcal{L}| = 5$ active sessions, each with a rate within $[10, 100]$.

We assume that there are $M = 5$ bands that can be used for the entire network (see Table 5.2).

[3] We use $\alpha = 0.85$ in our simulation results.

Table 5.2: Available bands among all nodes in the network in the simulation study.

Band Index	Spectrum Range (MHz)	Bandwidth (MHz)
I	$[1240, 1300]$	60.0
II	$[1525, 1710]$	185.0
III	$[902, 928]$	26.0
IV	$[2400, 2483.5]$	83.5
V	$[5725, 5850]$	125.0

Bands I and II are among the least-utilized (less than 2%) spectrum bands found in [64] and bands III, IV, and V are unlicensed ISM bands used for 802.11. Recall that available bands at each CR node is a *subset* of these five bands based on its location and the available bands at any two nodes in the network may not be identical. In the simulation, this is done by randomly selecting a subset of bands from the pool of five bands for each node. Further, we assume bands I to V can be divided into up to $3, 5, 2, 4$, and 4 sub-bands although other desirable divisions can be used. Note that the size of each sub-band may be unequal and is part of the optimization problem.

We assume that the transmission range at each node is 20 and that the interference range is 30, although other settings can be used. The path loss index α is assumed to be 4. The threshold Q_T is assumed to be η. Thus, we have $Q_I = \left(\frac{20}{30}\right)^{\alpha} Q_T = \frac{16}{81}\eta$ and the transmission power spectral density $Q = (20)^{\alpha} Q_T = 1.6 \cdot 10^5 \eta$.

Note that it is possible that there is no feasible solution for a specific data set. This could be attributed to dis-connectivity in the network (due to random network topology), resource bottleneck in a hot area, etc. Thus, we only report results based on those data sets that have feasible solutions.

We first present simulation results for 100 data sets for 20-node networks that can produce feasible solutions. For each data set, the network topology, source/destination pair and bit rate of each session, and available frequency bands at each node are randomly generated. We use the SF Algorithm to determine the cost, which is the total required bandwidth in the objective function. As discussed, we compare this result with the lower bound developed in Section 5.3. The running

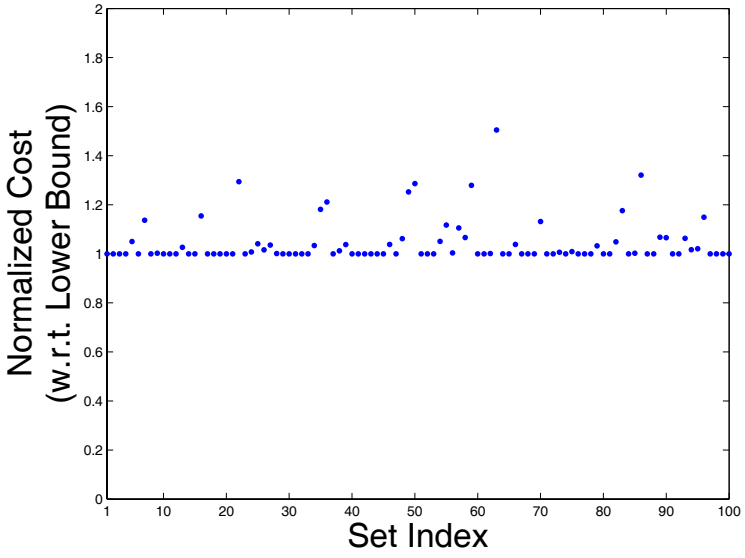


Figure 5.4: Normalized cost (with respect to lower bound) for 100 data sets of 20-node networks.

time for each simulation is less than 10 seconds on a Pentium 3.4 GHz machine.

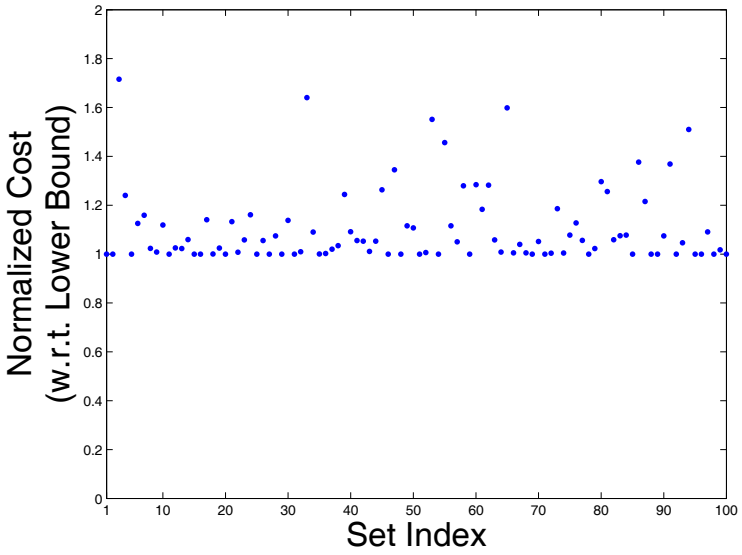
Figure 5.4 shows the normalized costs obtained by the SF algorithm with respect to the lower bound costs for 100 data sets. The average normalized cost among the 100 simulations is 1.04 and the standard derivation is 0.07. There are two observations that can be made from this figure. First, since the ratio of the solution obtained by SF (upper bound of optimal solution) to the lower bound solution is close to 1 (in many cases, they coincide with each other), the gap between them is very small. Second, since the optimal solution (unknown) is between the solution obtained by the SF algorithm and the lower bound, the lower bound is very tight; more importantly, *the SF solution must be even closer to the optimum.*

To get a sense of how the actual (rather than normalized) numerical results appear in the simulations, we list the first 40 sets of results in Table 5.3. Note that in many cases, the result obtained by the SF algorithm is identical to the respective lower bound obtained via relaxation. This indicates that the solution found by SF is optimal.

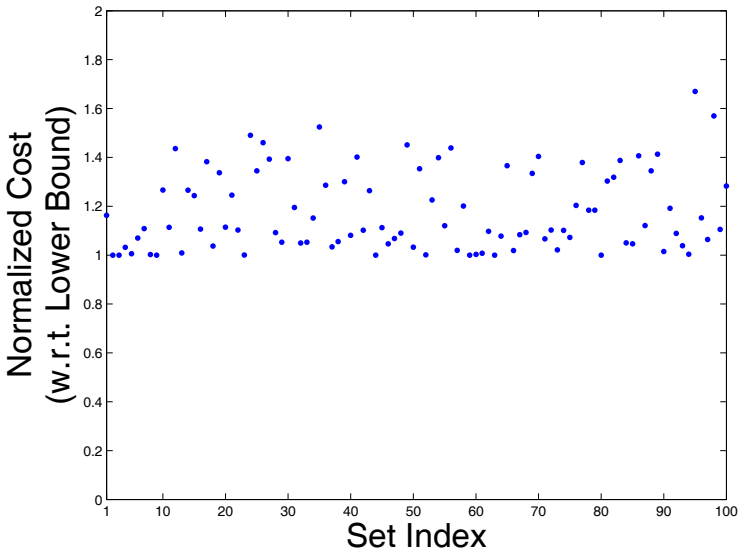
Simulation results for 100 random data sets for 30-node and 40-node networks that produce feasible solutions are displaced in Figs. 5.5(a) and (b), respectively. For 30-node networks, the

Table 5.3: Simulation results for the first 40 data sets of 20-node networks.

Data Set Index	Lower Bound	Result by SF	Data Set Index	Lower Bound	Result by SF
1	138.33	138.33	21	156.43	156.43
2	156.12	156.12	22	238.41	308.51
3	173.53	173.53	23	184.78	184.78
4	189.70	189.70	24	241.42	243.22
5	203.05	213.18	25	135.39	140.96
6	184.37	184.37	26	247.30	251.18
7	160.45	182.33	27	280.80	290.85
8	232.23	232.23	28	353.98	354.17
9	223.00	223.53	29	260.56	260.56
10	182.13	182.13	30	127.06	127.06
11	220.20	220.20	31	170.35	170.35
12	277.83	277.83	32	207.74	207.74
13	130.54	134.05	33	183.59	183.59
14	172.62	172.62	34	138.33	143.00
15	256.96	256.96	35	270.76	319.71
16	178.73	178.73	36	325.59	394.43
17	152.08	152.08	37	288.72	288.72
18	359.03	359.03	38	244.77	247.74
19	150.61	150.61	39	215.72	223.83
20	164.97	164.97	40	126.05	126.05



(a) 100 data sets of 30-node networks.



(b) 100 data sets of 40-node networks.

Figure 5.5: Normalized cost (with respect to lower bound).

average normalized cost among the 100 simulations is 1.10 and the standard derivation is 0.16. For 40-node networks, the average normalized cost among the 100 simulations is 1.18 and the standard derivation is 0.16. Thus, the SF solutions are also close to the optimal solutions.

5.6 Related Work

CR is based on *software defined radio* (SDR) [83]. Since its inception, SDR development has witnessed rapid advances. Standards bodies such as IEEE 802 Standards Committee, the SDR Forum, the Object Management Group have been instrumental in promoting open standards for SDR commercialization. Among others, the Software Communications Architecture core framework is the result from standardization efforts on SDR. The IEEE 802.22 working group is in the process of developing a standard for a CR interface for license-exempt devices on a non-interfering basis in spectrum that is allocated to the TV Broadcast Service. CR employs all the SDR technologies, plus the additional capability of spectrum sensing and cognition (learning and adaptation).

In CR research community, there have been extensive activities devoted to effective sharing of spectrum or spectrum allocation. For a multi-user single-hop communication in a network environment, a number of approaches have been proposed. For example, in [22, 72], game theory was applied to study spectrum sharing, while in [50, 51], pricing mechanism was used. In [34], Etkin et al. studied a utility maximization problem and solved it under certain condition. In [74], Peng et al. studied the spectrum assignment problem with the aim of maximizing the total utility. In these efforts, routing is not part of the problem.

For the multi-hop networking problem with CRs, there is limited amount of work to date available in the literature. In [112], Zhao et al. designed a distributed coordination approach for spectrum sharing. They showed that this approach offers throughput improvement over a dedicated channel approach. In [101], Ugarte and McDonald studied the network capacity problem for multi-hop CR networks and found an upper bound, although it is not clear how tight this bound is. In [108], Xin et al. studied how to assign frequency bands at each node to form a topology such

that a certain performance metric can be optimized. A layered graph was proposed to model frequency bands available at each node and to facilitate topology formation and achieve optimization objective. The authors considered the so-called fixed channel approach whereby the radio is assumed to operate on only one channel at a specific time. In [95], Steenstrup studied three different frequency assignment problems: common broadcast frequencies, non-interfering frequencies for simultaneous transmissions, and frequencies for direct source-destination communications. Each is viewed as a graph-coloring problem, and both centralized and distributed algorithms were presented. Within these limited efforts, there remains a lack of results on fundamental theoretical performance limits for multi-hop CR networks.

A closely related line of research is the so-called *multi-channel multi-radio* (MC-MR) networks (e.g., [3,31,57,58,81]). It is important to understand that a CR is vastly more powerful and flexible than MC-MR technology. First, the MC-MR platform employs a traditional *hardware-based* radio technology (i.e., signal processing, modulation, etc., are all implemented in the hardware), and thus each radio can only operate on a single channel at a time and *there is no switching of channels on the packet level.* As a result, the number of concurrent channels that can be used at a wireless node is limited by the number of hardware-based radios. In contrast, the radio technology in CR is software-based; a CR is capable of switching frequency bands on the packet level. As a result, the number of concurrent frequency bands that can be shared by a single CR is typically much larger than that which can be supported by MC-MR. Second, due to the nature of hardware-based radio technology in MC-MR, a common assumption in MC-MR is that there is a set of "common channels" available for every node in the network; each channel typically has the same bandwidth. However, such an assumption is hardly true for CR networks, in which each node may not have an identical set of frequency bands and each band is likely to be of unequal size. Due to this difference, CR is required to work on a set of frequency bands that are scattered over widely-separated slices of the frequency spectrum with different bandwidths. In summary, these important differences between MC-MR and CR warrant that the algorithmic design for a CR network is substantially more complex than that for MC-MR. In some sense, an MC-MR-based wireless network can be considered as a special case of a CR-based wireless network. Thus, algorithms designed for CR

networks can be tailored to address MC-MR networks, while the converse is not true.

5.7 Conclusions

In this chapter, we conducted a systematic study on the important problem of multi-hop networking with CR nodes. The nature of the problem calls for a characterization and modeling of multi-layer behaviors and constraints. We characterized behaviors and constraints for a multi-hop CR network from multiple layers, including the modeling of spectrum sharing and sub-band division, scheduling and interference constraints, and flow routing. We formulated an optimization problem with the objective of minimizing the required network-wide radio spectrum resource for a set of user sessions. Since the problem formulation is an MINLP, we developed a lower bound to estimate the objective function. Subsequently, we developed a novel sequential fixing algorithm to the cross-layer optimization problem. Simulation results showed that results obtained by this algorithm are very close to the lower bound, thus confirming that they are near-optimal.

Chapter 6

Optimal Power Control for CR Networks

6.1 Introduction

Cognitive radio (CR) is a revolution in radio technology and is enabled by recent advances in RF design, signal processing, and communication software [83]. Fundamental characteristics of CR are that transmitted waveforms are defined by software and that received waveforms are demodulated by software. This is in contrast to traditional hardware-based radios in which processing is done entirely by custom-made hardware circuitry. CR promises unprecedented flexibility in radio communications and is viewed as an enabling technology for dynamic spectrum access. Its capability has been recognized by the military and commercial sector and is now under intensive research and development by the DoD's Joint Tactical Radio System (JTRS) program [54] and wireless industry.

Since transmitted waveform is defined by software, a CR is capable of reconfiguring RF (on the fly) and switching to newly-selected frequency bands. From networking perspective, the emergence of CR offers new challenges in algorithm and protocol design. It is important to realize that a CR is vastly more powerful and flexible than the so-called *multi-channel multi-radio* (MC-MR) technology that has been actively researched in recent years (see e.g., [3, 31, 57, 58, 81] and refer-

ence therein). Note that MC-MR remains *hardware-based* radio technology: each radio can only operate on a single channel at a time and the number of concurrent channels that can be used at a wireless node is limited by the number of radio interfaces. In addition, an MC-MR based wireless network typically assumes there is a set of "common channels" available for all nodes in the network. This assumption is hardly true for CR networks since each node may have a different set of frequency bands based on its particular location. These important differences between MC-MR and CR warrant that the algorithm design for a CR network is substantially more complex than that under MC-MR. From algorithm design perspective, an MC-MR based wireless network can be viewed as a special case of a CR-based wireless network. Thus, algorithms designed for CR networks can be easily tailored to address MC-MR networks while the converse is not true.

In this chapter, we consider how to jointly perform power control, scheduling, and flow routing for a CR-based multi-hop wireless network. Specifically, we consider how to support a set of user communication sessions (each with a rate requirement) by jointly optimizing power control, scheduling, and flow routing such that the required radio resource and interference area in the network is minimized. We extend the current "protocol model" for interference modeling (more discussion in Section 6.2). Since power control directly affects a node's transmission range and interference range, we find that it has profound impact on scheduling feasibility, bandwidth efficiency, and problem complexity. We develop a formal mathematical model for scheduling feasibility under the influence of power control. Based on this model, the joint power control, scheduling, and flow routing problem is subsequently formulated as a *mixed integer nonlinear programming* (MINLP) problem. We develop a solution procedure based on the so-called *branch-and-bound* technique. Using numerical results, we validate this solution procedure and offer additional insights on the impact of power control and variable transmission/interference range on CR-based wireless networks.

The main contributions of this chapter are the following:

- By extending the existing protocol interference model, we have developed a formal mathematical model for scheduling feasibility under the influence of power control (constraints

(C-1'), (C-2'), and (C-3)), which takes into considerations of *variable* transmission and interference ranges. This new model generalizes existing deterministic transmission range and interference range model and can be applied for a broad range of problems where power control is part of the optimization space.

- We have formulated a joint power control, scheduling, and flow routing problem to support a set of user communication sessions in the network with the objective of minimizing radio resource requirement and interference area. Subsequently, we have developed an efficient solution procedure based on branch-and-bound technique and convex hull relaxation to solve this cross-layer optimization problem. The solution obtained via our approach yields a $(1-\varepsilon)$-optimal solution, where ε is a small error tolerance reflecting our desired accuracy in the final solution.

- By applying the solution procedure on sample random networks, we have demonstrated quantitatively that power control has significant impact on scheduling feasibility, bandwidth efficiency, and bandwidth-footprint product (BFP). This confirms the critical need of incorporating power control under the protocol interference model for future wireless network research.

The remainder of this chapter is organized as follows. In Section 6.2, we review related work on cross-layer wireless network research involving interference modeling, which is crucial in building tractable mathematical models and problem formulation. In Section 6.3, we examine the impact of power control on transmission range, interference range, and the necessary and sufficient condition for successful transmission. In Section 6.4, we develop a detailed mathematical model for power control, scheduling, and flow routing. In Section 6.5, we formulate the cross-layer optimization problem and Section 6.6 offers a solution procedure to this optimization problem. In Section 6.7, we use numerical results to validate the efficacy of the solution procedure and offer quantitative results to demonstrate the impact of power control. Section 6.8 concludes this chapter.

6.2 Related Work

Related work on MC-MR and its relationship with CR-based networks have been discussed in Section 6.1. In the rest of this section, we focus our review on cross-layer wireless network research involving interference modeling, which is crucial in building mathematical models and problem formulation.

Currently, there are two popular models for interference in wireless networks, namely, the *physical model* and the *protocol model.* Under the physical model (see, e.g., [8, 20, 23, 33, 44, 58]), a transmission is considered successful if and only if the signal to interference and noise ratio exceeds a certain threshold, where the interference includes all other concurrent transmissions. Since the calculation of a link's capacity involves not only the transmission power on this link, but also the transmission power on interference links, it is quite difficult to obtain an optimal solution whenever link capacity is involved. In [8], Behzad and Rubin found that for the special case when each node uses the same transmission power, then maximum transmission power should be used. However, for the general case where each node can adjust its transmission power independently, optimal solution is not available. In [33], Elbatt and Ephremides proposed a two-step approach with the aim of using a minimum power vector while supporting as many users as possible; routing is given a priori instead of being part of the optimization problem. In [20], Chen and Lee proposed a layered (de-coupled) approach for QoS scheduling and power control; while in [23], Cruz and Santhanam proposed a two-step approach to minimize a power cost function that first optimizes link scheduling and power control, and then optimizes routing. Due to de-coupling in the solution procedure, these approaches only yield sub-optimal solutions. In [44, 58], although the lower and upper bounds for transport capacity are obtained (asymptotically), the exact optimal value remains unknown.

Under the so-called protocol model [44], a transmission is considered successful if and only if the receiving node is in the transmission range of the corresponding transmitting node and is out of the interference range of all other transmitting nodes. Under the protocol model, related work can be classified into two categories. In the first category, e.g., [44, 58] (under the so-called

"random networks") and [3, 52, 57, 61, 79, 98], the transmission power is fixed. This corresponds to the special case of $Q = 1$ to be discussed later in this chapter. Since both the transmission range and interference range are fixed here, the interference relationship among the links are easy to identify. In the second category, i.e., [44,58] (under the so-called "arbitrary networks") and [12], power control is considered but only for wireless networks with certain special assumptions that may not reflect wireless networks considered in practice. For example, although power control is considered in [12], there is an assumption that hopping sequences (scheduling) for different link transmissions is already performed such that there is no interference between different links. In [44, 58], power control is considered for the so-called "arbitrary networks," within which the location for each node and its traffic pattern can be arbitrarily changed to optimize transport capacity. Although results for arbitrary networks do yield some theoretical insights, they are far from reflecting what happens for wireless networks in practice, which are mostly random networks [44].

In this chapter, we consider the case of power control under the protocol model for general wireless networks (without any special assumptions), i.e., random networks [44]. Such approach captures the most general characteristics of wireless network encountered in practice.

6.3 Power Control

In this section, we examine the impact of power control on scheduling feasibility, bandwidth efficiency, and problem complexity in a multi-hop CR network. But first, we need to have a clear understanding on the notion of transmission and interference ranges in a wireless network as well as the necessary and sufficient condition for successful transmission.

6.3.1 Transmission and Interference Ranges

We follow the protocol model as discussed in Section 6.2. In a wireless network, each node is associated with a transmission range and an interference range. Both transmission and interference

ranges directly depend on a node's transmission power and propagation gain. In this context, we assume a data transmission from node i to node j is successful only if the received power at node j exceeds a power threshold, say P_T. Suppose node i's transmission power is p and denote the transmission range of this node as $R_T(p)$. Then based on $g_{ij} \cdot p \geq P_T$ and (5.1), we can calculate the transmission range of this node as follows:

$$R_T(p) = \left(\frac{p}{P_T}\right)^{1/\alpha} . \tag{6.1}$$

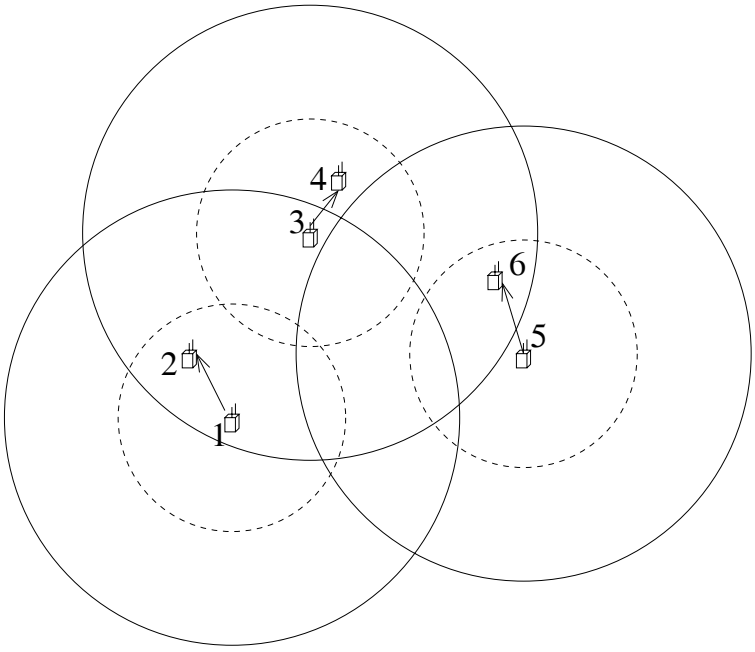
Similarly, we assume that an interference power is non-negligible only if it exceeds a threshold, say P_I at a receiver. Denote the interference range of a node as $R_I(p)$. Then following the same taken as the derivation for the transmission range, we can obtain the interference range of a node as follows:

$$R_I(p) = \left(\frac{p}{P_I}\right)^{1/\alpha} . \tag{6.2}$$

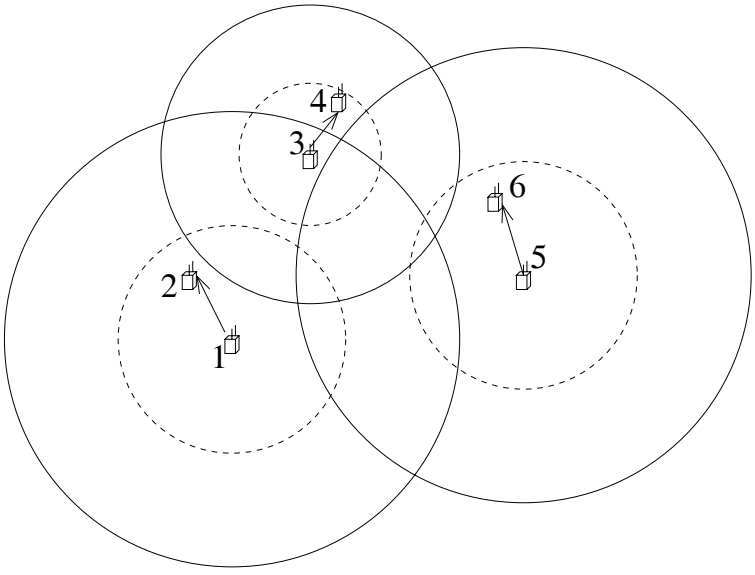
Note that the interference range is greater than the transmission range at a node, i.e., $R_I(p) > R_T(p)$ since $P_I < P_T$. As an example, in Fig. 6.1(a), we have three unicast communication links $1 \rightarrow 2, 3 \rightarrow 4$, and $5 \rightarrow 6$. For each transmitting node (1, 3, and 5), there are two circles centered around it, with the inner circle (dashed) representing transmission range and the outer circle (solid) representing interference range.

6.3.2 Necessary and Sufficient Condition for Successful Transmission

Scheduling for transmission at each node in the network can be done either in time domain or frequency domain. In this chapter, we consider scheduling in the frequency domain in the form of frequency bands. In a CR network, each node has a set of frequency bands that it may use for transmission and receiving. Suppose that band m is available at both node i and node j and denote p_{ij}^m the transmission power from node i to node j in frequency band m. Then to schedule a successful transmission from node i to node j, the following necessary and sufficient condition, expressed as two constraints, must be met. The first constraint (C-1) is that receiving node j must



(a) No power control.



(b) With power control.

Figure 6.1: A 3-link example network.

be physically within the transmission range of node i, i.e.,

(C-1) $$d_{ij} \leq R_T(p_{ij}^m) = \left(\frac{p_{ij}^m}{P_T}\right)^{1/\alpha} .$$

The second constraint (C-2) is that the receiving node j must not fall in the interference range of any other node k ($k \in \mathcal{N}$, $k \neq i$) that is transmitting in the same band, i.e.,

(C-2) $$d_{jk} \geq R_I(p_{kh}^m) = \left(\frac{p_{kh}^m}{P_I}\right)^{1/\alpha} .$$

where h is the intended receiving node of transmitting node k.

As an example, Fig. 6.1(a) shows a network with three links ($1 \rightarrow 2, 3 \rightarrow 4$, and $5 \rightarrow 6$). The transmission range and interference range for each transmitting node (1, 3, and 5) are shown in the figure. Clearly, each receiving node falls in the transmission range of its respective transmitting node. Further, we can see that both receiving nodes 2 and 6 fall in the interference range of node 3. Thus, when link $3 \rightarrow 4$ is using a frequency band m for transmission, link $1 \rightarrow 2$ and link $5 \rightarrow 6$ should not use the same band. It should also be noted that when link $3 \rightarrow 4$ is not using a frequency band m, both link $1 \rightarrow 2$ and link $5 \rightarrow 6$ may use band m. This is because that receiving node 2 does not fall in node 5's interference range and receiving node 6 does not fall in node 1's interference range.

6.3.3 Impact of Power Control

Most related work under the protocol model (see Section 6.2) assume fixed transmission power at each node. That is, for a given channel, the radio either does not transmit or transmit at its full power (say P_{max}). Such assumption clearly helps to simplify the problem since the transmission range and interference range are both deterministic in this case. As a result, efficient channel assignment algorithm or other optimization problems can be developed or solved.

In this chapter, we investigate the case that each node in a CR network has power control capability. That is, in addition to being able to transmit at full power P_{max}, a node can also transmit at any intermediate power between 0 and P_{max}. As one might expect, such freedom on power

control has significant impact on design space. In the rest of this section, we illustrate the impact of power control on scheduling feasibility, bandwidth efficiency, and problem complexity.

Scheduling Feasibility. Constraint (C-2) says that a receiving node j (corresponding to transmitting node i on band m) must not fall in the interference range of any other transmitting node k that is transmitting in the same band. In the absence of power control, the interference ranges for all the nodes in the network are fixed and there is hardly much we can do once we encounter an infeasibility scenario in scheduling. On the other hand, when power control capability is available, we could adjust the transmission power and thus reduce the interference range of some node so as to achieve scheduling feasibility.

We use an example to illustrate this important point. Recall that in Fig. 6.1(a), where there is no power control (and thus the transmission ranges and interference ranges for transmitting nodes 1, 3, and 5 are deterministic and shown in the figure), once link $3 \rightarrow 4$ is active (i.e., in transmission) in a frequency band m, then neither link $1 \rightarrow 2$ nor link $5 \rightarrow 6$ shall be active in the same frequency band. If frequency band m is the only band available for these links, then it is impossible to schedule transmissions on links $1 \rightarrow 2$ and $5 \rightarrow 6$. Now consider that each node has the power control capability and can adjust transmission power anywhere in $[0,\ P_{\text{max}}]$. In this setting, node 3 could reduce its transmission power while still maintaining data transmission to node 4 (see Fig. 6.1(b)). Now receiving nodes 2 and 6 are no longer in node 3's interference range. As a result of this, both transmitting nodes 1 and 5 can transmit in the same frequency band m simultaneously. In other words, power control could enable scheduling feasibility.

Bandwidth Efficiency. Power control capability has the additional benefit of reducing spectrum bandwidth requirement in the network. This is a direct consequence of the scheduling feasibility benefit discussed above.

Continuing the previous example, in Fig. 6.1(a), when there is no power control, one frequency band is inadequate to support all three links and we need at least two frequency bands in the network. However, when there is power control, in Fig. 6.1(b), we only need one frequency band to support all three links. In general, one should expect once power control is employed, the spectrum

Table 6.1: Notation in Chapter 6.

Symbol	Definition
c_{ij}^m	Link capacity of link $i \rightarrow j$ under p_{ij}^m
p_{ij}^m	The transmission power from node i to node j on band m
P_{max}	The maximum transmission power at a transmitter
P_T	The minimum threshold of power to decode a transmission at a receiver
P_I	The maximum threshold of power for interference to be negligible at a receiver
q_{ij}^m	The transmission power level from node i to node j on band m
$\mathbf{q}$	The vector of variables q_{ij}^m, $i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m$
Q	The number of transmission power levels at a transmitter
$R_T(p), R_I(p)$	The transmission and interference ranges under transmission power p
$R_T^{\text{max}}, R_I^{\text{max}}$	The maximum transmission and interference ranges (under full transmission power P_{max})
W	Bandwidth of a frequency band
x_{ij}^m	Binary indicator to mark whether or not band m is used by link $i \rightarrow j$
$\mathbf{x}$	The vector of variables x_{ij}^m for $i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m$
LB_z, UB_z	The lower and upper bounds of problem z
LB, UB	The minimum lower and upper bounds among all problems
ε	A small positive constant reflecting our desired accuracy in the final solution
Ω_z	The set of all possible values of $(\mathbf{x}, \mathbf{q})$ in problem z
ψ_z	The solution obtained by local search for problem z

bandwidth requirement can be reduced significantly for meeting the same communication objective than the case when there is no power control.

Problem Complexity. When power control is allowed, the transmission and interference ranges of a node can be adjusted since both of them are functions of this node's transmission power (see (6.1) and (6.2)). It is not hard to realize that the design space for algorithm or optimization solution procedure becomes substantially larger than the case when there is no power control. As a result, mathematical modeling, problem formulation, and solution procedure are also expected to be much more complex and interesting. In the remainder of this chapter, we explore this problem in the context of multi-hop CR-based ad hoc networks.

6.4 Mathematical Modeling

We consider an ad hoc network consisting of a set $\mathcal{N}$ of nodes. Unlike MC-MR networks where the set of available frequency bands at each node is identical, in a CR network, the set of available frequency bands at each node depends on its location and may not be the same. For example, at node i, its available frequency bands may consist of bands I, III, and V while at a different node j, its available frequency bands may consist of bands I, IV, and VI, and so forth. More formally, denote $\mathcal{M}_i$ the set of available frequency bands at node i. For simplicity, we assume the bandwidth of each frequency band is W.[1] Denote $\mathcal{M}$ the set of all frequency bands present in the network, i.e., $\mathcal{M} = \bigcup_{i\in\mathcal{N}} \mathcal{M}_i$. Table 6.1 lists all notation that are used only in this chapter.

6.4.1 Scheduling and Power Control

In Section 6.3, we discussed the necessary and sufficient condition for successful transmission and impact of power control on scheduling feasibility. In this section, we formalize these ideas into a mathematical model in the general context of multi-hop CR networks.

Suppose that band m is available at both node i and node j, i.e., $m \in \mathcal{M}_{ij}$, where $\mathcal{M}_{ij} = \mathcal{M}_i \bigcap \mathcal{M}_j$. Denote

$$x_{ij}^m = \begin{cases} 1 & \text{If node } i \text{ transmits data to node } j \text{ on frequency band } m, \\ 0 & \text{otherwise.} \end{cases}$$

As mentioned earlier, we consider scheduling in the frequency domain and thus once a band $m \in \mathcal{M}_i$ is used by node i for transmission to node j, this band cannot be used again by node i to transmit to a different node. That is,

(C-3) $$\sum_{j\in\mathcal{T}_i^m} x_{ij}^m \leq 1 \,,$$

where $\mathcal{T}_i^m$ is the set of nodes that are within the transmission range from node i under full power P_{max} on band m.

[1] The case of heterogeneous bandwidth for each frequency band can be easily extended.

Denote $R_T^{\max}$ the maximum transmission range of a node when it transmits at full power $P_{\max}$. Then based on (6.1), we have

$$R_T^{\max} = R_T(P_{\max}) = \left(\frac{P_{\max}}{P_T}\right)^{1/\alpha} .$$

Thus, we have

$$P_T = \frac{P_{\max}}{(R_T^{\max})^{\alpha}} .$$

Then for a node transmitting at a power $p \in [0,\ P_{\max}]$, its transmission range is

$$R_T(p) = \left(\frac{p}{P_T}\right)^{1/\alpha} = \left(\frac{p(R_T^{\max})^{\alpha}}{P_{\max}}\right)^{1/\alpha} = \left(\frac{p}{P_{\max}}\right)^{1/\alpha} R_T^{\max} . \tag{6.3}$$

Similarly, denote $R_I^{\max}$ the maximum interference range of a node when it transmits at full power $P_{\max}$. Then following the same token, we have

$$\begin{aligned} R_I^{\max} &= R_I(P_{\max}) = \left(\frac{P_{\max}}{P_I}\right)^{1/\alpha} \\ P_I &= \frac{P_{\max}}{(R_I^{\max})^{\alpha}} . \end{aligned}$$

For a node transmitting at a power $p \in [0,\ P_{\max}]$, its interference range is

$$R_I(p) = \left(\frac{p}{P_{\max}}\right)^{1/\alpha} R_I^{\max} . \tag{6.4}$$

Recall that $\mathcal{T}_i^m$ denotes the set of nodes that are within the transmission range from node i under full power $P_{\max}$ on band m. More formally, we have $\mathcal{T}_i^m = \{j : d_{ij} \leq R_T^{\max}, j \neq i, m \in \mathcal{M}_j\}$. Similarly, denote $\mathcal{I}_j^m$ the set of nodes that can make interference on node j on band m under full power $P_{\max}$, i.e., $\mathcal{I}_j^m = \{k : d_{jk} \leq R_I^{\max}, \mathcal{T}_k^m \neq \emptyset\}$. The physical meaning of $\mathcal{T}_k^m \neq \emptyset$ in the above definition for $\mathcal{I}_j^m$ is that the interfering node k can use band m for a valid transmission to a node in $\mathcal{T}_k^m$.

Note that the definitions of $\mathcal{T}_i^m$ and $\mathcal{I}_j^m$ are both based on full transmission power $P_{\max}$. When power level p is below $P_{\max}$, the corresponding transmission and interference ranges will be smaller. As a result, it is necessary to keep track of the set of nodes fall in the transmission

range and the set of nodes that can produce interference whenever transmission power changes at a node.

Recall the two constraints (C-1) and (C-2) for successful transmission from node i to node j and (6.3) and (6.4), respectively, we have

$$d_{ij} \leq R_T(p_{ij}^m) = \left(\frac{p_{ij}^m}{P_{\max}}\right)^{1/\alpha} R_T^{\max} \tag{6.5}$$

$$d_{jk} \geq R_I(p_{kh}^m) = \left(\frac{p_{kh}^m}{P_{\max}}\right)^{1/\alpha} R_I^{\max} \quad (k \in \mathcal{I}_j^m, k \neq i, h \in \mathcal{T}_k^m)\ . \tag{6.6}$$

Based on (6.5) and (6.6), we have the following requirements for the transmission link $i \to j$ and interfering link $k \to h$:

$$p_{ij}^m \begin{cases} \in \left[\left(\frac{d_{ij}}{R_T^{\max}}\right)^{\alpha} P_{\max},\ P_{\max}\right] & \text{If } x_{ij}^m = 1, \\ = 0 & \text{If } x_{ij}^m = 0. \end{cases}$$

$$p_{kh}^m \leq \begin{cases} \left(\frac{d_{kj}}{R_I^{\max}}\right)^{\alpha} P_{\max} & \text{If } x_{ij}^m = 1, \\ P_{\max} & \text{If } x_{ij}^m = 0. \end{cases} \quad (k \in \mathcal{I}_j^m, k \neq i, h \in \mathcal{T}_k^m)\ .$$

Mathematically, these requirements can be re-written as

(C-1') $$p_{ij}^m \in \left[\left(\frac{d_{ij}}{R_T^{\max}}\right)^{\alpha} P_{\max} x_{ij}^m,\ P_{\max} x_{ij}^m\right]$$

(C-2') $$p_{kh}^m \leq P_{\max} - \left(1 - \left(\frac{d_{kj}}{R_I^{\max}}\right)^{\alpha}\right) P_{\max} x_{ij}^m \quad (k \in \mathcal{I}_j^m, k \neq i, h \in \mathcal{T}_k^m)\ .$$

Recall that we consider scheduling in the frequency domain and in (C-3), we state that once a band m is used by node i for transmission to node j, this band cannot be used again by node i to transmit to a different node. In addition, for successful scheduling in frequency domain, the following two constraints must also hold.

(C-4) For a band $m \in \mathcal{M}_j$ that is available at node j, this band cannot be used for both transmission and receiving. That is, if band m is used at node j for transmission (or receiving), then it cannot be used for receiving (or transmission).

(C-5) Similar to constraint (C-3) on transmission, node j cannot use the same band $m \in \mathcal{M}_j$ for receiving from multiple nodes.

Note that (C-4) can be viewed as "self-interference" avoidance constraint where at the same node j, its transmission to another node h on band m interferes its reception from node i in the same band. It turns out that the above two constraints are mathematically *embedded* in (C-1') and (C-2'). That is, once (C-1') and (C-2') are satisfied, then both constraints (C-4) and (C-5) are also satisfied. This result is formally stated in the following lemma.

Lemma 8. *If transmission powers on every transmission link and interference link satisfy (C-1') and (C-2') in the network, then (C-4) and (C-5) are also satisfied.*

Proof. We first prove that (C-1') and (C-2') leads to (C-4).

- To show that (C-1') and (C-2') lead to (C-4), we first let $k = j$ in (C-2'). Then (C-2') degenerates into $p_{jh}^m \leq P_{\max} - P_{\max} x_{ij}^m$ since $d_{jj} = 0$. If node j is receiving from node i on band m, i.e., $x_{ij}^m = 1$, then $p_{jh}^m \leq P_{\max} - P_{\max} x_{ij}^m = 0$. Since $p_{jh}^m \geq \left(\frac{d_{jh}}{R_T^{\max}}\right)^n P_{\max} x_{jh}^m$ in (C-1'), we have that x_{jh}^m must be 0. That is, if node j is receiving from node i on band m, then node j cannot transmit to a node h in the same band.

- Now suppose node j is transmitting to node h on band m, i.e., $x_{jh}^m = 1$. We will show that this will lead to $x_{ij}^m = 0$. This can be proved by contradiction. That is, if $x_{ij}^m = 1$, then we have just proved in the above paragraph that $x_{jh}^m = 0$. But this contradicts our initial assumption that $x_{jh}^m = 1$. Therefore, x_{ij}^m must be 0. That is, if node j is transmitting to node h on band m, then node j cannot use the same band for receiving from a node i.

Combining the above two results, we have shown that (C-4) holds.

We now prove that (C-1') and (C-2') also leads to (C-5). Again the proof is based on contradiction. Suppose that (C-5) does not hold. Then node j can receive from two different nodes i and k on the same band m, i.e., $x_{ij}^m = 1$ and $x_{kj}^m = 1$. Note that link $k \to j$ can be viewed as an

interfering link to link $i \to j$. This corresponds to letting $h = j$ in (C-2'). Then from (C-2'), since $x_{ij}^m = 1$, we have

$$p_{kj}^m \leq \left(\frac{d_{kj}}{R_I^{\max}}\right)^{\alpha} P_{\max} \ .$$

On the other hand, by (C-1'), we have

$$p_{kj}^m \geq \left(\frac{d_{kj}}{R_T^{\max}}\right)^{\alpha} P_{\max} \ .$$

However, the above two inequalities cannot hold simultaneously (contradiction) since we have $R_I^{\max} > R_T^{\max}$. Thus, the initial assumption that (C-5) does not hold is incorrect and the proof is complete. □

The significance of Lemma 8 is that since (C-4) and (C-5) are embedded in (C-1') and (C-2'), it is sufficient to consider constraints (C-1'), (C-2'), and (C-3) for scheduling and power control.

6.4.2 Flow Routing

We consider an ad hoc network consisting of a set $\mathcal{N}$ of nodes. Among these nodes, there is a set $\mathcal{L}$ of active user communication (unicast) sessions. Denote $s(l)$ and $d(l)$ the source and destination nodes of session $l \in \mathcal{L}$ and $r(l)$ the rate requirement (in b/s) of session l. To route these flows from its respective source node to destination node, it is necessary to employ multi-hop due to limited transmission range of a node. Further, to have better load balancing and flexibility, it is desirable to employ multi-path (i.e., allow flow splitting) between a source node and its destination node. This is because a single path is overly restrictive and usually does not yield optimal solution. Then, flow balance constraints (5.5) and (5.6) must be satisfied.

In addition to the above flow balance equations at each node $i \in \mathcal{N}$ for session $l \in \mathcal{L}$, the aggregated flow rates on each radio link cannot exceed this link's capacity. Under p_{ij}^m, we have

$$\sum_{l \in \mathcal{L}}^{s(l) \neq j, d(l) \neq i} f_{ij}(l) \leq \sum_{m \in \mathcal{M}_{ij}} c_{ij}^m = \sum_{m \in \mathcal{M}_{ij}} W \log_2 \left(1 + \frac{g_{ij}}{\eta W} p_{ij}^m\right) \ , \tag{6.7}$$

where η is the ambient Gaussian noise density. Note that the denominator inside the log function contains only ηW. This is due to the use of protocol interference modeling, i.e., when node i is transmitting to node j on band m, then the interference range of all other nodes in this band should not contain node j. Thus, protocol modeling significantly helps to simplify the calculation of link capacity c_{ij}^m.

6.5 Problem Formulation

Objective Function. In the last section, we presented a mathematical model for constraints on scheduling, power control, and flow routing. Under these constraints, a number of objective functions can be considered for problem formulation. A commonly used objective is to maximize network capacity, which can be expressed as maximizing a scaling factor for all the rate requirements of the communication sessions in the network (see, e.g. [3, 57]). In this chapter, we consider another objective called "bandwidth-footprint-product" (BFP), which better characterizes the spectrum and space occupancy for a CR network. The BFP was first introduced by Liu and Wang in [61]. The so-called "footprint" refers the interference area of a node under a given transmission power, i.e., $\pi \cdot (R_I(p))^2$. Since each node in the network will use a number of bands for transmission and each band will have a certain footprint corresponding to its transmission power, an important objective is to minimize network-wide BFP, which is the sum of BFPs among all the nodes in the network. That is, our objective is to minimize

$$\sum_{i \in \mathcal{N}} \sum_{m \in \mathcal{M}_i} \sum_{j \in \mathcal{T}_i^m} W \cdot \pi (R_I(p_{ij}^m))^2 \ ,$$

which is equal to $\pi (R_I^{\max})^2 \sum_{i \in \mathcal{N}} \sum_{m \in \mathcal{M}_i} \sum_{j \in \mathcal{T}_i^m} W \left(\frac{p_{ij}^m}{P_{\max}} \right)^{2/n}$. Since $\pi (R_I^{\max})^2$ is a constant factor, we can remove it from the objective function.

Discretization of Transmission Powers. For power control, we allow the transmission power to be adjusted between 0 and $P_{\max}$. In practice, the transmission power can only be tuned into a finite number of discrete levels between 0 and $P_{\max}$. To model this discrete version of power control,

we introduce an integer parameter Q that represents the total number of power levels to which a transmitter can be adjusted, i.e., $0, \frac{1}{Q}P_{\max}, \frac{2}{Q}P_{\max}, \cdots, P_{\max}$. Denote $q_{ij}^m \in \{0, 1, 2, \cdots, Q\}$ the integer power level for p_{ij}^m, i.e., $p_{ij}^m = \frac{q_{ij}^m}{Q}P_{\max}$. Then (C-1'), (C-2'), and (6.7) can be re-written as follows:

$$q_{ij}^m \in \left[\left(\frac{d_{ij}}{R_T^{\max}}\right)^{\alpha} Qx_{ij}^m,\ Qx_{ij}^m\right] \tag{6.8}$$

$$q_{kh}^m \le Q - \left(1 - \left(\frac{d_{kj}}{R_I^{\max}}\right)^{\alpha}\right) Qx_{ij}^m \qquad (k \in \mathcal{I}_j^m, k \ne i, h \in \mathcal{T}_k^m) \tag{6.9}$$

$$\sum_{l\in\mathcal{L}}^{s(l)\ne j, d(l)\ne i} f_{ij}(l) \le \sum_{m\in\mathcal{M}_{ij}} W \log_2\left(1 + \frac{g_{ij}P_{\max}}{\eta WQ} q_{ij}^m\right)$$

Problem Formulation. Putting together the objective function and all the constraints for scheduling, power control, and flow routing, we have the following formulation:

$$\begin{aligned}
\text{Min} \quad & \sum_{i\in\mathcal{N}} \sum_{m\in\mathcal{M}_i} \sum_{j\in\mathcal{T}_i^m} W \left(\frac{q_{ij}^m}{Q}\right)^{2/\alpha} \\
\text{s.t.} \quad & \sum_{j\in\mathcal{T}_i^m} x_{ij}^m \le 1 && (i \in \mathcal{N}, m \in \mathcal{M}_i) \\
& q_{ij}^m - \left(\frac{d_{ij}}{R_T^{\max}}\right)^{\alpha} Qx_{ij}^m \ge 0 && (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \qquad (6.10) \\
& q_{ij}^m - Qx_{ij}^m \le 0 && (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \qquad (6.11) \\
& \sum_{h\in\mathcal{T}_k^m} q_{kh}^m + \left(1 - \left(\frac{d_{kj}}{R_I^{\max}}\right)^{\alpha}\right) Qx_{ij}^m \le Q && (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m, \\
& && k \in \mathcal{I}_j^m, k \ne i) \qquad (6.12)
\end{aligned}$$

$$\sum_{l\in\mathcal{L}}^{s(l)\ne j, d(l)\ne i} f_{ij}(l) - \sum_{m\in\mathcal{M}_{ij}} W \log_2\left(1 + \frac{g_{ij}P_{\max}}{\eta WQ} q_{ij}^m\right) \le 0 \quad (i \in \mathcal{N}, j \in \mathcal{T}_i)$$

$$\begin{aligned}
\sum_{j\in\mathcal{T}_i} f_{ij}(l) &= r(l) && (l \in \mathcal{L}, i = s(l)) \\
\sum_{j\in\mathcal{T}_i}^{j\ne s(l)} f_{ij}(l) - \sum_{k\in\mathcal{T}_i}^{k\ne d(l)} f_{ki}(l) &= 0 && (l \in \mathcal{L}, i \in \mathcal{N}, i \ne s(l), d(l)) \\
x_{ij}^m \in \{0,1\}, q_{ij}^m &\in \{0, 1, 2, \cdots, Q\} && (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \\
f_{ij}(l) &\ge 0 && (l \in \mathcal{L}, i \in \mathcal{N}, i \ne d(l), j \in \mathcal{T}_i, j \ne s(l)) \ ,
\end{aligned}$$

where $W, g_{ij}, R_T^{\max}, R_I^{\max}, P_{\max}, \eta, r(l)$, and Q are all constants and x_{ij}^m, q_{ij}^m, and $f_{ij}(l)$ are all optimization variables. In this formulation, (6.10) and (6.11) come from (6.8). Inequality (6.12) is

based on (6.9) by noting that in (C-3), there is at most one $h \in \mathcal{T}_k^m$ such that $x_{kh}^m = 1$. As a result, based on (6.11), there is at most one $h \in \mathcal{T}_k^m$ such that $q_{kh}^m > 0$. Thus, (6.9) can be rewritten as

$$\sum_{h \in \mathcal{T}_k^m} q_{kh}^m \leq Q - \left(1 - \left(\frac{d_{kj}}{R_I^{\max}}\right)^{\alpha}\right) Q x_{ij}^m \quad (k \in \mathcal{I}_j^m, k \neq i) \,,$$

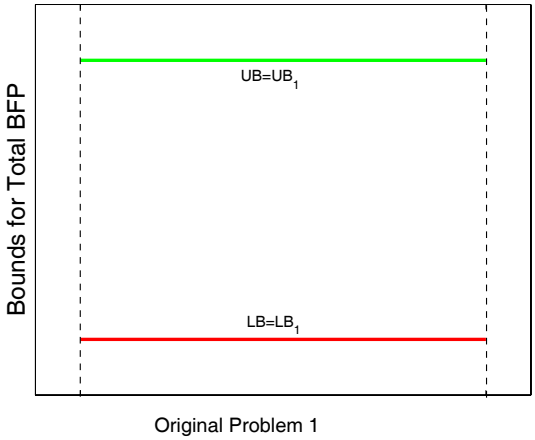
which is equivalent to (6.12).

This optimization problem is in the form of *mixed-integer nonlinear programming* (MINLP) problem, which is NP-hard in general [40]. In the next section, we develop a solution procedure based on the branch-and-bound approach [71] and the convex hull relaxation to solve this MINLP optimization problem.

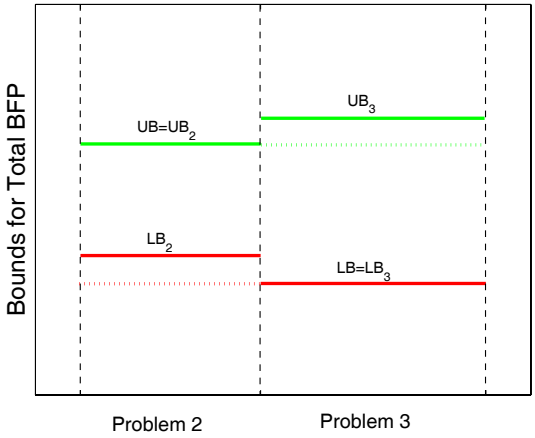
6.6 A Solution Procedure

6.6.1 Overview

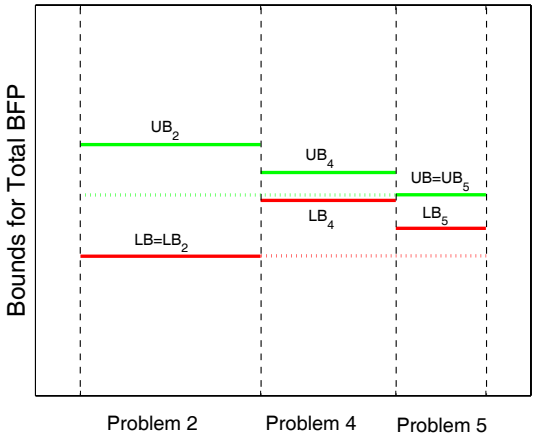
We find that the so-called *branch-and-bound* approach [71] is most effective in solving our optimization problem. Under this approach, we aim to provide a $(1-\varepsilon)$-optimal solution, where ε is a small positive constant reflecting our desired accuracy in the final solution. Initially, branch-and-bound analyzes partition variables, i.e., all discrete variables and all variables in a nonlinear term. For our problem, these variables include all x_{ij}^m's and q_{ij}^m's. Then branch-and-bound determines the value set for each partition variable. That is, we have $x_{ij}^m \in \{0, 1\}$ and $q_{ij}^m \in \{0, 1, 2, \cdots, Q\}$. By using some *relaxation* technique, branch-and-bound obtains a linear relaxation for the original MINLP problem and its solution provides a *lower bound* (LB) to the objective function. As we shall show shortly, this critical step is made possible by the convex hull relaxation for nonlinear discrete terms. With the relaxation solution as a starting point, branch-and-bound uses a *local search* algorithm to find a feasible solution to the original problem, which provides an upper bound (UB) for the objective function (see Fig. 6.2(a) for an example). If the obtained lower and upper bounds are within ε to each other, i.e., $LB \geq (1-\varepsilon)UB$, we are done.



(a) Iteration 1.



(b) Iteration 2.



(c) Iteration 3.

Figure 6.2: Illustration of branch-and-bound solution procedure.

Branch-and-bound Procedure

1. Initialization:
2. Let the initial best solution $\psi_\varepsilon = \emptyset$ and the initial upper bound $UB = \infty$.
3. Determine initial value set for each partition variable.
4. Let the initial problem list include only the original problem, denoted as problem 1.
5. Build a linear relaxation and obtain the relaxation solution $\hat{\psi}_1$.
6. The objective value of $\hat{\psi}_1$ is a lower bound LB_1 to problem 1.
7. Iteration:
8. Suppose problem z has the minimum LB_z among all problems in the problem list.
9. Update lower bound $LB = LB_z$.
10. Find a feasible solution ψ_z from $\hat{\psi}_z$ via a local search algorithm and denote its objective value
11. as UB_z.
12. If $(UB_z < UB)$ {
13. Update $\psi_\varepsilon = \psi_z$ and $UB = UB_z$.
14. If $LB \geq (1-\varepsilon)UB$, we stop with the $(1-\varepsilon)$-optimal solution ψ_ε.
15. Otherwise, remove all problems z' with $LB_{z'} \geq (1-\varepsilon)UB$ from the problem list. }
16. Select a variable with the maximum relaxation error and divide its value set into two sets by
17. its value in $\hat{\psi}_z$.
18. Build two new problems z_1 and z_2 based on these two sets.
19. Remove problem z from the problem list.
20. Obtain LB_{z1} and LB_{z2} for problems z_1 and z_2 via their linear relaxations.
21. If $LB_{z1} \leq (1-\varepsilon)UB$, add problem z_1 into the problem list.
22. If $LB_{z2} \leq (1-\varepsilon)UB$, add problem z_2 into the problem list.
23. If the problem list is empty, we stop with the $(1-\varepsilon)$-optimal solution ψ_ε.
24. Otherwise, go to the next iteration.

Figure 6.3: A general framework of the branch-and-bound solution procedure.

If the relaxation errors for nonlinear terms are not small, then the upper bound UB could be far away from the lower bound LB. To close this gap, we must have a tighter linear relaxation, i.e., with smaller relaxation errors. This could be achieved by further narrowing down the value sets of partition variables. Specifically, branch-and-bound selects a partition variable with the maximum relaxation error and divides its value set into two sets by its value in the relaxation solution (see Fig. 6.2(b)). Then the original problem (denoted as problem 1) is divided into two new problems (denoted as problem 2 and problem 3). Again, branch-and-bound performs relaxation and local search on these two new problems. Now we have lower bounds LB_2 and LB_3 for problems 2 and 3, respectively. We also have feasible solutions that provide upper bounds UB_2 and UB_3 for problems 2 and 3, respectively. Since the relaxations in problems 2 and 3 are both tighter than that in problem 1, we have $\min\{UB_2, UB_3\} \leq UB_1$ and $LB_2, LB_3 \geq LB_1$. For a minimization problem, the lower bound of the original problem is updated from $LB = LB_1$ to $LB = \min\{LB_2, LB_3\}$. Also, the best feasible solution to the original problem is the solution with a smaller UB_i. Then the upper bound of the original problem is updated from $UB = UB_1$ to $UB = \min\{UB_2, UB_3\}$. As a result, we now have smaller gap between UB and LB. If $LB \geq (1-\varepsilon)UB$, we are done. Otherwise, we choose a problem with the minimum lower bound (Problem 3 in Fig. 6.2(b)) and perform partition for this problem.

Note that during the iteration process for branch-and-bound, if we find a problem z with $LB_z \geq (1-\varepsilon)UB$, then we can conclude that this problem cannot provide much improvement on UB (see problem 4 in Fig. 6.2(c)). That is, further branch on this problem will not yield much improvement and we can thus remove this problem from further consideration. Eventually, once we find $LB \geq (1-\varepsilon)UB$, branch-and-bound procedure terminates. It has been shown that under very general conditions, a branch-and-bound solution procedure always converges [89]. Moreover, although the worst-case complexity of such a procedure is exponential, the actual running time could be fast if all partition variables are integer variables (e.g., the problem considered in this chapter).

Figure 6.3 shows the general framework of branch-and-bound procedure. Several key components in the branch-and-bound procedure are problem specific and remain to be solved. These include

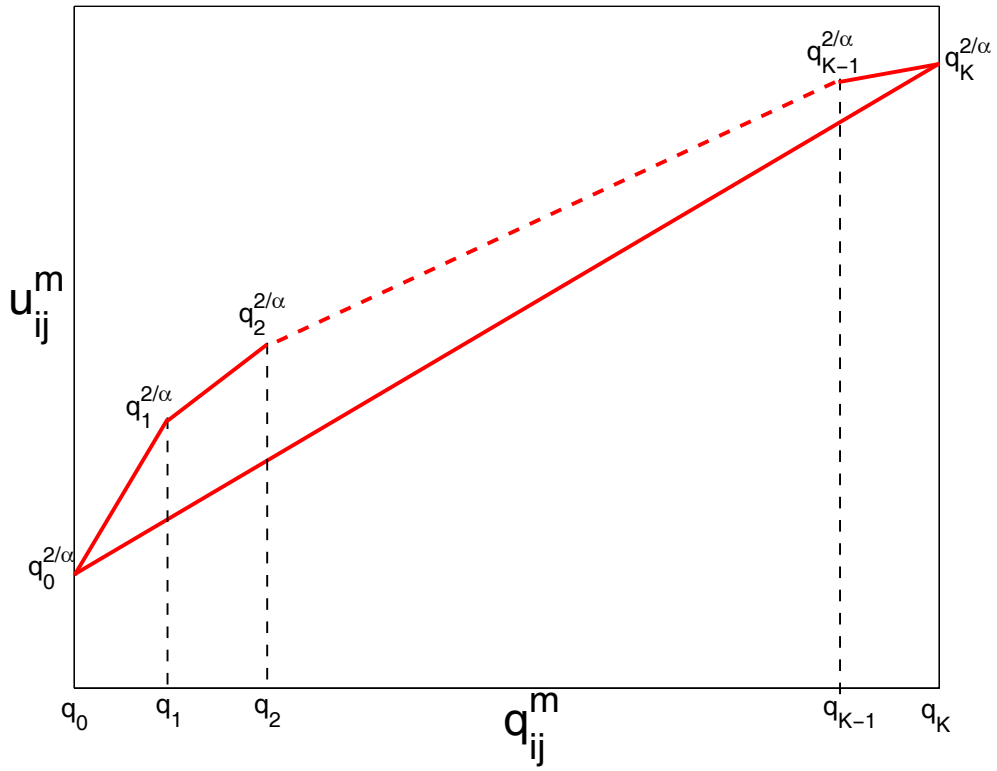


Figure 6.4: Illustration of convex hull for a discrete term.

- How to obtain a linear relaxation?
- How to perform local search?
- How to choose a variable for partition?

In the rest of this section, we develop these important components in the branch-and-bound solution procedure.

6.6.2 Linear Relaxation

During each iteration of the branch-and-bound procedure, we need a relaxation technique to obtain a lower bound of the objective function (see line 5 in Fig. 6.3). For a nonlinear discrete term, we propose to use a convex hull relaxation. That is, we introduce a new variable u_{ij}^m for nonlinear

discrete term $(q_{ij}^m)^{2/\alpha}$. Suppose $q_{ij}^m \in \{q_0, q_1, \cdots, q_K\}$, where $q_0 = (q_{ij}^m)_L \leq q_1 \leq \cdots \leq q_K = (q_{ij}^m)_U$. The convex hull (see Fig. 6.4) can be formulated as

$$\begin{aligned} u_{ij}^m - \tfrac{(q_K)^{2/\alpha}-(q_0)^{2/\alpha}}{q_K-q_0} q_{ij}^m &\geq \tfrac{q_K(q_0)^{2/\alpha}-(q_K)^{2/\alpha}q_0}{q_K-q_0} \\ u_{ij}^m - \tfrac{(q_k)^{2/\alpha}-(q_{k-1})^{2/\alpha}}{q_k-q_{k-1}} q_{ij}^m &\leq \tfrac{q_k(q_{k-1})^{2/\alpha}-(q_k)^{2/\alpha}q_{k-1}}{q_k-q_{k-1}} \quad (1 \leq k \leq K) \, . \end{aligned}$$

Similarly, we can introduce a new variable v_{ij}^m for nonlinear discrete term $\log_2\left(1 + \frac{g_{ij}P^{\max}}{\eta W Q} q_{ij}^m\right)$ and construct convex hull constraints for v_{ij}^m.

Denote $\mathbf{x}$ and $\mathbf{q}$ as the vector for variables x_{ij}^m and q_{ij}^m, respectively. We have the following linear relaxation for problem z.

$$\begin{array}{lll} \text{Min} & \sum_{i\in\mathcal{N}} \sum_{m\in\mathcal{M}_i} \sum_{j\in\mathcal{T}_i^m} \frac{W}{Q^{2/\alpha}} u_{ij}^m & \\ \text{s.t.} & \text{Convex hull constraints for } u_{ij}^m & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \\ & \sum_{j\in\mathcal{T}_i^m} x_{ij}^m \leq 1 & (i \in \mathcal{N}, m \in \mathcal{M}_i) \\ & q_{ij}^m - \left(\frac{d_{ij}}{R_T^{\max}}\right)^{\alpha} Q x_{ij}^m \geq 0 & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \\ & q_{ij}^m - Q x_{ij}^m \leq 0 & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \\ & \sum_{h\in\mathcal{T}_k^m} q_{kh}^m + \left(1 - \left(\frac{d_{kj}}{R_I^{\max}}\right)^{\alpha}\right) Q x_{ij}^m \leq Q & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m, k \in \mathcal{I}_j^m, k \neq i) \\ & \sum_{l\in\mathcal{L}}^{s(l)\neq j, d(l)\neq i} f_{ij}(l) - \sum_{m\in\mathcal{M}_{ij}} W v_{ij}^m \leq 0 & (i \in \mathcal{N}, j \in \mathcal{T}_i) \\ & \text{Convex hull constraints for } v_{ij}^m & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \\ & \sum_{j\in\mathcal{T}_i} f_{ij}(l) = r(l) & (l \in \mathcal{L}, i = s(l)) \\ & \sum_{j\in\mathcal{T}_i}^{j\neq s(l)} f_{ij}(l) - \sum_{k\in\mathcal{T}_i}^{k\neq d(l)} f_{ki}(l) = 0 & (l \in \mathcal{L}, i \in \mathcal{N}, i \neq s(l), d(l)) \\ & u_{ij}^m, v_{ij}^m \geq 0 & (i \in \mathcal{N}, m \in \mathcal{M}_i, j \in \mathcal{T}_i^m) \\ & f_{ij}(l) \geq 0 & (l \in \mathcal{L}, i \in \mathcal{N}, i \neq d(l), j \in \mathcal{T}_i, j \neq s(l)) \\ & (\mathbf{x}, \mathbf{q}) \in \Omega_z \ , & \end{array}$$

where Ω_z is the set of all possible values of $(\mathbf{x}, \mathbf{q})$ in problem z. For example, Ω_1 for the original problem 1 is $\{(\mathbf{x}, \mathbf{q}) : 0 \leq x_{ij}^m \leq 1, 0 \leq q_{ij}^m \leq Q\}$.

6.6.3 Local Search Algorithm

A linear relaxation for a problem z can be solved in polynomial time. Denote the relaxation solution as $\hat{\psi}_z$, which provides a lower bound to problem z, although may not be feasible. We now show how to obtain a feasible solution ψ_z based on $\hat{\psi}_z$ (see line 10 in Fig. 6.3).

We use the same routing solution, i.e., let $\mathbf{f} = \hat{\mathbf{f}}$. To obtain a feasible solution, we need to determine the value for $\mathbf{x}$ and $\mathbf{q}$ in solution ψ_z such that the routing solution $\mathbf{f}$ is feasible, i.e., (6.7) should hold for each link $i \to j$.

Initially, each q_{ij}^m is set as the smallest value $(q_{ij}^m)_L$ in its value set and x_{ij}^m is fixed as 0 or 1 if its value set only has one element 0 or 1, respectively. Based on these q_{ij}^m's, we can compute the capacity $\sum_{m\in\mathcal{M}_i} W \log_2\left(1+\frac{g_{ij}P\max}{\eta WQ}q_{ij}^m\right)$ for each link $i \to j$. The requirement on a link $i \to j$ is $\sum_{l\in\mathcal{L}}^{s(l)\neq j, d(l)\neq i} f_{ij}(l)$. For each link with a requirement larger than its capacity, we will try to satisfy (6.7) by increasing q_{ij}^m under its value set limitation. After we do this for all links, then we can calculate the objective value of solution ψ_z. Otherwise, if the requirement of any link cannot be satisfied, then we fail to find a feasible solution and set the objective value as ∞. The details of this local search algorithm is shown in Fig. 6.5.

6.6.4 Selection of Partition Variables

If the relaxation error for a problem z is not small, the gap between its lower and upper bounds may be large. To narrow the gap, we generate two new sub-problems z_1 and z_2 from problem z. We hope that these two new problems have smaller relaxation errors. Subsequently, they may have tighter bounds than bounds for problem z. Thus, we identify a variable based on its relaxation error (line 16, Fig. 6.3).

Note that $\mathbf{x}$ variables are for scheduling while $\mathbf{q}$ variables are for power control. It is better to determine whether to use a band or not before we determine the transmission power on this band. Thus, we first select one of $\mathbf{x}$ variables. In particular, for the relaxation solution $\hat{\psi}_z$, the relaxation

Local Search Algorithm

1. Initialization:
2. Set x_{ij}^m as 0 or 1 if its value set only have one element 0 or 1, respectively.
3. Set $q_{ij}^m = (q_{ij}^m)_L$.
4. For each link $i \to j$, compute the capacity $\sum_{m \in \mathcal{M}_i} W \log_2 \left(1 + \frac{g_{ij} P_{\max}}{\eta W Q} q_{ij}^m\right)$ and the
5. requirement $\sum_{l \in \mathcal{L}}^{s(l) \neq j, d(l) \neq i} f_{ij}(l)$.
6. Iteration:
7. Among all links that requirement is larger than capacity, choose a link $i \to j$ with the
8. maximum requirement.
9. If there is no such link, we are done. Otherwise, we will try to increase its capacity as follows.
10. We first increase q_{ij}^m on a used band m in the non-increasing order of $\hat{q}_{ij}^m$ and under the
11. limitations that $q_{ij}^m \leq \lceil \hat{q}_{ij}^m \rceil$ and $q_{ij}^m \leq (q_{ij}^m)_U$.
12. If the achieved capacity is not enough, we then try to find and use an available but currently
13. unused band m in the non-increasing order of $\hat{q}_{ij}^m$.
14. A band m is available if for any node k using this band, node j is not in its interference range.
15. We increase q_{ij}^m under the limitations that $q_{ij}^m \leq \lceil \hat{q}_{ij}^m \rceil$ and $q_{ij}^m \leq (q_{ij}^m)_U$.
16. We also need to set $x_{ij}^m = 1$, $x_{ih}^m = 0$ for $h \in \mathcal{T}_i, h \neq j$, and let $(q_{kh}^m)_U \leq \left\lfloor \left(\frac{d_{kj}}{R_I^{\max}}\right)^{\alpha} Q \right\rfloor$ for
17. $k \in \mathcal{I}_j^m, k \neq i, h \in \mathcal{T}_k^m$.
18. If the achieved capacity is still not enough, we finally increase q_{ij}^m on a used band m in the
19. non-increasing order of $\hat{q}_{ij}^m$ and under the limitation that $q_{ij}^m \leq (q_{ij}^m)_U$.
20. If the achieved capacity is still not enough, we claim that link $i \to j$ cannot be satisfied.

Figure 6.5: A local search algorithm.

error of a discrete variable x_{ij}^m is $\min\{\hat{x}_{ij}^m, 1-\hat{x}_{ij}^m\}$. We choose an x_{ij}^m with the maximum relaxation error among all $\mathbf{x}$ variables and let its value set in problems z_1 and z_2 be $\{0\}$ and $\{1\}$, respectively.[2] It should be note that the new value set of x_{ij}^m may impose constraints on other variables. That is, if the new value set of x_{ij}^m is $\{0\}$, then we have $q_{ij}^m = 0$ based on (6.11). If the new value set of x_{ij}^m is $\{1\}$, then we have $x_{ih}^m = 0$ for $h \in \mathcal{T}_i, h \neq j$ based on (C-3), $q_{ij}^m \geq \left(\frac{d_{ij}}{R_T^{\max}}\right)^\alpha Q$ based on (6.10), and $q_{kh}^m \leq \left(\frac{d_{kj}}{R_I^{\max}}\right)^\alpha Q$ for $k \in \mathcal{I}_j^m, k \neq i, h \in \mathcal{T}_k$ based on (6.12).[3]

If all $\mathbf{x}$ variables are already selected, then we select one of $\mathbf{q}$ variables. In particular, for the relaxation solution $\hat{\psi}_z$, the relaxation error of a discrete variable q_{ij}^m is $\min\{\hat{q}_{ij}^m - \lfloor \hat{q}_{ij}^m \rfloor, \lfloor \hat{q}_{ij}^m \rfloor + 1 - \hat{q}_{ij}^m\}$; the relaxation error of a nonlinear discrete term $u_{ij}^m = (q_{ij}^m)^{2/\alpha}$ is $|\hat{u}_{ij}^m - (\hat{q}_{ij}^m)^{2/\alpha}|$; and the relaxation error of a nonlinear discrete term $v_{ij}^m = \log_2\left(1 + \frac{q_{ij}P\max}{\eta WQ} q_{ij}^m\right)$ is $\left|\hat{v}_{ij}^m - \log_2\left(1 + \frac{q_{ij}P\max}{\eta WQ}\hat{q}_{ij}^m\right)\right|$. If the maximum relaxation error is related to a discrete variable q_{ij}^m or a nonlinear discrete term u_{ij}^m or v_{ij}^m, we choose q_{ij}^m as the branch variable. Assuming the value set of q_{ij}^m in problem z is $\{q_0, q_1, \cdots, q_K\}$, its value set in problems z_1 and z_2 will be $\{q_0, q_1, \cdots, \lfloor \hat{q}_{ij}^m \rfloor\}$ and $\{\lfloor \hat{q}_{ij}^m \rfloor + 1, \lfloor \hat{q}_{ij}^m \rfloor + 2, \cdots, q_K\}$, respectively. Again, the new value set of q_{ij}^m may have limitations on other variables. In particular, if the new value set of q_{ij}^m is $\{0\}$, then we have $x_{ij}^m = 0$ based on (6.10). If the new value set of q_{ij}^m does not include 0, then we have $x_{ij}^m = 1$ based on (6.11).

6.7 Numerical Results

In this section, we present some important numerical results. The purpose of this effort is to validate the efficacy of the solution procedure and to offer additional insights on power control that are not obvious from our theoretical development.

For ease of exposition, we normalize all units for distance, bandwidth, rate, and power based on (5.1) and (6.7) with appropriate dimensions. We consider a 20-node ad hoc network with each

[2]Since the value set for this x_{ij}^m only has one element, this x_{ij}^m can be replaced by a constant in the new problem. As a result, some constraints may also be removed.

[3]If the constraints make the new value set of a variable q_{kh}^m empty, then the corresponding new problem is clearly infeasible. As a result, we can remove it from the problem list.

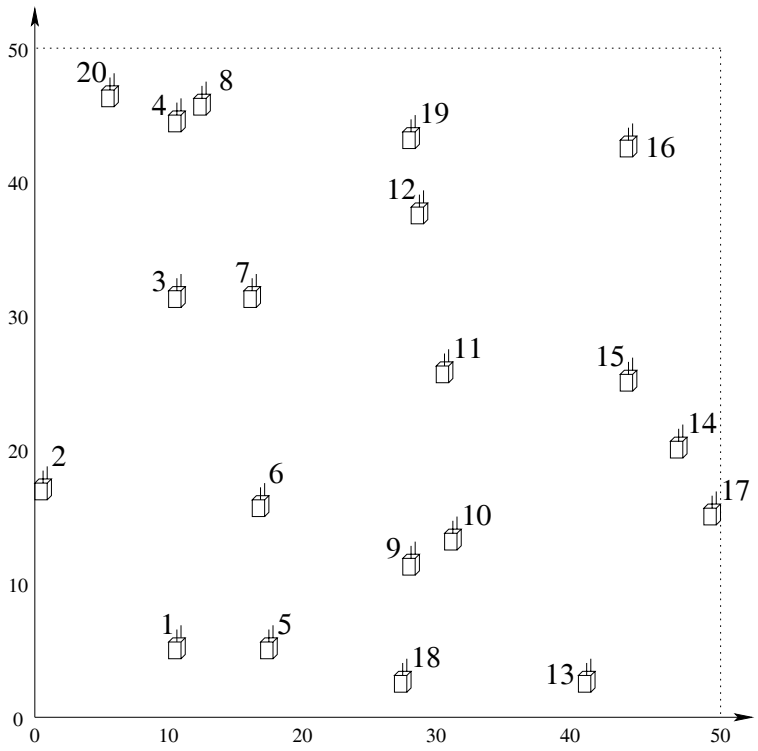


Figure 6.6: A 20-node ad hoc network.

node randomly located in a 50x50 area. An instance of network topology is given in Fig. 6.6 with each node's location listed in Table 6.2. We assume there are $|\mathcal{M}| = 10$ frequency bands in the network and each band has a bandwidth of $W = 50$. Each node may only have a subset of these frequency bands. In the simulation, this is done by randomly selecting a subset of bands for each node from the pool of 10 bands. Table 6.2 shows the available bands for each node.

We assume that, under the maximum transmission power, the transmission range on each node is 20 and the interference range is twice of the interference range (i.e., 40). Both transmission range and interference range will be smaller when transmission power is less than maximum. The pass loss index α is assumed to be 4. The threshold P_T is assumed to be $\eta W = 50\eta$. Thus, we have $P_I = \left(\frac{R_T^{\max}}{R_I^{\max}}\right)^{\alpha} P_T W = \frac{50}{16}\eta$ and the maximum transmission power $P_{\max} = (R_T^{\max})^{\alpha} P_T W = 8 \cdot 10^6 \eta$.

Within the network, we assume there are $|\mathcal{L}| = 5$ user communication sessions, with source

Table 6.2: Each node's location and available frequency bands for the 20-node network.

Node Index	Location	Available Bands
1	$(10.5, 4.3)$	I, II, III, IV, V, VI, VII, VIII, IX, X
2	$(1.7, 17.3)$	II, III, IV, V, VI, VII, X
3	$(10.7, 30.8)$	I, III, IV, V, VI, VII, VIII, IX, X
4	$(10.2, 45.3)$	I, III, IV, V, VI, VII, VIII, IX, X
5	$(17.8, 4)$	I, II, V, VI, VII, VIII, IX
6	$(17.2, 15.2)$	I, II, IV, VIII
7	$(16.9, 30.8)$	I, II, III, IV, V, VI, VII, VIII, IX, X
8	$(12.3, 47.3)$	I, III, IV, V, VII, VIII, IX
9	$(28.2, 11.5)$	I, III, V, VII
10	$(32.1, 13.8)$	I, II, III, IV, VI, VII, VIII, IX, X
11	$(30.4, 25.6)$	I, II, III, V, VI, VIII, IX, X
12	$(29.7, 36)$	I, II, III, IV, VI, VI
13	$(41.7, 3.1)$	I, II, III, V, VI, VIII, IX, X
14	$(41.7, 3.1)$	I, IV, V, VIII, IX, X
15	$(43.3, 25.3)$	II, III, IV, V, VI, VII, VIII, IX, X
16	$(44.1, 42.7)$	I, II, IV, VI, VII, VIII, IX, X
17	$(49.6, 15.8)$	I, II, III, IV, V, VI, VII, VIII
18	$(28.7, 2.5)$	I, II, III, VI, VII, VIII, IX, X
19	$(28, 43.5)$	II, IV, V, VI, VIII
20	$(5, 46.9)$	II, IV, V, VI, VII

Table 6.3: Source node, destination node, and rate requirement of the 5 sessions.

Source Node	Destination Node	Rate Requirement
7	16	28
8	5	12
15	13	56
2	18	75
9	11	29

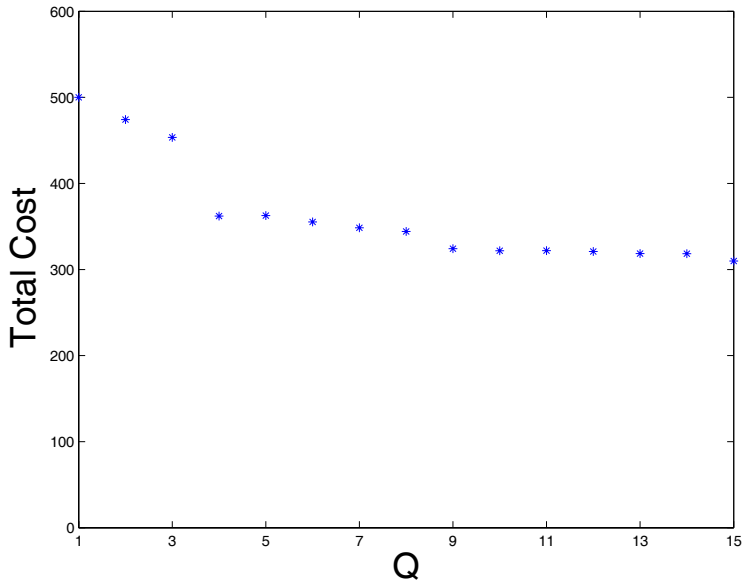


Figure 6.7: Total cost as a function of discretization levels of power control.

node and destination node randomly selected and the rate of each session is randomly generated within $[10, 100]$. Table 6.3 specifies an instance of the source node, destination node, and rate requirement for the 5 sessions in the network.

We set the desired approximation error bound ε to be 0.05 in the solution procedure (Section 6.6), which guarantees that the obtained solution is within 5% from the optimum.

6.7.1 Power Control and Level of Granularity on BFP

In this set of numerical results, we apply the solution procedure to the 20-node network described above for different level of power control granularity (Q). Note that $Q = 1$ corresponds to the case that there is no power control, i.e., a node uses its peak power P_{max} for transmission. When Q is sufficiently large, the discrete nature of power control diminishes and power control becomes continuous between $[0,\ P_{\text{max}}]$.

Figure 6.7 shows the results from our solution procedure. First, we note that power control has a significant impact on BFP. Comparing the case when there is no power control ($Q = 1$) and the case

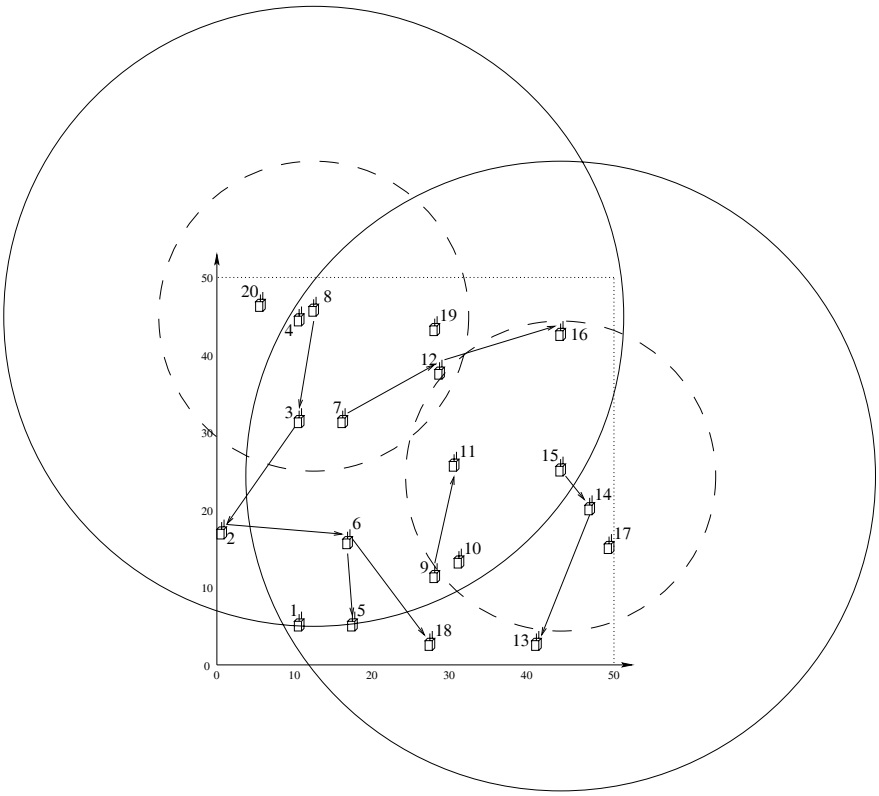
of $Q = 15$, we find that there is nearly a 40% reduction in the total cost (objective value). Second, the cost (objective) is a non-increasing function of Q. However, when Q becomes sufficiently large (e.g., 10 in this network setting), further increase in Q will not have much reduction on cost. This suggests that for practical purpose, the number of required granularity levels to achieve a reasonably good result does not need to be a large number.

To illustrate the difference in routing under different Q, we record the flow routing topology for $Q = 1$ (no power control) and $Q = 10$ in Fig. 6.8 (a) and (b), respectively. In addition to the obvious difference in flow routing topology starting from node 2, we also notice that, due to power control, the frequency band used on some link in Fig. 6.8(a) and (b) are also different. For example, in Fig. 6.8(a), we plot the transmission range and interference range for nodes 8 and 15. Since there is no power control, node 15 has to use the maximum transmission power P_{max} even to its neighboring node 14. Since node 3 is in the interference range of node 15, nodes 8 and 15 cannot use the same frequency band. On the contrary, in Fig. 6.8(b), since each node is capable of adjusting transmission power ($Q = 10$), both nodes 8 and node 15 can use smaller transmission powers to achieve the same connectivity and bit rates. As a result, in Fig. 6.8(b), node 3 is no longer in the interference range of node 15. Thus, in the solution to $Q = 10$, both nodes 8 and 15 are allowed to use the same frequency band. This observation from numerical results validate the benefit on scheduling feasibility brought by power control.

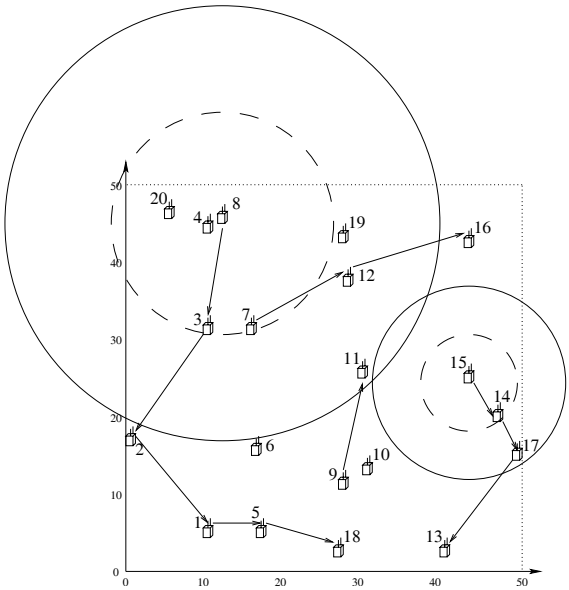
6.7.2 Results on Scheduling Feasibility and Bandwidth Efficiency

To better understand the impact of power control on scheduling feasibility and bandwidth efficiency, we change the available frequency bands in Table 6.2 into the following setting. Each node in the 20-node network has an identical set of frequency bands, i.e., $\mathcal{M}_i = \mathcal{M}$. This special case is the setting for MC-MR.

Clearly, when $|\mathcal{M}|$ is very small, feasible solution may not exist for either power control case ($Q = 10$) or no power control case ($Q = 1$). We now increase the available frequency bands $|\mathcal{M}|$



(a) No power control ($Q = 1$)



(b) With power control ($Q = 10$)

Figure 6.8: Routing topology for the 5 communication sessions in the 20-node network.

at each node and find that starting from $|\mathcal{M}| = 5$, we can obtain a feasible solution for the power control case ($Q = 10$). However, for $|\mathcal{M}| = 5$, there is still no feasible solution for the no power control case ($Q = 1$).

It is not hard to see that if we continue to add available frequency bands into the network, we will eventually have a feasible solution for the no power control case ($Q = 1$). In our numerical results, we find that only when $|\mathcal{M}| = 9$, we find a first feasible solution for the no power control case. Recall that we only need $|\mathcal{M}| = 5$ for a feasible solution with power control. This example verifies that not only power control offers greater design space for scheduling feasibility, it also conserves the number of frequency bands for the network.

6.8 Conclusions

In this chapter, we investigated how to support user communication sessions by jointly performing power control, scheduling, and flow routing for a CR-based multi-hop wireless network. We developed a formal mathematical model for scheduling feasibility under the influence of power control. This model extends the existing so-called protocol model for wireless networks (where transmission range and interference range are deterministic) and thus can be used for a broad class of problems where power control (and thus transmission range and interference range) is part of the optimization space. We formulated a cross-layer optimization problem encompassing power control, scheduling, and flow routing and developed an efficient solution procedure based on branch-and-bound technique and convex hull relaxation. Through numerical results, we demonstrated quantitatively that power control and variable transmission/interference range have significant impact on scheduling feasibility, bandwidth efficiency, and bandwidth-footprint product. These results affirm the critical need of incorporating power control under the existing protocol interference modeling for future wireless network research.

Chapter 7

Summary and Future Work

7.1 Summary

Recently, many new types of wireless networks have emerged for both civil and military applications. Among these new types of wireless networks, we focus on wireless sensor networks and ad hoc networks. Wireless sensor networks consist of battery-powered nodes to provide in-situ, unattended, high-precision, and real-time observation over a vast area. Wireless ad hoc networks are characterized by the absence of infrastructure support. Nodes in an ad hoc network are able to self-organize into a multi-hop network.

For these new wireless networks, many advanced physical layer technologies have been developed to improve performance. In particular, CR has the ability of sensing available bands and switching among bands. It is enabled by recent advances in RF design, signal processing, and communication software. CR promises unprecedented flexibility in radio communications and is viewed as an enabling technology for dynamic spectrum access.

Due to the unique characteristics of these new networks, it is necessary to consider new performance limits, e.g., network lifetime for sensor networks and bandwidth footprint product for CR networks. Such performance limits are important as they can be used as benchmarks for the

design of distributed algorithms and protocols. However, due to some unique requirements at the application layer as well as some unique characteristics associated at the physical layer, existing analytical approaches developed for traditional wireless networks are not adequate. In this dissertation, we focus on the design of new algorithms and optimization techniques to address theoretical performance limits associated with these new wireless networks.

For these new wireless networks, we find that it is necessary to consider constraints at multiple layers, e.g., power control at the physical layer, scheduling at the link layer, and routing at the network layer. The formulation of these problems are usually in very complex form (nonlinear or non-convex) and are mathematically challenging. In this dissertation, we address these difficult problems via the development of some novel algorithms that are both efficient and optimal (or near-optimal). The main contributions of this dissertation are the following.

1. For node lifetime problem, we show how to equitably maximize node lifetime for each node in a sensor network under the lexicographic max-min (LMM) criteria. We develop a polynomial-time algorithm based on the parametric analysis (PA) technique, which is computationally more efficient than the slack variable based approach in previous work. In addition to providing an efficient polynomial-time algorithm for the LMM-optimal node lifetime vector computation, we also show how to obtain a corresponding routing solution among the remaining alive nodes at each stage such that the LMM-optimal node lifetime vector can indeed be achieved. Another contribution of this work is that our approach can also be employed to address the LMM rate allocation problem, which is similar in structure to the LMM node lifetime problem. More important, we find a duality relationship between the LMM node lifetime problem and the LMM rate allocation problem, which enables one to develop solutions to both problems by solving only one of the two problems.

2. For base station placement problem, we aim to optimally place the base station to maximize sensor network lifetime. We design a $(1-\varepsilon)$-approximation algorithm for this problem, which is based on subarea division and fictitious cost point (FCP) representation. Under the given small error bound ε, we first divide the infinite search space for base station location

into a finite number of subareas. We then represent each subarea by an FCP, which is an N-tuple cost vector with each component representing the upper bound of cost to a sensor node in the network. The achievable network lifetime for each FCP can be obtained by a linear program (LP). We prove that any point in the subarea corresponding the best FCP is a $(1-\varepsilon)$-optimal solution for base station placement. The complexity of our approximation algorithm is significantly lower than the state-of-the-art algorithm proposed in previous work.

3. For mobile base station problem, we aim to find the optimal base station moving trajectory as well as time-dependent flow routing solution to maximize lifetime of sensor networks. We contribute an approximation algorithm regarding the movement of a mobile base station. The foundation of our result hinges upon a novel time-to-space transformation for problem formulation. That is, to achieve the maximum network lifetime, we only need to determine the sojourn time for the base station at each point and the flow routing solution when the base station is at this point. Based on this transformation, we present an intermediate result which says that when the location of the base station is constrained to be on a set of pre-determined points, then the optimal solution can be obtained via an LP. For the general mobile base station problem where the location of the base station is un-constrained, we transform this infinite search space into a finite set of FCPs based on the given small error bound ε. Then we can obtain the sojourn time at each FCP (or the corresponding subarea) and the flow routing solution when base station is at this FCP. We prove that such a solution is $(1-\varepsilon)$-optimal. This is the first theoretical result on mobile base station problem.

4. We consider how to minimize the total required bandwidth in a CR network to support a given set of communication sessions. For this problem, it is necessary to consider a cross-layer design, which includes spectrum allocation, scheduling and interference avoidance, and multi-hop multi-path routing. This problem can be formulated as a mixed integer nonlinear program (MINLP). To solve this problem, we propose a sequential fixing (SF) procedure to fix scheduling variables iteratively. Once we have fixed all scheduling variables, the values to all other variables can be obtained by an LP. Simulation results show that the solutions

obtained by SF algorithm are close to a lower bound developed in this work. Thus, these solutions are even closer to the minimum required bandwidth and are near-optimal.

5. We further consider the case that each CR node has power control capability. That is, in addition to scheduling and routing, power control is also part of the optimization space. In this case, we aim to minimize the total required bandwidth footprint product (BFP) to support a set of user sessions. We follow the so-called protocol model and develop a formal mathematical model for scheduling with power control. This model extends existing deterministic interference model and can be used for a wide range of problems. Based on this model, we have formulated a joint power control, scheduling, and routing problem. Our efficient solution procedure to this cross-layer optimization problem is based on the branch-and-bound framework and convex hull relaxation. The solution obtained via our approach yields a $(1-\varepsilon)$-optimal solution, where ε is a small positive constant reflecting our desired accuracy in the final solution. This is the first theoretical result on this problem.

7.2 Future Research Direction

The following problems remain open and will be my future research.

- There are two major interference models for wireless networks, i.e., the protocol model (see Chapters 5 and 6) and the physical model. Under the protocol model, we focus on the interference relationship between two transmissions, i.e., a receiver is interfered by another transmitter if the interference from this transmitter cannot be neglected (over a threshold). Under the physical model, we consider signal to interference plus noise ratio (SINR) directly, i.e., a receiver is interfered if SINR at this receiver is below a threshold. A common agreed understanding is that the physical model is more accurate, despite that it is more complex. In our future work, we will consider how to determine the performance limit under the physical model via nonlinear optimization techniques. We will explore the relationship between the physical model and the protocol model. Another question we would like to answer is that

whether it is possible to set up a suitable threshold for the protocol model such that it yields a similar performance to the physical model.

- Distributed algorithms are very important in network implementation. We will consider how to develop distributed algorithms for the solutions presented in this dissertation. A promising direction for distributed cross-layer optimization problem is subgradient method [7]. The essence of this approach is that for these performance limit problems, the subgradient method may be decomposed into sub-components, each of which is suitable for local computation. However, if the optimization problem is non-convex, the subgradient method may not be able to find the global optimal solution. We will explore some innovative ideas based on subgradient method to provide optimal or near-optimal solutions.

Bibliography

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks (Elsevier),* vol. 38, no. 4, pp. 393–422, 2002.

[2] S.M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications,* vol. 16, issue 8, pp. 1451–1458, Oct. 1998.

[3] M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proc. ACM Mobicom,* pp. 58–72, Cologne, Germany, Aug. 28–Sep. 2, 2005.

[4] BARON Global Optimization Software, http://www.andrew.cmu.edu/user/ns1b/baron/baron.html.

[5] S. Basagni, A. Carosi, E. Melachrinoudis, C. Petrioli, and Z.M. Wang, "A new MILP formulation and distributed protocols for wireless sensor networks lifetime maximization," in *Proc. IEEE International Conference on Communications,* pp. 3517–3524, Istanbul, Turkey, June 11–15, 2006.

[6] M.S. Bazaraa, J.J. Jarvis, and H.D. Sherali, *Linear Programming and Network Flows,* third edition, John Wiley & Sons, 2004.

[7] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms,* third edition, John Wiley & Sons, Inc., New York, NY, 2006.

[8] A. Behzad and I. Rubin, "Impact of power control on the performance of ad hoc wireless networks," in *Proc. IEEE Infocom,* pp. 102–113, Miami, FL, March 13–17, 2005.

[9] D. Bertsekas and R. Gallager, *Data Networks,* Prentice Hall, Englewood Cliffs, NJ, 1992.

[10] M. Bhardwaj, T. Garnett, and A.P. Chagdrakasan, "Upper bounds on the lifetime of sensor networks," in *Proc. IEEE International Conference on Communications,* pp. 785–790, Helsinki, Finland, June 11–13, 2001.

[11] M. Bhardwaj and A.P. Chandrakasan, "Bounding the lifetime of sensor networks via optimal role assignments," in *Proc. IEEE Infocom,* pp. 1587–1596, New York, NY, June 23–27, 2002.

[12] R. Bhatia and M. Kodialam, "On power efficient communication over multi-hop wireless networks: joint routing, scheduling and power control," in *Proc. IEEE Infocom,* pp. 1457–1466, Hong Kong, China, March 7–11, 2004.

[13] D. Blough and S. Paolo, "Investigating upper bounds on network lifetime extension for cell-based energy conservation techniques in stationary ad hoc networks," in *Proc. ACM Mobicom,* pp. 183–192, Atlanta, GA, Sep. 23–28, 2002.

[14] A. Bogdanov, E. Maneva, and S. Riesenfeld, "Power-aware base station positioning for sensor networks," in *Proc. IEEE Infocom,* pp. 575–585, Hong Kong, China, March 7–11, 2004.

[15] T.X. Brown, H.N. Gabow, and Q. Zhang, "Maximum flow-life curve for a wireless ad hoc network," in *Proc. ACM Mobihoc,* pp. 128–136, Long Beach, CA, Oct. 4–5, 2001.

[16] D. Cabric, S.M. Mishra, and R.W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers,* pp. 772–776, Pacific Grove, CA, Nov. 7–10, 2004.

[17] J.-H. Chang and L. Tassiulas, "Routing for maximum system lifetime in wireless ad-hoc networks," in *Proc. Allerton Conference on Communication, Control, and Computing,* pp. 22–31, Monticello, IL, Sep. 22–24, 1999.

[18] J.-H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," in *Proc. IEEE Infocom,* pp. 22–31, Tel Aviv, Israel, March 26–30, 2000.

[19] J.-H. Chang and L. Tassiulas, "Fast approximate algorithms for maximum lifetime routing in wireless ad-hoc networks," in *Proc. IFIP-TC6 Networking Conference,* pp. 702–713, Paris, France, May 14–19, 2000.

[20] C.C. Chen and D.S. Lee, "A joint design of distributed QoS scheduling and power control for wireless networks," in *Proc. IEEE Infocom,* Barcelona, Catalunya, Spain, April 23–29, 2006.

[21] J. Chou, D. Petrovis, and K. Ramchandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *Proc. IEEE Infocom,* pp. 1054–1062, San Francisco, CA, March 30–April 3, 2003.

[22] N. Clemens and C. Rose, "Intelligent power allocation strategies in an unlicensed spectrum," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 37–42, Baltimore, MD, Nov. 8–11, 2005.

[23] R.L. Cruz and A.V. Santhanam, "Optimal routing, link scheduling and power control in multi-hop wireless networks," in *Proc. IEEE Infocom,* pp. 702–711, San Francisco, CA, March 30–April 3, 2003.

[24] DARPAR Grand Challenge 2005, URL: http://www.darpa.mil/grandchallenge/index.html.

[25] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications,* second edition, Chapter 4, Springer–Verlag, New York, NY, 1998.

[26] M.F. Demirkol and M.A. Ingram, "Power-controlled capacity for interfering MIMO links," in *Proc. IEEE Vehicular Technology Conference,* pp. 187–191, Atlantic City, NJ, Oct. 7–11, 2001.

[27] M.F. Demirkol and M.A. Ingram, "Stream control in network with interfering MIMO links," in *Proc. IEEE Wireless Communications and Networking Conference,* pp. 343–348, New Orleans, LA, March 16–20, 2003.

[28] P. Desnoyers, D. Ganesan, and P. Shenoy, "TSAR: A two tier sensor storage architecture using interval skip graphs," in *Proc. ACM Embedded Networked Sensor Systems,* pp. 39–50, San Diego, CA, Nov. 2–4, 2005.

[29] S.S. Dhillon and K. Chakrabarty, "Sensor placement for effective coverage and surveillance in distributed sensor networks," in *Proc. IEEE Wireless Communications and Networking Conference,* pp. 1609–1614, New Orleans, LA, March 16–20, 2003.

[30] S. Doshi, S. Bhandare, and T.X. Brown, "An on-demand minimum energy routing protocol for a wireless ad hoc network," *ACM Mobile Computing and Communications Review,* vol. 6, no. 3, pp. 50–66, July 2002.

[31] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. ACM Mobicom,* pp. 114–128, Philadelphia, PA, Sep. 26–Oct. 1, 2004.

[32] A. Efrat, S. Har-Peled, and J. Mitchell, "Approximation algorithms for two optimal location problems in sensor networks," in *Proc. IEEE Broadband Communications, Networks, and Systems,* pp. 714–723, Boston, MA, Oct. 3–7, 2005.

[33] T. Elbatt and A. Ephremides, "Joint scheduling and power control for wireless ad-hoc networks," in *Proc. IEEE Infocom,* pp. 976–984, New York, NY, June 23–27, 2002.

[34] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 251–258, Baltimore, MD, Nov. 8–11, 2005.

[35] FCC, "Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies, notice of proposed rule making and order," FCC 03-322.

[36] G.J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications,* vol. 6, no. 3, pp. 311–355, March 1998.

[37] S.R. Gandham, M. Dawande, R. Prakash, and S. Venkatesan, "Energy efficient schemes for wireless sensor networks with multiple mobile base stations," in *Proc. IEEE Globecom,* pp. 377–381, San Francisco, CA, Dec. 1–5, 2003.

[38] G. Ganesan and Y.G. Li, "Cooperative spectrum sensing in cognitive radio networks," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 137–143, Baltimore, MD, Nov. 8–11, 2005.

[39] A. Gashemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 131–136, Baltimore, MD, Nov. 8–11, 2005.

[40] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness,* W.H. Freeman and Company, New York, NY, 1979.

[41] D. Gesbert, M. Shafi, D. Shiu, P.J. Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications,* vol. 21, issue 3, pp. 281–302, April 2003.

[42] A. Goldsmith, S.A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications,* vol. 21, issue 1, pp. 684–702, June 2003.

[43] B. Gomez, A.T. Campbell, M. Naghshineh, and C. Bisdikian, "Conserving transmission power in wireless ad hoc networks," in *Proc. IEEE International Conference on Network Protocols,* pp. 24–34, Riverside, CA, Nov. 11–14, 2001.

[44] P. Gupta and P.R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory,* vol. 46, no. 2, pp. 388–404, March 2000.

[45] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications,* vol. 23, issue 2, pp. 201–220, Feb. 2005.

[46] W. Heinzelman, *Application-specific Protocol Architectures for Wireless Networks,* Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, June 2000.

[47] Y.T. Hou, H. Tzeng, and S.S. Panwar, "A generalized max-min rate allocation policy and its distributed implementation using the ABR flow control mechanism," in *Proc. IEEE Infocom,* pp. 1366–1375, San Francisco, CA, March 29–April 2, 1998.

[48] Y.T. Hou, Y. Shi, and H.D. Sherali, "Rate allocation in wireless sensor networks with network lifetime requirement," in *Proc. ACM Mobihoc,* pp. 67–77, Roppongi Hills, Tokyo, Japan, May 24–26, 2004.

[49] Y.T. Hou, Y. Shi, H.D. Sherali, and S.F. Midkiff, "Prolonging sensor network lifetime with energy provisioning and relay node placement," in *Proc. IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks,* pp. 295–304, Santa Clara, CA, Sep. 26–29, 2005.

[50] J. Huang, R.A. Berry, and M.L. Honig, "Spectrum sharing with distributed interference compensation," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 88–93, Baltimore, MD, Nov. 8–11, 2005.

[51] O. Ileri, D. Samardzija, T. Sizer, and N.B. Mandayam, "Demand responsive pricing and competitive spectrum allocation via a spectrum server," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 194–202, Baltimore, MD, Nov. 8–11, 2005.

[52] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. ACM Mobicom,* pp. 66–80, San Diego, CA, Sep. 14–19, 2003.

[53] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," in *Proc. ACM Sigcomm,* pp. 145–158, Portland, OR, Aug. 30–Sep. 3, 2004.

[54] Joint Tactical Radio System, http://jtrs.army.mil/.

[55] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Maximum lifetime data gathering and aggregation in wireless sensor networks," in *Proc. IEEE International Conference on Networking,* pp. 685–696, Atlanta, GA, Aug. 26–29, 2002.

[56] K. Kar, M. Kodialam, T.V. Lakshman, and L. Tassiulas, "Routing for network capacity maximization in energy-constrained ad-hoc networks," in *Proc. IEEE Infocom,* pp. 673–681, San Francisco, CA, Mar. 30–Apr. 3, 2003.

[57] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM Mobicom,* pp. 73–87, Cologne, Germany, Aug. 28–Sep. 2, 2005.

[58] P. Kyasanur and N.H. Vaidya, "Capacity of multi-channel wireless networks: impact of number of channels and interfaces," in *Proc. ACM Mobicom,* pp. 43–57, Cologne, Germany, Aug. 28–Sep. 2, 2005.

[59] Q. Li, J. Aslam, and D. Rus, "Online power-aware routing in wireless Ad-hoc networks," in *Proc. ACM Mobicom,* pp. 97–107, Rome, Italy, July 16–21, 2001.

[60] J. Liu, F. Zhao, P. Cheung, and L. Guibas, "Apply geometric duality to energy-efficient non-local phenomenon awareness using sensor networks," *IEEE Wireless Communications,* vol. 11, issue 6, pp. 62–68, Dec. 2004.

[61] X. Liu and W. Wang, "On the characteristics of spectrum-agile communication networks," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 214–223, Baltimore, MD, Nov. 8–11, 2005.

[62] J. Luo and J.-P. Hubaux, "Joint mobility and routing for lifetime elongation in wireless sensor networks," in *Proc. IEEE Infocom,* pp. 1735–1746, Miami, FL, March 13–17, 2005.

[63] M. McHenry, "Spectrum white space measurements," presentation to New America Foundation BroadBand Forum, June 20, 2003. http://www.newamerica.net/Download_Docs/pdfs/Doc_File_185_1.pdf.

[64] M. McHenry and D. McCloskey, "New York City Spectrum Occupancy Measurements September 2004," available at http://www.sharedspectrum.com/inc/content/measurements/nsf/NYC_report.pdf.

[65] N. Megiddo, "Linear-time almorithm for linear programming in R^3 and related problems," *SIAM Journal on Computing,* vol. 12, pp. 759–776, 1983.

[66] A. Michail and A. Ephremides, "Energy efficient routing for connection-oriented traffic in ad-hoc wireless networks," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications,* pp 762–766, London, UK, Sep. 18–21, 2000.

[67] A. Misra and S. Banerjee, "MRPC: maximizing network lifetime for reliable routing in wireless environments," in *Proc. IEEE Wireless Communications and Networking Conference,* pp. 800–806, Orlando, FL, March 17–21, 2002.

[68] S.M. Mishra, A. Sahai, and R.W. Brodersen, "Cooperative sensing among cognitive radios," in *Proc. IEEE International Conference on Communications,* pp. 1658–1663, Istanbul, Turkey, June 11–15, 2006.

[69] J. Mitola III, *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio,* Ph.D. thesis, KTH Royal Institute of Technology, 2000.

[70] "Ultra Wideband (UWB) Frequently Asked Questions (FAQ)", Multispectral Solutions, Inc., July 2003, http://www.multispectral.com/.

[71] G.L. Nemhauser and L.A. Wolsey, *Integer and Combinatorial Optimization,* John Wiley & Sons, New York, NY, 1999.

[72] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 269–278, Baltimore, MD, Nov. 8–11, 2005.

[73] J. Pan, Y.T. Hou, L. Cai, Y. Shi, and S.X. Shen, "Topology control for wireless sensor networks," in *Proc. ACM Mobicom,* pp. 286–299, San Diego, CA, Sep. 14–19, 2003.

[74] C. Peng, H. Zheng, and B.Y. Zhao, "Utilization and fairness in spectrum assignment for opportunistic spectrum access," *ACM/Springer Mobile Networks and Applications,* vol. 11, issue 4, pp. 555–576, Aug. 2006.

[75] S.S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense sensor network," *IEEE Signal Processing Magazine,* vol. 19, no. 2, pp. 51–60, March 2002.

[76] M.B. Pursley, H.B. Russell, and J.S. Wysocarski, "Energy-efficient routing in frequency-hop radio networks with partial-band interference," in *Proc. IEEE Wireless Communications and Networking Conference,* pp. 79–83, Chicago, IL, Sep. 23–28, 2000.

[77] R.C. Qiu, H. Liu, and X. Shen, "Ultra-wideband for multiple access communications," *IEEE Communications Magazine,* vol. 43, issue 2, pp. 80–87, Feb. 2005.

[78] R. Ramanathan and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit power adjustment," in *Proc. IEEE Infocom,* pp. 404–413, Tel-Aviv, Israel, March 26–30, 2000.

[79] K. Ramachandran, E. Belding-Royer, K. Almeroth, and M. Buddhikot, "Interference-aware channel assignment in multi-radio wireless mesh networks," in *Proc. IEEE Infocom,* Barcelona, Catalunya, Spain, April 23–29, 2006.

[80] K. Ramchandran, "Distributed sensor networks: opportunities and challenges in signal processing and communications," presentation at *NSF Workshop on Distributed Communications and Signal Processing for Sensor Networks,* Evanstaon, IL, Dec. 2002, available at http://www.ece.northwestern.edu/˜pappas/nsf_workshop/presentations/ ramchandran_workshop_DDSP.ppt.

[81] A. Raniwala and T. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. IEEE Infocom,* pp. 2223–2234, Miami, FL, March 13–17, 2005.

[82] T.S. Rappaport, *Wireless Communications: Principles and Practice,* second edition, Prentice Hall, Upper Saddle River, NJ, Dec. 2001.

[83] J.H. Reed, *Software Radio: A Modern Approach to Radio Engineering,* Prentice Hall, Upper Saddle River, NJ, May 2002.

[84] J.H. Reed, *An Introduction to Ultra Wideband Communication Systems,* Prentice Hall, Upper Saddle River, NJ, April 2005.

[85] T. Renk, C. Kloeck, and F.K. Jondral, "A cognitive approach to the detection of spectrum holes in wireless networks," in *Proc. IEEE Consumer Communications and Networking Conference,* pp. 1118–1122, Las Vegas, NV, Jan. 11–13, 2007.

[86] V. Rodoplu and T.H. Meng, "Minimum energy mobile wireless networks," *IEEE Journal on Selected Areas in Communications,* vol. 17, issue 8, pp. 1333–1344, Aug. 1999.

[87] A. Sankar and Z. Liu, "Maximum lifetime routing in wireless ad-hoc networks," in *Proc. IEEE Infocom,* pp. 1089–1097, Hong Kong, China, March 7–11, 2004.

[88] H.D. Sherali, W.P. Adams, and P.J. Driscoll, "Exploiting special structures in constructing a hierarchy of relaxations for 0-1 mixed integer problems," *Operations Research,* vol. 46, no. 3, pp. 396–405, 1998.

[89] H.D. Sherali and W.P. Adams, *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems,* Kluwer Academic Publishers, Dordrecht, 1999.

[90] Y. Shi and Y.T. Hou, "Optimal power control for multi-hop software defined radio networks," in *Proc. IEEE Infocom,* pp. 1694–1702, Anchorage, AL, May 6–12, 2007.

[91] G.T. Sibley, M.H. Rahimi, and G.S. Sukhatme, "Robomote: A tiny mobile robot platform for large-scale ad-hoc sensor networks," in *Proc. IEEE International Conference on Robotics and Automation,* pp. 1143–1148, Washington, DC, May 11–15, 2002.

[92] S. Singh, M. Woo, and C.S. Raghavendra, "Power-aware routing in mobile ad hoc networks," in *Proc. ACM Mobicom,* pp. 181–190, Dallas, TX, Oct. 25–30, 1998.

[93] K. Sohrabi, J. Gao, V. Ailawadhi, and G. Pottie, "Protocols for self-organizing of a wireless sensor network," *IEEE Personal Communications Magazine,* vol. 7, pp. 16–27, Oct. 2000.

[94] V. Srinivasan, P. Nuggehalli, C.F. Chiasserini, and R. Rao, "Cooperation in wireless ad hoc networks," in *Proc. IEEE Infocom*, pp. 808–817, San Francisco, CA, Mar. 30–Apr. 3, 2003.

[95] M.E. Steenstrup, "Opportunistic use of radio-frequency spectrum: A network perspective," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 638–641, Baltimore, MD, Nov. 8–11, 2005.

[96] I. Stojmenovic and X. Lin, "Power-aware localized routing in wireless networks," *IEEE Transactions on Parallel and Distributed Systems,* vol. 12, no. 11, pp. 1122–1133, Nov. 2001.

[97] K. Sundaresan and R. Sivakumar, "Routing in ad hoc networks with MIMO links," in *Proc. IEEE International Conference on Network Protocols,* pp. 85–98, Boston, MA, Nov. 6–9, 2005.

[98] J. Tang, G. Xue, C. Chandler, and W. Zhang, "Interference-aware routing in multihop wireless networks using directional antennas," in *Proc. IEEE Infocom,* pp. 751–760, Miami, FL, March 13–17, 2005.

[99] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory,* vol. 44, no. 2, pp. 744–765, March 1998.

[100] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications,* vol. 10, no. 6, pp. 585–596, Nov. 1999.

[101] D. Ugarte and A.B. McDonald, "On the capacity of dynamic spectrum access enabled networks," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 630–633, Baltimore, MD, Nov. 8–11, 2005.

[102] E. Visostky, S. Kuffner, and R. Peterson, "On collaborative detection of TV transmissions in support of dynamic spectrum sharing," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 338–345, Baltimore, MD, Nov. 8–11, 2005.

[103] G. Wang, G. Cao, and T. La Porta, "Movement-assisted sensor deployment," in *Proc. IEEE Infocom,* pp. 2469–2479, Hong Kong, China, March 7–11, 2004.

[104] Z. Wang, S. Basagni, E. Melachrinoudis, and C. Petrioli, "Exploiting sink mobility for maximizing sensor networks lifetime," in *Proc. IEEE Hawaii International Conference on System Sciences,* Waikoloa, HA, Jan. 3–6, 2005.

[105] R. Wattenhofer, L. Li, P. Bahl, and Y.-M. Wang, "Distributed topology control for power efficient operation in multihop wireless ad hoc networks," in *Proc. IEEE Infocom,* pp. 1388–1397, Anchorage, AK, April 22–26, 2001.

[106] E. Welzl, "Smallest enclosing disks," *Lecture Notes in Computer Science,* vol. 555, pp. 359–370, 1991.

[107] J. Wu and S. Yang, "SMART: A scan-based movement-assisted sensor deployment method in wireless sensor networks," in *Proc. IEEE Infocom,* pp. 2313–2324, Miami, FL, March 13–17, 2005.

[108] C. Xin, B. Xie, and C.-C. Shen, "A novel layered graph model for topology formation and routing in dynamic spectrum access networks," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 308–317, Baltimore, MD, Nov. 8–11, 2005.

[109] K. Xu, H. Hassanein, and G. Takahara, "Relay node deployment strategies in heterogeneous wireless sensor networks: multiple-hop communication case," in *Proc. IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks,* pp. 575–585, Santa Clara, CA, Sep. 26–29, 2005.

[110] M. Younis, M. Bangad, and K. Akkaya, "Base-station repositioning for optimized performance of sensor networks," in *Proc. IEEE Vehicular Technology Conference,* pp. 2956–2960, Orlando, FL, Oct. 4–9, 2003.

[111] H. Zhang and J. Hou, "On deriving the upper bound of α-lifetime for large sensor networks," on *Proc. ACM Mobihoc,* pp. 121–132, Tokyo, Japan, May 24–26, 2004.

[112] J. Zhao, H. Zheng, and G. Yang, "Distributed coordination in dynamic spectrum allocation networks," in *Proc. IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks,* pp. 259–268, Baltimore, MD, Nov. 8–11, 2005.

[113] W. Zhao, M. Ammar, and E. Zegura, "Controlling the mobility of multiple data transport ferries in a delay-tolerant network," in *Proc. IEEE Infocom,* pp. 1407–1418, Miami, FL, March 13–17, 2005.

[114] Y. Zou and K. Chakrabarty, "Sensor deployment and target localization based on virtual forces," in *Proc. IEEE Infocom,* pp. 1293–1303, San Francisco, CA, March 30–April 3, 2003.

[115] G. Zussman and A. Segall, "Energy efficient routing in ad hoc disaster recovery networks," in *Proc. IEEE Infocom,* pp. 682–691, San Francisco, CA, March 30–April 3, 2003.

Vita

Yi Shi was born in Hefei, Anhui, China, 1975. He received his B.S. degree in computer science from Department of Special Class for Gifted Young, University of Science and Technology of China, Hefei, Anhui, China, in 1998. In 2001, he received his M.S. degree in computer science from Institute of Software, Chinese Academy of Sciences, Beijing, China. He came to study in the U.S. in August 2001. He obtained a second M.S. degree from Department of Computer Science, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, in 2003. He then enrolled in the Ph.D. program in the Bradley Department of Electrical and Computer Engineering.

Yi is a Graduate Research Assistant (GRA) in the Complex Network Systems Research (CNSR) Lab headed by his advisor Prof. Y. Thomas Hou. His research focuses on algorithms and optimization for cognitive radio wireless networks, MIMO and cooperative communication networks, sensor networks, and ad hoc networks. His work has appeared in some highly selective international conferences (*ACM Mobicom*, *ACM Mobihoc*, and *IEEE Infocom*) and IEEE journals.

Yi was a member of the Technical Program Committee of IEEE Workshop on Networking Technologies for Software Defined Radio (SDR) Networks (held in conjunction with IEEE SECON 2006), Reston, VA, Sept. 25, 2006. He was a guest lecturer for a number of courses at Virginia Tech, such as ECE 6504 *Advanced Foundations of Networking* (Spring 2005), ECE 4614 *Telecommunication Networks* (Spring 2006), and ECE/CS 6570 *Advanced Foundation of Networking* (Fall 2007).

While an undergraduate, Yi was on a three-member team for China at the International Mathematical Contest in Modeling in 1997 and achieved the Meritorious Award. In 1998, he served as advisor on one team and his team won the Meritorious Award. Yi was a recipient of Chinese Government Award for Outstanding Students Abroad in 2006. He was a recipient of Student Travel Grants of IEEE Infocom 2007 and IEEE SECON 2007.

Journal Papers

1. **Y. Shi**, Y.T. Hou, and H.D. Sherali, "Cross-layer optimization for data rate utility problem in UWB-based ad hoc networks," submitted to *IEEE Transactions on Mobile Computing,* accepted with minor revision.

2. **Y. Shi** and Y.T. Hou, "Power control for multi-hop cognitive radio networks," submitted to *IEEE Transactions on Mobile Computing,* under review.

3. **Y. Shi** and Y.T. Hou, "An approximation algorithm for optimal base station placement in sensor networks," submitted to *IEEE Transactions on Sensor Networks,* under review.

4. **Y. Shi**, Y.T. Hou, and A. Efrat, "Algorithm design for a class of base station location problems in sensor networks," to appear in *ACM/Springer Wireless Networks.*

5. **Y. Shi**, Y.T. Hou, H.D. Sherali, and S.F. Midkiff, "Optimal data routing for UWB-based sensor networks," *IEEE Journal on Selected Areas in Communications,* vol. 24, issue 4, pp. 857–863, April 2006.

6. Y.T. Hou, **Y. Shi**, and H.D. Sherali, "Spectrum Sharing for Multi-hop Networking with Cognitive Radios," *IEEE Journal on Selected Areas in Communications,* vol. 26, issue 1, Jan. 2008.

7. Y.T. Hou, **Y. Shi**, and H.D. Sherali, "Rate allocation and network lifetime problems for wireless sensor networks," to appear in *IEEE Transactions on Networking.*

8. Y.T. Hou, **Y. Shi**, J. Pan, and S.F. Midkiff, "Maximizing the lifetime of wireless sensor networks through optimal single-session flow routing," *IEEE Transactions on Mobile Computing,* vol. 5, no. 9, pp. 1255–1266, Sep. 2006.

9. Y.T. Hou, **Y. Shi**, H.D. Sherali, and S.F. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *IEEE Transactions on Wireless Communications,* vol. 4, no. 5, pp. 2579–2590, Sep. 2005.

10. Y.T. Hou and **Y. Shi**, "Variable bit rate flow routing in wireless sensor networks," *IEEE Transactions on Wireless Communications,* vol. 6, no. 6, pp. 2140–2148, June 2007.

11. Y.T. Hou, **Y. Shi**, and H.D. Sherali, "On node lifetime problem for energy-constrained wireless sensor networks," *ACM/Springer Mobile Networks and Applications,* vol. 10, no. 6, pp. 865–878, Dec. 2005.

12. Y.T. Hou, **Y. Shi**, and H.D. Sherali, "Optimal base station selection for anycast routing in wireless sensor networks," *IEEE Transactions on Vehicular Technology,* vol. 55, issue 3, pp. 813–821, May 2006.

13. Y.T. Hou, **Y. Shi**, H.D. Sherali, and J.E. Wieselthier, "Multicast communications in ad hoc networks using directional antennas: A lifetime-centric approach," *IEEE Transactions on Vehicular Technology.* vol. 56, issue 3, pp. 1333–1344, May 2007.

14. J. Liu, Y.T. Hou, **Y. Shi**, and H.D. Sherali, "Maximizing the capacity of multiuser MIMO networks with interference," to appear in *IEEE Transactions on Wireless Communications.*

15. J. Pan, L. Cai, Y.T. Hou, **Y. Shi**, and S.X. Shen, "Optimal base-station locations in two-tiered wireless sensor networks," *IEEE Transactions on Mobile Computing,* vol. 4, no. 5, pp. 458–473, Sep./Oct. 2005.

16. J. Pan, Y.T. Hou, L. Cai, **Y. Shi**, and S.X. Shen, "Serialized optimal relay schedules in two-tiered wireless sensor networks," *Computer Communications,* vol. 29, no. 4, pp. 511–524, Feb. 2006.

Conference Papers

1. **Y. Shi** and Y.T. Hou, "A distributed optimization algorithm for multi-hop cognitive radio networks," to appear in *Proc. IEEE Infocom 2008,* Phoenix, AZ, April 15–17, 2008.

2. **Y. Shi** and Y.T. Hou, "Theoretical results on base station movement problem for sensor network," to appear in *Proc. IEEE Infocom 2008,* Phoenix, AZ, April 15–17, 2008.

3. **Y. Shi** and Y.T. Hou, "Optimal power control for multi-hop software defined radio networks," in *Proc. IEEE Infocom,* pp. 1694–1702, Anchorage, AL, May 6–12, 2007.

4. **Y. Shi**, Y.T. Hou, H.D. Sherali, and S.F. Midkiff, "Cross-layer optimization for routing data traffic in UWB-based sensor networks," in *Proc. ACM Mobicom,* pp. 299–312, Cologne, Germany, Aug. 28–Sep. 2, 2005.

5. Y.T. Hou, **Y. Shi**, and H.D. Sherali, "Optimal spectrum sharing for multi-hop software defined radio networks," in *Proc. IEEE Infocom,* pp. 1–9, Anchorage, AL, May 6–12, 2007.

6. Y.T. Hou, **Y. Shi**, H.D. Sherali, and J.E. Wieselthier, "Online lifetime-centric multicast routing for ad hoc networks with directional antennas," in *Proc. IEEE Infocom,* pp. 761–772, Miami, FL, March 13–17, 2005.

7. Y.T. Hou, **Y. Shi**, and H.D. Sherali, "Rate allocation in wireless sensor networks with network lifetime requirement," in *Proc. ACM Mobihoc,* pp. 67–77, Roppongi Hills, Tokyo, Japan, May 24–26, 2004.

8. J. Pan, Y.T. Hou, L. Cai, **Y. Shi**, and S.X. Shen, "Topology control for wireless sensor networks," in *Proc. ACM Mobicom,* pp. 286–299, San Diego, CA, Sep. 14–19, 2003.

9. **Y. Shi** and Y.T. Hou, "Network capacity of UWB-based sensor networks," in *Proc. IEEE International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks,* Vancouver, British Columbia, Canada, Aug. 14–17, 2007.

10. **Y. Shi** and Y.T. Hou, "Approximation algorithm for base station placement in wireless sensor networks," in *Proc. IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks,* pp. 512–517, San Diego, CA, June 18–21, 2007.

11. **Y. Shi**, Y.T. Hou, H.D. Sherali, and S. Kompella, "Cross-layer optimization for UWB-based ad hoc networks," in *Proc. IEEE Military Communications Conference,* pp. 1–7, Washington, DC, Oct. 23–25, 2006.

12. **Y. Shi**, Y.T. Hou, and A. Efrat, "Algorithm design for base station placement problems in sensor networks," in *Proc. IEEE International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks,* Waterloo, Ontario, Canada, Aug. 7–9, 2006.

13. Y.T. Hou, **Y. Shi**, J. Pan, S.F. Midkiff, and K. Sohraby, "Single-beam flow routing for wireless sensor networks," in *Proc. IEEE Global Telecommunications Conference,* pp. 3263–3268, St Louis, MO, Nov. 28–Dec. 2, 2005.

14. Y.T. Hou, **Y. Shi**, H.D. Sherali, and S.F. Midkiff, "Prolonging sensor network lifetime with energy provisioning and relay node placement," in *Proc. IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks,* pp. 295–304, Santa Clara, CA, Sep. 26–29, 2005.

15. Y.T. Hou, **Y. Shi**, and H.D. Sherali, "On base station selection for anycast flow routing in energy-constrained wireless sensor networks," in *Proc. IEEE International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks,* Orlando, FL, Aug. 22–24, 2005.

16. Y.T. Hou, **Y. Shi**, J.H. Reed, and K. Sohraby, "Flow routing for variable bit rate source nodes in energy-constrained wireless sensor networks," in *Proc. IEEE International Conference on Communications,* pp. 3057–3062, Seoul, Korea, May 16–20, 2005.

17. Y.T. Hou, **Y. Shi**, J. Pan, and S.F. Midkiff, "Lifetime-optimal data routing in wireless sensor networks without flow splitting," in *Workshop on Broadband Advanced Sensor Networks,* San Jose, CA, Oct. 25, 2004.

18. Y.T. Hou, **Y. Shi**, and H.D. Sherali, "On lexicographic max-min node lifetime for wireless sensor networks," in *Proc. IEEE International Conference on Communications,* pp. 3790–3796, Paris, France, June 20–24, 2004.

19. Y.T. Hou, **Y. Shi**, and J. Pan, "A lifetime-aware single-session flow routing algorithm for energy-constrained wireless sensor networks," in *Proc. IEEE Military Communications Conference,* pp. 603–608, Boston, MA, Oct. 13–16, 2003.

20. Y.T. Hou, **Y. Shi**, J. Pan, and S.F. Midkiff, "Optimal single-session flow routing for wireless sensor networks," in *Proc. IEEE Vehicular Technology Conference,* pp. 2855–2859, Orlando, FL, Oct. 6–9, 2003.

21. J. Liu, Y.T. Hou, **Y. Shi**, and H.D. Sherali, "Optimization of multiuser MIMO networks with interference," in *Proc. IEEE Global Telecommunications Conference,* pp. 1–6, San Franscisco, CA, Nov. 27–Dec. 1, 2006.

22. J. Liu, T.Y. Park, Y.T. Hou, **Y. Shi**, and H.D. Sherali, "Cross-layer optimization of MIMO-based mesh networks under orthogonal channels," in *Proc. IEEE Wireless Communications and Networking Conference,* pp. 49–54, Hong Kong, China, March 11–15, 2007.

23. J. Pan, L. Cai, Y.T. Hou, **Y. Shi**, and S.X. Shen, "Locating base-stations for video sensor networks," in *Proc. IEEE Vehicular Technology Conference,* pp. 3000–3004, Orlando, FL, Oct. 6–9, 2003.