Resource Allocation in Relay Enhanced Broadband Wireless Access Networks

by

Preetha Thulasiraman

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

Waterloo, Ontario, Canada, 2010

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

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

Abstract

The use of relay nodes to improve the performance of BWA networks has been the subject of intense research activities in recent years. Relay enhanced BWA networks are anticipated to support multimedia traffic (i.e., voice, video, and data traffic). In order to guarantee service to network users, efficient resource distribution is imperative. Wireless multihop networks are characterized by two inherent dynamic characteristics: 1) the existence of wireless interference and 2) mobility of user nodes. Both mobility and interference greatly influence the ability of users to obtain the necessary resources for service. In this dissertation we conduct a comprehensive research study on the topic of resource allocation in the presence of interference and mobility. Specifically, this dissertation investigates the impact interference and mobility have on various aspects of resource allocation, ranging from fairness to spectrum utilization. We study four important resource allocation algorithms for relay enhanced BWA networks. The problems and our research achievements are briefly outlined as follows.

First, we propose an interference aware rate adaptive subcarrier and power allocation algorithm using maximum multicommodity flow optimization. We consider the impact of the wireless interference constraints using Signal to Interference Noise Ratio (SINR). We exploit spatial reuse to allocate subcarriers in the network and show that an intelligent reuse of resources can improve throughput while mitigating the impact of interference. We provide a sub-optimal heuristic to solve the rate adaptive resource allocation problem. We demonstrate that aggressive spatial reuse and fine tuned-interference modeling garner advantages in terms of throughput, end-to-end delay and power distribution.

Second, we investigate the benefits of decoupled optimization of interference aware routing and scheduling using SINR and spatial reuse to improve the overall achievable throughput. We model the routing optimization problem as a linear program using maximum concurrent flows. We develop an optimization formulation to schedule the link traffic such that interference is mitigated and time slots are reused appropriately based on

spatial TDMA (STDMA). The scheduling problem is shown to be NP-hard and is solved using the column generation technique. We compare our formulations to conventional counterparts in the literature and show that our approach guarantees higher throughput by mitigating the effect of interference effectively.

Third, we investigate the problem of multipath flow routing and fair bandwidth allocation under interference constraints for multihop wireless networks. We first develop a novel isotonic routing metric, RI^3M, considering the influence of inter-flow and intra-flow interference. Second, in order to ensure QoS, an interference-aware max-min fair bandwidth allocation algorithm, LMX:M^3F, is proposed where the lexicographically largest bandwidth allocation vector is found among all optimal allocation vectors while considering constraints of interference on the flows. We compare with various interference based routing metrics and interference aware bandwidth allocation algorithms established in the literature to show that RI^3M and LMX:M^3F succeed in improving network performance in terms of delay, packet loss ratio and bandwidth usage.

Lastly, we develop a user mobility prediction model using the Hidden Markov Model (HMM) in which prediction control is transferred to the various fixed relay nodes in the network. Given the HMM prediction model, we develop a routing protocol which uses the location information of the mobile user to determine the interference level on links in its surrounding neighborhood. We use SINR as the routing metric to calculate the interference on a specific link (link cost). We minimize the total cost of routing as a cost function of SINR while guaranteeing that the load on each link does not exceed its capacity. The routing protocol is formulated and solved as a minimum cost flow optimization problem. We compare our SINR based routing algorithm with conventional counterparts in the literature and show that our algorithm reinforces routing paths with high link quality and low latency, therefore improving overall system throughput.

The research solutions obtained in this dissertation improve the service reliability and QoS assurance of emerging BWA networks.

Acknowledgements

I first want to extend my sincerest thanks and gratitude to my PhD supervisor, Dr. Xuemin (Sherman) Shen. Dr. Shen has been very supportive, both in professional matters and personal matters and has guided and helped me throughout the course of my PhD study. His insights into the technical issues related to my research have been of great help in setting the directions of my work. I will always admire and respect his professionalism and integrity.

I want to express my thanks to my friend and colleague, Dr. Jiming Chen, whose helpful discussions and insights during the later stages of my PhD study has furthered my knowledge in my research area.

I want to thank Dr. Hossam Hassanein, Dr. Sagar Naik, Dr. Liang-Liang Xie and Dr. Kumaraswamy Ponnambalam for serving on my thesis and defense committee. Thanks also to Dr. Mahesh V. Tripunitara who served as a member of my PhD seminar and to Dr. Patrick Lam who served as the delegate during my PhD defense. I would also like to extend my thanks to all members of the Broadband Communication Research (BBCR) group for their discussions and suggestions in our weekly group meetings. I also want to thank the administrative staff of the Electrical and Computer Engineering department, Ms. Wendy Boles, Ms. Annette Dietrich, Ms. Lisa Hendel, and Ms. Karen Schooley for their help and cooperation.

I want to convey my thanks to my sisters for their assistance and support throughout various stages of my education.

The successes that I have achieved thus far are largely due to my parents, Dr. K. Thulasiraman and Santha Thulasiraman. The sacrifices they made for the sake of my education, and their love and confidence in me is immeasurable. They are my supporters, teachers, and role models that have guided me throughout my life.

Finally, I want to express my heartfelt thank you to my husband, Venkatesh. Every good thing that has happened in my life is because of him. His support, encouragement,

patience and endless love have carried me throughout my studies.

Dedication

This dissertation is dedicated to my parents, Dr. K. and Santha Thulasiraman, and my husband Venkatesh for their love, constant support, and guidance.

Contents

List of Tables

List of Figures

List of Acronyms

ACK	Acknowledgement
AWGN	Additive White Gaussian Noise
BBR	Bandwidth Blocking Ratio
BS	Base Station
BWA	Broadband Wireless Access
CBR	Constant Bit Rate
CBT	Channel Busy Time
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CTS	Clear To Send
DMPR	Disjoint Multipath Routing
EM	Expectation/Maximization
FDMA	Frequency Division Multiple Access
FIFO	First In First Out
HMM	Hidden Markov Model
IR	Interference Ratio
LOS	Line Of Sight
LTE	Long-Term Evolution
MCF	Multicommodity Flow\Maximum Concurrent Flow
MMF	Max-Min Fairness
MPR	Multipath Routing
NLOS	Non-Line of Sight
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PER	Packet Error Ratio
QoS	Quality of Service
RREQ	Route Request
RTS	Request To Send
SINR	Signal to Interference-Noise Ratio
SNR	Signal to Noise Ratio
STDMA	Spatial Time Division Multiple Access
TMDA	Time Division Multiple Access

Chapter 1

Introduction

The communications landscape has been changing dramatically in recent years under the increasing pressure of rapid technological development and intense competition. Thus, wireless networks are becoming more pervasive, accelerated by new wireless communication technologies, inexpensive wireless equipment and broader Internet access availability. Broadband wireless access (BWA) networks are one such technology that are fast becoming a viable solution to provide ubiquitous communications.

1.1 Overview of Relay Enhanced Broadband Wireless Access Networks

BWA networks are designed to support fixed and mobile users with heterogeneous and high traffic rate requirements. In such networks, a single base station (BS) is deployed to cover a cellular area. In such a large area, users at the cell edge often experience bad channel conditions. Moreover, in urban regions, shadowing by various obstacles can degrade the signal quality in some areas. Emerging broadband wireless applications require increasingly high throughput and more stringent quality-of-service (QoS) requirements. As real-time applications (e.g., voice over IP and video streaming) rapidly grow, BWA

networks are expected to achieve efficient communications. Increasing capacity along with coverage in conventional networks dictates the dense deployment of base stations. Increasing the number of base stations is an expensive solution and increasing the base station power only increases the intercell interference. To meet the goal of low cost network deployment for both short range and long range coverage, the use of relay nodes has been shown to be a promising solution [1, 2]. Broadband cellular multihop networks consist of fixed infrastructure relay nodes whose sole priority is to forward data to and from the users to the BS. Deploying relays is a feasible solution since typical relays are cheaper than base stations and they do not need their own wired backhaul.

The introduction of relay nodes has several performance benefits. First, a relay works on behalf of the BS to increase the network coverage. While conventional cellular systems normally cover a diameter of 2-5km, a relay normally covers a region (subcell) with diameter 200-500m. If the density of relay stations is somewhat high, most user-terminals will be close to one or more relays than to a BS. This has two primary advantages: the radio propagation paths are shortened so that the pathloss is lowered, and the path essentially can be routed around obstacles to mitigate effects of shadowing [1]. This results in higher data rates on the links between relays and users, thereby increasing throughput. Also, from the point of view of the user, the relay acts like a BS and so by having intermediate points of traffic aggregation, the capacity per area element can be balanced [3]. Second, because relay stations are closer to the individual user terminals, the transmit power required for a relay to transmit to a user and vice versa is significantly lower than for a BS, thereby allowing for energy saving. Thus, the practical rationale for the deployment of relay enhanced BWA networks is to ensure that the QoS of a user in terms of data rate, delay, outage probability, etc. does not wholly depend on the its location and distance from the base station. Fig. 1.1 illustrates the general architecture of a relay enhanced BWA network and its various uses and advantages for overcoming inherent transmission gaps.

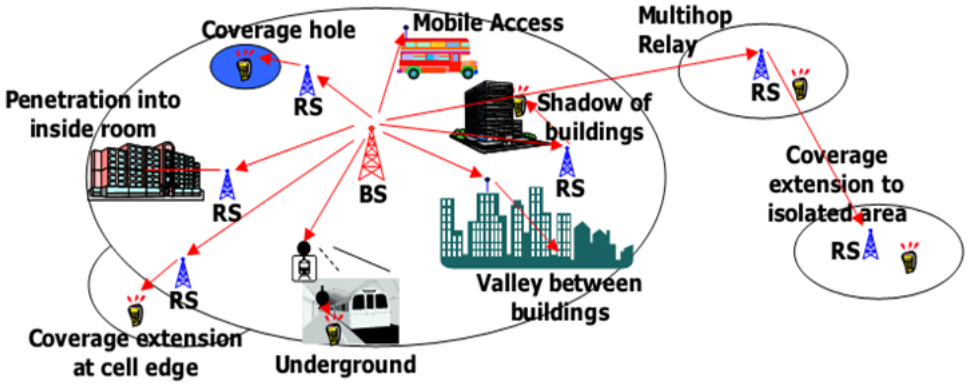


Figure 1.1: Illustration of a relay enhanced wireless access network and its various advantages

1.1.1 OFDMA Based Broadband Wireless Access Standards

Researchers in both academia and industry have accepted orthogonal frequency division multiple access (OFDMA) as the most appropriate air-interface for the emerging broadband wireless access networks and standards. OFDMA is a multi-user version of the popular orthogonal frequency-division multiplexing (OFDM) digital modulation scheme which splits the available system bandwidth into orthogonal subcarriers. This allows simultaneous low data rate transmission from several users on different portions of the broadband spectrum. The advantage of OFDMA is that it bridges frequency division multiple access (FDMA) and time division multiple access (TDMA) by dynamically assigning subcarrriers (FDMA) in different time slots (TDMA). The main advantage of OFDMA is that it allows for multiuser diversity by allowing subcarriers to be shared among multiple users [4]. OFDMA is a very versatile technology that is seen as the communication paradigm for various next generation wireless networks.

Recently, the notion of multihop relay systems in OFDMA based networks has been

attracting a plethora of attention. The relay based extension of OFDMA based BWA networks is used for the purpose of extending and enhancing the network throughput by leveraging the increase in data over multiple hops. In this regard, relay enhanced broadband networks for suburban/urban areas has generated a great deal of interest. For example, the IEEE 802.16j multihop relay standard for mobile WiMax [5, 6] in metropolitan areas has garnered attention in research labs across the world particularly in developing nations and the Long-Term Evolution (LTE) advanced standard [7] has introduced the concept of relay nodes to improve QoS and data rate transmission.

Both LTE and WiMax are 4th generation (4G) broadband network technologies enabling the delivery of last mile wireless access as an alternative to DSL and cable. The WiMax standard was developed to provide high data rates over wide geographic regions to fixed stationary sites. For this reason WiMax is commonly referred to as a wireless metropolitan area network (WMAN). Originally, WiMax network's deployment was a point-to-multipoint (PMP) architecture where the base stations are the central, controlling units. They are connected to the service provider's core network and provides the wireless interface for the user and relay nodes. The wireless links between the base station, users and relays can have both line of sight (LOS) or non-line of sight (NLOS) characteristics. The PHY layer of IEEE 802.16 specifies a multicarrier transmission scheme based on OFDMA, thereby supporting resource allocation in time and frequency by allowing time slots and subcarriers to be allocated simultaneously. Thus, the synergy between OFDMA and relaying techniques offers a promising technology for providing high data rate to users everywhere, anytime. To this end, an OFDMA-based relay-enhanced network comprising various forms of infrastructure-based or dedicated relays is envisaged in the next-generation networks. The combination of OFDMA with relaying techniques provides rich opportunities for cost-effective and high-performance networks [8]. To exploit such opportunities, intelligent resource management algorithms are required, particularly in the presence of inherent wireless characteristics such as interference and mobility.

1.2 Resource Allocation in the Presence of Interference

Interference is the major limiting factor in the performance of wireless multihop networks. Sources of interference include simultaneous transmissions within a certain range as well as concurrent use of the same frequency channel for transmission. Interference is severe in urban areas due to the large number of base stations and mobile users. Interference has been recognized as a major bottleneck in increasing network capacity and throughput and is often responsible for dropped transmissions [9]. Interference experienced at individual nodes (relays and users) is impacted by variations in network size (number of nodes), network density (relative positions of nodes) and traffic per node. There are two widely used models to characterize interference in a wireless network, namely, the *protocol model* and the *physical model*. The protocol model, also known as the unified disk graph model, has been widely used by researchers in the wireless networking community as a way to simplify the mathematical characterization of the physical layer. Under the protocol model, a successful transmission occurs when a node falls inside the transmission range of its intended transmitter and falls outside the interference ranges of other non-intended transmitters. The setting of the transmission range is based on a signal-to-noise-ratio (SNR) threshold. The setting of the interference range is a heuristic approximation and remains an open problem [10]. Under the protocol model, the impact of interference from a transmitting node is binary and is solely determined by whether or not a receiver falls within the interference range of this transmitting node. That is, if a node falls in the interference range of a nonintended transmitter, then this node is considered to be interfered and thus cannot receive correctly from its intended transmitter; otherwise, the interference is assumed to be negligible [11]. Various graph based approaches have been developed for modeling interference using the protocol model. The most common and widely used model is the conflict graph model [12]. The nodes in the conflict graph, $\mathcal{G}_c$,

represent edges in the original connectivity graph $\mathcal{G}$. An edge is placed between two nodes in the conflict graph if the corresponding links in the connectivity graph interfere. Due to such simplification, the protocol model has been widely used in developing algorithms and protocols in wireless networks [13, 14, 15].

The physical model, also known as the SINR model, is based on considerations arising from practical transceiver designs of communication systems that treat interference as noise. Under the physical model, a transmission is successful if and only if signal-to-interference-and-noise-ratio (SINR) at the intended receiver exceeds a threshold so that the transmitted signal can be decoded with an acceptable bit error probability. Furthermore, capacity calculation is based on SINR (via Shannons formula), which takes into account interference due to simultaneous transmissions by other nodes. This model is less restrictive than the protocol interference model as it may occur that a message from node u to node v is correctly received even if there is a simultaneous transmitting node w close to v (for instance, because node u is using a much larger transmit power than node w). As a result, higher network capacity can be achieved by applying the physical interference model. However, the use of the SINR model is computationally more complex and requires various optimization and heuristic techniques to be used to obtain a solution. Nonetheless, it has been shown that despite the computational complexity, the SINR model provides a more practical and realistic assessment of wireless interference [11, 16].

In wireless communications, resource management is vital in controlling how scarce resources can be allocated, distributed, and utilized among all nodes in a system. Unlike wired links which have a constant link capacity, wireless links are relatively vulnerable due to fading over frequency and interference over time. Interference aware resource allocation involves striking a good balance between fair and efficient distribution of spectral resources throughout the network while concurrently mitigating the resulting interference. One of the major difficulties associated with interference mitigation is the lack of

predictability of interference coming from other links that have simultaneous transmissions combined with channel variability. In order to develop efficient resource allocation algorithms that are cognizant of interference, certain potential issues must be addressed, outlined as follows.

- *Channel Assignment*: In all wireless networks, channels are the basic resource to be allocated. Frequency allocation is a major factor in mitigating both co-channel interference and intra-channel interference. Commonly, channel assignment protocols that consider interference use variations of the graph coloring technique [14, 17, 18]. The objective of edge-coloring is to find the minimum number of colors needed to color the edges of the graph such that edges with the same color are not incident on any node. A coloring approach does not necessarily guarantee that two links of the same color will not interfere since they could be within the interference range of each other. Even if they are not in each other's interference range, it is still possible that the SINR strength may not be sufficient enough for successful simultaneous transmissions. In OFDMA networks, the complexity of allocating subcarriers increases exponentially with the number of subcarriers and the number of links in the network. Since an optimal solution to this problem is too complex, suboptimal heuristics have to be used [4, 19].

- *Frequency Reuse*: Since wireless channels are a limited resource, efficient use of resources is important. The concept of frequency reuse can be exploited to utilize the resources efficiently. Reuse is impacted by interference and is dependent on the spatial separation between two links using the same channel for transmissions. Two simultaneous transmissions can employ the same channel(s) if they are far away enough from each other such that the co-channel interference level is below a required threshold. With effective frequency reuse, system capacity can be increased if sufficient interference mitigation is implemented. In OFDMA networks, subcarriers can be reused by using spatial separation such that links can transmit

on the same subcarrier as long as they do not strongly interfere with each other.

- *Fairness in Allocation*: Fairness and system throughput, influenced by wireless interference, are major objectives of resource allocation in wireless networks. Wireless multihop networks are limited by two main resources: bandwidth and network capacity. Achieving high throughput and fair allocation of resources among competing users (or flows) in wireless networks is one of the most important problems in data communications. However, these two objectives may conflict with each other [20]. In resource allocation, two situations must be avoided: 1) a flow must not be starved because of inefficient resources for transmission (i.e., bandwidth); and 2) a flow must not be provided more resources than necessary since only some of the resources may be used and the remaining will be wasted. Resources can be utilized efficiently if only the terminal with the best channel condition transmits, whereby the maximum throughput can be acquired. Such an opportunistic transmission, however, gives rise to unfairness and possibly violates the QoS requirements of some wireless nodes. The concept of fairness in wireless networks is a QoS policy and can be applied to various design issues such as scheduling and routing [21, 22].

- *Mobility*: In addition to wireless interference, an inherent characteristic in wireless networks is the movement of users. Mobility presents significant technical challenges. Due to the uncertainty of user movements, it is difficult to efficiently allocate and reserve resources. However, mobility has been shown to improve capacity and throughput [23]. Resource allocation algorithms that exploit the mobility of users can improve the spectral efficiency of the network. When users move, their radio channels vary due to shadowing and multipath fading. They also encounter other users that they can connect to directly if their terminals are equipped with short-range radios. Therefore, mobility creates opportunities that can be used for

resource-efficient communication.

1.3 Motivations and Objectives

The various issues discussed in Section 1.2 are impacted by various parameters and system dynamics. In this section, we present our motivation and research objectives in terms of developing a framework for interference aware resource allocation.

- *Subcarrier Allocation and Spatial Reuse*: OFDMA networks pose an interesting set of resource allocation problems, particularly 1) routing: how to select paths that minimize interference and increase throughput? 2) subcarrier assignment: what is the set of subcarriers that each user should operate on? and 3) power allocation: what is the optimal power for the nodes transmitting on specific subcarriers? These problems are inter-related and form a challenging cross-layer problem across the network and MAC layers. Research on subcarrier allocation in OFDMA networks focuses on assigning a set of subcarriers to each link such that no subcarrier is assigned to more than one link [24, 25, 26]. These studies rely on the fact that inherently in OFDMA networks, the number of subcarriers is usually large enough so that each link can use a different subcarrier, guaranteeing no two links are transmitting on the same subcarrier and thereby eliminating inter-carrier interference. However, using all the subcarriers that are available is not an efficient method of assigning subcarriers as this may result in an overuse of resources, thereby limiting network performance. It has been shown that spatial reuse of resources (i.e., subcarriers (channels)) provides gains in capacity and throughput in wireless networks [27, 28]. Exploiting the benefits of spatial reuse in subcarrier assignment has the potential to show that some subcarriers may be better for a specific node in terms of channel gain than others. It may be beneficial to have two nodes using the same subcarrier if that subcarrier provides a better transmission medium for both

nodes. Thus, an interference aware subcarrier assignment algorithm taking into consideration spatial frequency reuse and interference aware routing is imperative to increase overall network capacity and throughput.

- *Power Allocation and Scheduling*: Subcarrier allocation in OFDMA networks cannot be investigated alone since various parameters such as power and time are all inter-related. Rate constrained transmissions such as real time voice and video streaming consume the largest traffic load in wireless networks. Rate adaptive resource allocation without regards to interference has been studied in detail for traditional cellular networks. In [29] the authors formulate the capacity maximizing subcarrier and power allocation problem and propose a hueristic allocation algorithm that shows significant performance improvement with respect to static FDMA resource allocation. Similarly, in [30] the authors optimally solve the capacity maximization problem and show that allocating each carrier to the user with the best channel on that carrier and then distributing the power to the carriers by waterfilling maximizes the capacity. Optimal subcarrier and power allocation subject to rate with general objectives such as proportional fairness or QoS constraints have also been studied in [31], [32], and [33]. However, resource allocation for cellular multiuser OFDMA systems with relay stations has not been studied sufficiently. In relay based networks such as our system model, rate adaptive resource allocation that deals with subcarrier and power allocation has generally focused on either a) maximizing throughput subject to either only the base station power constraint or only the relay power constraint while proposing various optimization approaches [34, 35, 36, 37, 38, 39, 40, 41]; or b) solving the subcarrier and power allocation problems separately rather than jointly [29]. In addition, rate adaptive resource allocation based on subcarrier and power distribution has not taken into account the limitations of interference on the various optimization constraints. Therefore, a rate adaptive subcarrier and power allocation scheme that maximizes

overall rate considering various QoS constraints is necessary to support today's multimedia oriented services.

- *Interference and Fairness in Routing and Bandwidth Allocation*: Efficient routing between pairs of nodes in communication networks is a basic problem of network optimization. Achieving high throughput and fair allocation of resources among competing users (or flows) in wireless networks is one of the most important problems in data communications and is directly coupled with routing between nodes. Throughput enhancement and fairness can not be simultaneously achieved, but rather must be balanced [20]. Max-min fairness (MMF) is considered to be an efficient approach that balances these two conflicting objectives by preventing starvation of any flow, and at the same time, increases the bandwidth of a flow as much as possible. In the wireless environment, allocation of bandwidth to paths sharing a set of links is further complicated by the inherent interference that is generated by simultaneous transmissions. Interference can be divided into two categories: inter-flow and intra-flow. Inter-flow interference is generated when two links belonging to different flows are active on the same channel at the same time. Intra-flow interference is when two links belonging to the same flow are active on the same channel at the same time. The effects of interference using the MMF approach have been quantified using graph theoretic approaches (i.e., conflict/contention graph) which ultimately exploits the protocol interference model (i.e., transmissions interfere only within a specific range) [15, 42]. Although, [15, 42] have provided a theoretical foundation for fairness in wireless networks, the reliance on such graph based models induces binary conflicts. The use of the SINR model in determining MMF bandwidth allocation and fair routing would provide a less restrictive and more realistic allocation of bandwidth to the various network paths. Therefore, a SINR based MMF routing and bandwidth allocation optimization formulation would serve to fairly distribute resources and reduce competition between simultaneous flows.

- *Multipath Routing Using SINR Constraints*: Discovering available relaying paths (routes) between a source and base station (BS) is a critical prerequisite for the success of multihop wireless networks. Multipath routing (MPR) has long been recognized as an effective strategy to achieve load balancing and increase reliability [43]. To improve the transmission reliability and avoid shared-link (or node) failures, the multiple paths can be selected to be link- or node-disjoint. In this case, the MPR approach is referred to as disjoint multipath routing (DMPR). DMPR provides better robustness and a greater degree of fault tolerance than compared to the generic MPR. Due to these advantages, DMPR schemes have been researched in the context of wireless networks in order to enhance network survivability [44, 45]. Several routing metrics to capture interference on routing paths have been introduced in the literature. However, the metrics developed have either 1) been based on extending existing routing metrics (i.e., expected transmission count (ETX)) or existing routing algorithms [46, 47] or 2) have integrated interference into variations of the shortest path routing scheme [48, 49]. In the above mentioned works, the interference that is quantified does not refer to the interference received from the physical layer (i.e., signal strength). Rather, there has been a consistent focus on the level of interference in terms of distance using the protocol interference model because of ease of implementation. Limited research on SINR based routing schemes exist. Furthermore, in terms of interference based multipath routing, research has focused on the use of straightforward methods to quantify interference. Specifically, in [50], the authors use an extension of the correlation factor (correlation factor is defined as the number of links connecting two paths) which captures interpath interference but provides little information about the level of interference between simultaneous transmissions. In addition to interference based routing, guaranteeing QoS provisions has also been investigated within this context [51, 52]. Providing fault tolerance and QoS provisioning in the presence of interference are

major issues that must be studied jointly in wireless systems in order to gauge a realistic sense of network performance, particulary in terms of throughput.

- *Mobility Prediction and the Impact of Interference*: In wireless networks, mobility is a major driver of network dynamics. The use of mobility patterns for analysis and simulation has attracted considerable attention in recent years. Several prediction schemes have been proposed and discussed in the literature. Generally, mobility prediction schemes fall under one of two categories: 1) prediction using clustering of nodes and 2) prediction using general Markov chains. Clustering has been used in various mobility prediction schemes [47, 53]. Such schemes predict future movements of nodes based on various parameters such as relative velocity and relative mobility [54, 55, 56, 57]. To form and maintain clusters requires message forwarding which can lead to high message overhead [53, 58]. The effectiveness of a clustering scheme depends on the number of clusters formed. If the number of clusters formed is either too large or too small, the advantages of clustering are lost. Mobility prediction models that depend on user movement history have been typically researched with the Markov model [59, 60, 61]. These schemes contain records of a user's next move, direction of travel, and other information. In [62] and [63], k-order Markov mobility prediction models were proposed. However, an immense amount of mobility history is required to generate high order k Markov prediction models. Other prediction schemes that use user mobility history have been proposed in [64, 65, 66]. A component of resource management is routing since the manner in which resources are distributed is imperative when forwarding data through multiple hops to the base station. It is evident that despite the existence of various mobility prediction models, most do not focus on the impact that mobility has on routing protocols. The combination of wireless interference and user mobility adds to the dynamic nature of the routing process. The conventional way of routing in wireless networks is to route packets on the minimum-cost path from

the source to the destination, where the cost can be defined to be distance, energy etc. Routing the data packets towards the base station on these minimum-cost paths is efficient provided the rate of information generation is low or the channel bandwidth is sufficiently high. However, if the nodes generate data constantly and bandwidth is limited, then routing data on the minimum-cost paths can overload wireless links close to the base station. A routing protocol that does not take the wireless channel bandwidth limitation into consideration might route the packets over highly congested and interfered links and paths. This will lead to increase in congestion, increased delay and packet losses, which in turn will cause retransmission of packets increasing energy consumption. Except for [47], which deals with a clustering approach to mobility management and interference aware routing, there is a lack of attention to mobility prediction using Markov based models for interference sensitive routing protocols in the literature. Thus, there is a need to bridge various aspects of mobility management and interference based resource allocation.

1.4 Research Contributions

In this thesis, we focus on the following research question:

Given a suburban/urban relay enhanced wireless cellular network, how can we design resource allocation algorithms and protocols that explicitly consider the impact of interference and mobility so as to provide users with service availability, QoS assurance, and fair spectrum utilization?

In this dissertation, we tackle the various issues involved in interference aware resource allocation. We study various resource allocation problems through the prism of SINR induced interference and its impact on network performance. The research contributions of this dissertation and their significance are summarized as follows:

- **Interference Aware Rate Adaptive Subcarrier and Power Allocation Us-**

ing Maximum Multicommodity Flow Optimization Method: This work has been reported in [67] and [68]. We study the joint problem of interference aware flow routing and rate adaptive subcarrier and power allocation. The contributions of this work are two-fold. First, we develop a routing approach based on maximum multicommodity flow (MCF) theory that determines paths with least interference using the physical interference model. The MCF approach has been a popular optimization approach for the throughput maximization problem. It has typically been used in wired networks (i.e., traffic engineering). In order to extend the MCF approach in wireless networks, it must be tailored to consider interference constraints. We propose a novel algorithm to solve the traditional MCF problem under interference constraints of wireless networks. The optimization problem formulation for the MCF proposed in this paper, denoted as interference-based MCF (MCFI), uses a SINR derived interference quantification method to maximize the flow from a user and to determine the least interfering paths. Second, we study the problem of rate adaptive subcarrier and power allocation with time and QoS constraints to maximize the overall rate while achieving proportional fairness amongst nodes under a total power constraint. The subcarriers are allocated using the concept of spatial reuse, and interference constraints derived from the interference model are considered in the optimization formulation. In addition, in order to synchronize transmissions, time slots (scheduling) are also allocated. We have shown that spatial reuse of resources (i.e., subcarriers) in an OFDMA network can improve network throughput to a certain limit after which throughput degradation is encountered. This indicates that by allowing users to transmit on the same subcarrier under certain conditions, is not detrimental to the network performance. We also showed that our rate adaptive subcarrier and power allocation algorithm improves the power distribution and transmission delay. This work is presented in Chapter 3.

- **Decoupled Optimization of Routing and Scheduling Using SINR Interference Constraints**: This work has been published in [69]. In this work we address the issue of joint routing and scheduling using Spatial TDMA (STDMA) which allows TDMA slots to be shared by simultaneous transmissions if they are sufficiently geographically separated. The joint optimization of routing and scheduling under SINR interference constraints can not be solved in a single optimization formulation because of its computational complexity. To overcome this limitation, we develop a decoupled optimization scheme that 1) routes traffic such that the achieved throughput is enhanced and 2) schedules concurrent transmissions such that system efficiency is increased and interference is mitigated. The contributions of this work are as follows: First, we formulate and solve a maximum concurrent flow linear program that maximizes the traffic routed in the network by finding the flows on each link on all paths by explicitly considering the impact of interference on the capacity using the physical interference model. We refer to this problem as MCF-ROPT. Second, given the flow for each link, we develop a STDMA scheduling scheme that schedules the flows in spatially reused time slots such that link traffic demand is satisfied and interference is mitigated. We incorporate the effect of reusing multiple carriers, meaning that two links transmitting on the same subcarrier can be scheduled in the same time slot as long as the SINR values for the receivers are satisfied. Our STDMA multicarrier traffic sensitive scheduling scheme is denoted as the SM-TSS problem. This work is presented in Chapter 4

- **Multipath Routing and Max-Min Fair QoS Provisioning Under Interference Constraints**: This work has been reported in [70], [71] and [72]. In this work we address the issues of routing and fair bandwidth allocation in the presence of interference using the SINR model. Our contributions in this work can be summarized as followed. First, we design an isotonic routing metric which is cognizant of interference and provides reliable multipath routing. The routing metric is used

to quantify the interference on the network links such that least interfering paths can be obtained. The Routing with Inter-flow and Intra-flow Interference Metric (RI^3M), captures both inter-flow and intra-flow interference while balancing link load. We prove the isotonicity of the routing metric through virtual network decomposition. Using the routing metric we derive link disjoint paths to sustain service availability and fault tolerance. Second, we develop an MMF optimization formulation for multipath (splittable) routing where paths are determined using a multi-commodity optimization formulation. The MMF algorithm finds the fairest (lexicographically largest) bandwidth allocation vector for the demands such that the MMF and SINR interference constraints are met. We refer to this algorithm as the Lexicographic MMF Multipath Flow (LMX:M^3F) algorithm. We show that our proposed routing metric and bandwidth allocation formulation improves bandwidth usage, throughput, and delay in comparison to existing interference aware fair bandwidth allocation approaches. This work is presented in Chapter 5.

- **SINR Based Routing Using Distributed Mobility Prediction**: This work is being prepared for publication [73]: The goal of this work is to provide a mobility prediction scheme in relay enhanced broadband wireless networks which is employed to design an interference aware routing algorithm. We first develop a distributed mobility prediction model using the Hidden Markov Model (HMM) which can dynamically track the mobility of users. Our distributed mobility prediction scheme transfers mobility management control to the various relay nodes. Second, we develop a SINR based routing algorithm which uses the location of a mobile user at time t to determine least interfering paths. Specifically, we develop the routing algorithm such that the link costs are derived from the SINR values and the chosen routes have minimum cost (minimum interference). In addition, we ensure that the capacity of each link is not violated when the traffic is routed. Our algorithm is formulated and solved using the minimum cost optimization method. This work is

presented in Chapter 6.

1.5 Outline of Dissertation

This dissertation is organized as follows. The system model used in this research is presented in Chapter 2. The interference aware rate adaptive subcarrier and power allocation using maximum multicommodity flow optimization problem is discussed in Chapter 3. The decoupled optimization of routing and scheduling using SINR constraints problem is discussed in Chapter 4. The problem of multipath routing and max-min fair QoS provisioning under interference constraints is presented in Chapter 5. Chapter 6 presents the problem of SINR based routing using distributed mobility prediction. Finally, conclusions and future work are given in Chapter 7. To better illustrate the interplay among our research accomplishments, the organizational flowchart of this dissertation is depicted in Fig. 1.2.

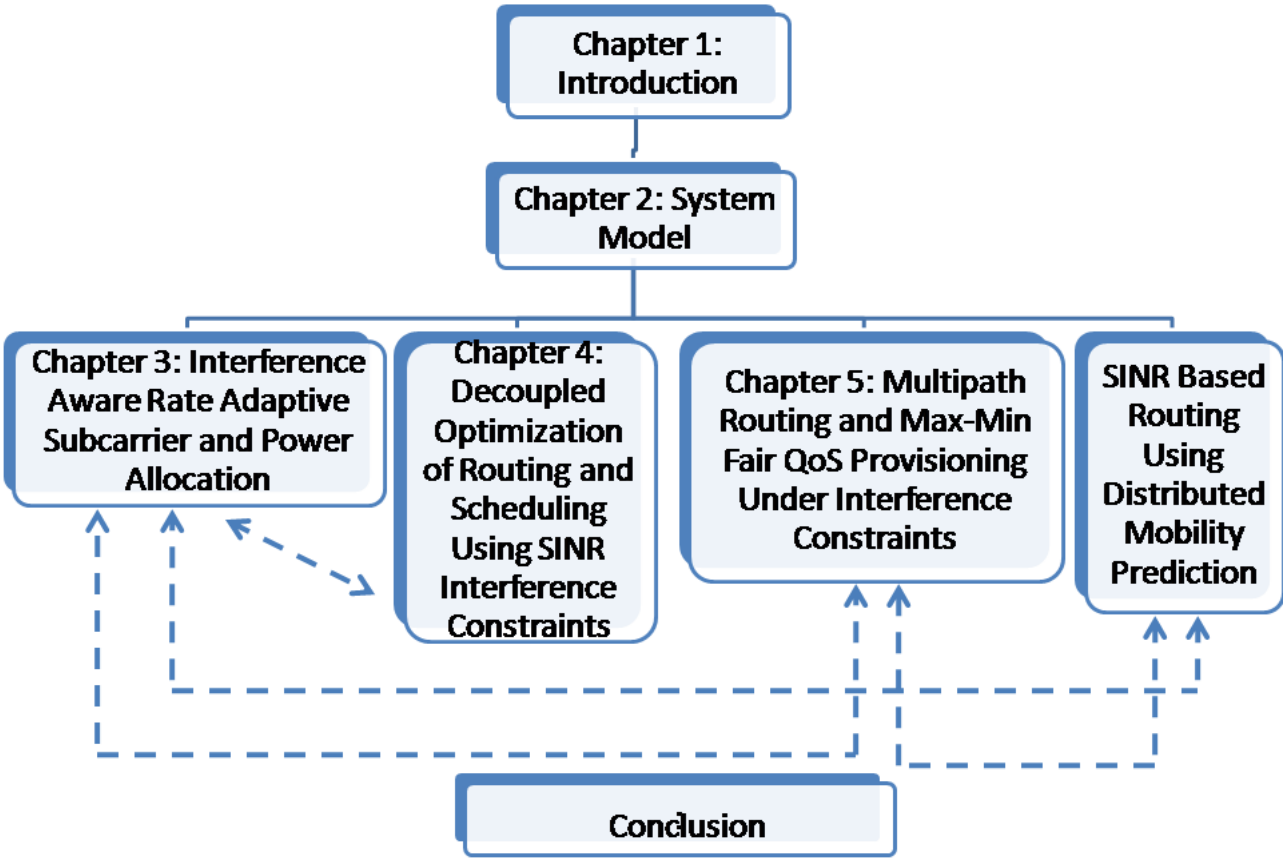


Figure 1.2: Illustration of the organizational flowchart of this dissertation, where solid (dashed) lines represent formal and direct correlations (partial correlations) between two research components, and arrows indicate the flow of communication.

Chapter 2

System Model

In this chapter, we define the system model and the interference model that will be used in the research problems studied in this dissertation.

2.1 Network Architecture

We consider a multihop relay enhanced cellular network (RCN), consisting of a base station, a set $\mathcal{R}$ of fixed relay stations and a set $\mathcal{N}$ of mobile users where each user is connected to either the base station or a relay station. Our network topology is based on the RCN model used in emerging BWA networks [74]. As shown in Fig. 2.1, the network architecture is based on three tiers of wireless devices: 1) user nodes which are the lowest tier have limited functionality (i.e., do not communicate with one another and have no routing capability); 2) relay nodes that route packets between the user and BS is the second tier. They also communicate with one another; and 3) the base station is the highest tier and is connected to the wired infrastructure.

In order to avoid single points of failure (i.e., failure of a relay node which will disrupt traffic flow), the relays are connected in a mesh manner so that multiple paths are available between the user and BS thereby increasing service availability and fault tolerance. Mesh networking is a promising technology for numerous applications (i.e., broadband

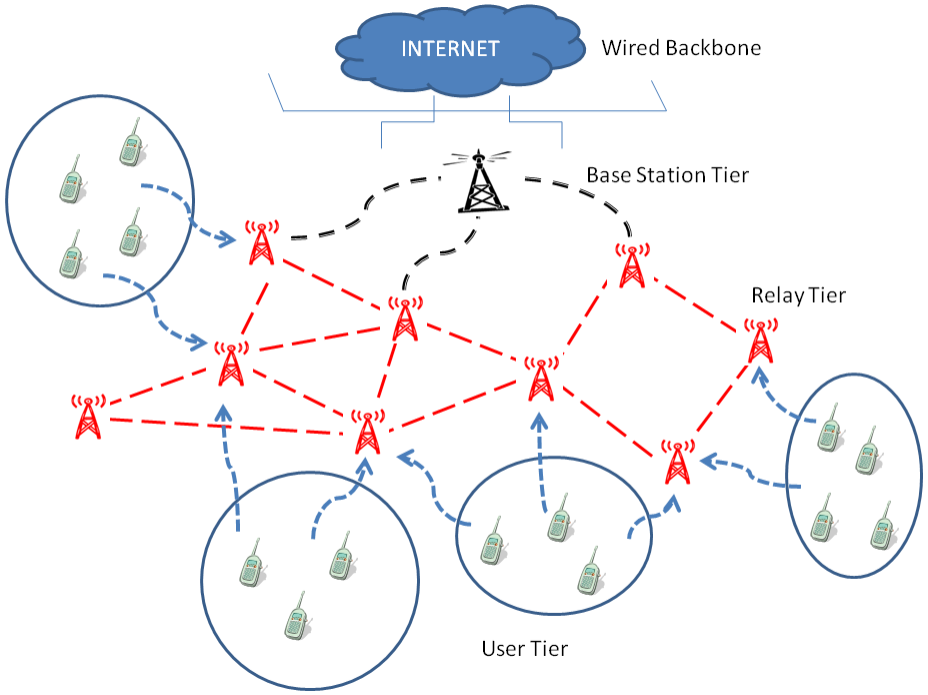


Figure 2.1: System model: a hierarchically organized relay enhanced cellular wireless network

networking) and has attracted significant attention as a cost effective way of deploying wireless broadband networks [75]. The combination of wireless mesh networks and relay networks has been discussed in the literature where the general structure of a mesh network has been incorporated with relaying aspects [76]. Our defined architecture uses a wireless relay network structure that is enhanced with mesh networking capabilities.

We assume that each relay node is equipped with omni-directional transceivers. The relays are used purely for packet forwarding (i.e., relays do not inject traffic into the network). We assume that each user/relay node has a maximum power level, P_{max}, where the P_{max} value is different for the user and relay nodes. We consider uplink traffic only where each user has some traffic to route. In the absence of user mobility, we assume that each node knows the geographic location of all the other nodes in the cell via location discovery schemes [77]. This information is necessary for the receivers to feedback SINR measurements to their respective transmitters.

2.2 Interference Model

We represent the network architecture by a communication graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes (relays, users and BS) and $\mathcal{E}$ is the set of edges. As discussed, in the literature there are two prominent interference models: *protocol model* and *physical model*. The protocol model states that two simultaneous transmissions will interfere only within a certain predefined interference range. The physical interference model is less restrictive than the protocol model. The physical interference model states that a transmission from node i to node j is successful if the SINR at j (the receiver) is above a certain threshold value. Therefore, the SINR is contingent upon other simultaneous transmissions. The SINR for a transmission from node i to node j is given as follows

$$SINR_{ij} = \frac{P_j(i)}{\eta + \sum_{k \in V'} P_j(k)} \geq \beta \tag{2.1}$$

where $P_j(i)$ is the received power at node j due to node i, η is the ambient noise power, V' is the subset of nodes in the network that are transmitting simultaneously, and β is the SINR threshold.

In this dissertation we implement the physical interference model to quantify the interference using protocol interference parameters. To be specific, we use the following variation of the protocol model which is incorporated into the SINR model to determine potentially interfering links. The transmission range represents the maximum distance up to which a packet can be received, while the interference range represents the maximum distance up to which simultaneous transmissions interfere. Let R_T^{max} (r_T^{max}) and R_I^{max} (r_I^{max}) represent the maximum transmission and interference ranges of each relay (user) node, respectively. All relay nodes use the same maximum transmission range (R_T^{max}) as do all the user nodes (r_T^{max}). Each wireless node i (either relay or user node) has a transmission region which is a circle in a 2D plane, centered at i with radius R_T^{max} (r_T^{max}). In the literature, the interference range is usually chosen to be twice as large as

the transmission range which is not necessarily a practical assumption [11]. The actual values of the transmission and interference ranges depend on the transmission power used by the nodes. To provide realistic limits for R_T^{max} (r_T^{max}) and R_I^{max} (r_I^{max}), we use a method called a "reality check" which links the parameters of the physical interference model and the protocol model. The reality check method, introduced in [10], essentially sets a realistic interference range in which links are assumed to interfere. For the protocol model, there are two parameters, the maximum transmission and interference ranges, R_T^{max} and R_I^{max}, respectively. Since the underlying physical layer mechanism is the same, the parameter R_T^{max} (r_T^{max}) should be consistent with the β parameter in the physical model, as shown in Eq. 2.1. Two nodes with distance R_T^{max} (r_T^{max}) should be able to communicate with each other under the maximum transmission power P_{max} and the SINR should be β. As a result, according to [10], R_T^{max} (r_T^{max}) is $\frac{P_{max}}{\beta}$, where P_{max} is the maximum power value for the relay node (user node).

Note that the maximum interference range, R_T^{max} (r_T^{max}), is a parameter introduced by the protocol model and there is no corresponding parameter in the physical model. The only requirement on R_I^{max} (r_I^{max}) is R_I^{max} (r_I^{max})>R_T^{max} (r_T^{max}), i.e., a lower bound for R_I^{max} (r_I^{max}) is R_T^{max} (r_T^{max}). Thus, if we set the interference range to be slightly higher than the transmission range, $R_T^{max}(r_T^{max}) = \frac{P_{max}}{\beta}$, then the solution is more realistic.

2.3 Chapter Summary

In this chapter, we presented the system model that is considered in the various research problems studied in this dissertation. Specifically, we focus on a multihop relay enhanced broadband wireless network consisting of relay nodes, users and a base station. To model interference, we consider SINR as the primary indicator of interference strength in the network. We incorporate protocol interference model parameters into the SINR model via a reality check method to provide a more realistic and practical implementation of

wireless interference.

Chapter 3

Interference Aware Rate Adaptive Subcarrier and Power Allocation Using Maximum Multicommodity Flow Optimization Method

In this work we develop two resource allocation algorithms considering the impact of wireless interference constraints using a novel weighted SINR conflict graph to quantify the interference among the various nodes: 1) interference aware routing using maximum multicommodity flow optimization method; and 2) rate adaptive joint subcarrier and power allocation algorithm under interference and QoS constraints. We exploit spatial reuse to allocate subcarriers in the network and show that an intelligent reuse of resources can improve throughput while mitigating the effect of interference. We provide a heuristic to solve the rate adaptive resource allocation problem. We demonstrate that aggressive spatial reuse and fine tuned-interference modeling has advantages in terms of throughput, end-to-end delay and power distribution. The work presented in this chapter has been published in [67] and [68].

3.1 Problem Preliminaries

We consider the network model given in Chapter 2, Section 2.1 in which all nodes, including users remain static. The wireless propagation effect is modeled by the radio propagation losses. The channel gain of a link will depend on the subcarrier used. Let $G_{n,k}$ be the channel gain of node n on subcarrier k. The OFDMA network under consideration has a total bandwidth of W which is divided into K subcarriers. We assume that the transmissions experience path loss, Rayleigh fading and log normal shadowing[1]. We consider frame by frame resource allocation. A frame is of duration T ms. Channel conditions and user population are assumed to be constant during a time frame. This assumption does not impose a serious restriction since the channel and user statistics are typically not available at a finer granularity than the frame durations. Rayleigh fading is assumed to be flat in each subcarrier and i.i.d for different users and subcarriers. We assume centralized scheduling and assume that the base station can perfectly obtain the channel conditions of all relay stations and user nodes.

3.2 Interference Based Maximum Multicommodity Flow (MCFI)

Given a relay enhanced cellular network, $\mathcal{G}$, each user $n \in \mathcal{N}$ has a traffic demand that must be routed to the BS. In this section, we develop a network optimization formulation that determines the routes to forward traffic of each user to the BS under physical interference constraints such that the maximum possible traffic is routed.

In order to quantify interference using the SINR model, we use a weighted conflict graph. In a conflict graph, a node is introduced for each link in the original network. An edge connects two nodes in the conflict graph if these two links interfere. An edge-based

[1]In the simulations we keep the users fixed but simulate the effects of mobility through Rayleigh fading and log-normal shadowing.

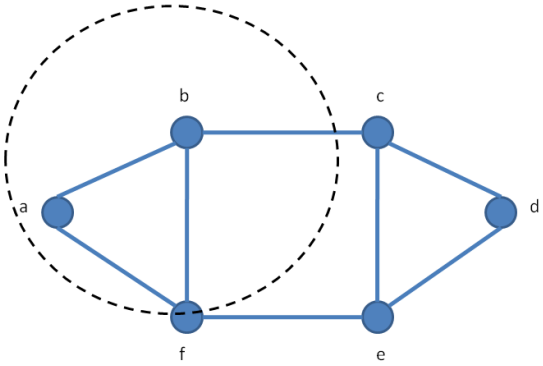
notion of the conflict graph for the physical interference model inserts a weighted edge between two nodes. Consider two links $e_1 = v_1w_1$ and $e_2 = v_2w_2$ (where e_1 and e_2 are the nodes in the conflict graph). We add a weighted edge between e_1 and e_2 if they potentially interfere with each other, where the weight of the link represents the fraction of the maximum permissible noise and interference level at the receiver node of e_2 that is contributed by activity on link e_1.

To determine the interfering links, we use the reality check method discussed in Chapter 2, Section 2.2. As was discussed, this method provides us the transmission range (i.e., $R_T^{max}(r_T^{max}) = \frac{P_{max}}{\beta}$) and thus we can set the interference range to be slightly greater than this value. Given the interference range, all links within that range will interfere and a weighted edge will exist between the interfering nodes in the conflict graph. We define a link weight $w(e_1, e_2)$ as follows

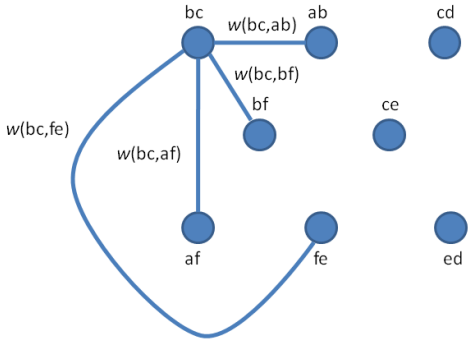
$$w(e_1, e_2) = \begin{cases} 0 & \text{if node } e_1 = e_2 \\ \frac{P_{w_1}(v_1)}{\frac{1}{\beta}P_{w_1}(v_2) - \eta} & \text{, otherwise} \end{cases} \tag{3.1}$$

Fig. 3.1 illustrates the mechanism of creating the weighted conflict graph (WCG) and determining the link weights based on interference range. Fig. 3.1(a) shows the connectivity network $\mathcal{G}$ and the interference range of node b. The conflict graph construction for link (b, c) is given in Fig. 3.1(b) where the potentially interfering links are those within the interference range of node b. The network of Fig. 3.1(b) shows only a partial construction of the WCG.

In order to determine the individual routing paths and compute the maximum achievable throughput, we use the maximum multicommodity flow approach (MCF) [78, 79] with interference constraints. The MCF is a variation of the multicommodity flow problem in which each pair of nodes (user-destination pairs) can send and receive flow simultaneously. The ratio of the flow between the user and BS to the predefined demand for that pair is the throughput. The interference based MCF (MCFI) is defined as fol-



(a) Connectivity graph $\mathcal{G}$



(b) Partial construction of weighted conflict graph, WCG, for link (b, c)

Figure 3.1: Illustration of the weighted conflict graph (WCG) construction from the original connectivity graph

lows [67, 68]. Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$, be a relay enhanced cellular network where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of links in the network. Note that $|\mathcal{V}| = |\mathcal{N}|+|\mathcal{R}|+1$. There is one base station in all networks considered in this paper. There are $|\mathcal{N}|$ user-BS pairs, where $\mathcal{N}$ is simply the set of users (i.e., there are a set of $|\mathcal{N}|$ commodities in the network). Each user is associated with a certain traffic demand that must be routed to the BS. We denote x_{ij}^n as the amount of flow from the nth commodity over link (i, j), normalized with respect to the capacity of the link. The link capacity is defined as

$$u_{ij} = Blog_2(1 + SINR_{ij}) \tag{3.2}$$

where B is the bandwidth of each subcarrier. The $SINR_{ij}$ is defined as in Eq. 2.1. We let f^n be the flow originating from user node n. We denote $Int(i, j)$ to be the set of links that interfere with link (i, j) according to the weighted conflict graph. The MCFI is formulated as follows

$$\text{maximize} \quad \sum_{n\in\mathcal{N}} f^n \tag{3.3}$$

subject to

$$\sum_{(i,j)\in\mathcal{E}} x_{ij}^n - \sum_{(j,i)\in\mathcal{E}} x_{ji}^n = \begin{cases} f^n & \text{if} \quad i = n \\ -f^n & \text{if} \quad i = BS \\ 0 & , \quad \text{otherwise} \end{cases} \quad ,\forall n \in \mathcal{N} \tag{3.4}$$

$$\sum_{n\in\mathcal{N}} x_{ij}^n + \sum_{(p,q)\in Int(i,j)} \sum_{n\in\mathcal{N}} x_{pq}^n \leq 1, \forall (i, j) \in \mathcal{E} \tag{3.5}$$

$$\sum_{n\in\mathcal{N}} [\sum_{(i,j)\in\mathcal{E}} x_{ij}^n + \sum_{(i,j)\in\mathcal{E}} x_{ji}^n] \leq 1 \tag{3.6}$$

$$x_{ij}^n \geq 0, \forall (i, j) \in \mathcal{E}, \forall n \in \mathcal{N} \tag{3.7}$$

The first constraint (Eq. 3.4) represents the flow conservation constraints at each node for each commodity. Eq. 3.5 is the link capacity constraint dictated by the interference model and the weighted conflict graph. The constraint in Eq. 3.6 is the node capacity

constraint in which the sum of the ingoing and outgoing flows should be less than the channel capacity. The linear program described above leads to a multicommodity flow problem which uses multiple paths to route each commodity from source to destination. In many wireless network protocols, however, data are generally routed along a single path to avoid some side-effects that occur due to multi-path routing. In single path routing, each edge can either carry the full traffic for a given connection or none of it. This constraint is given in Eq. 3.8.

$$x_{ij}^n = f^n \cdot y_{ij}^n, \forall n \in \mathcal{N}, \forall (i,j) \in \mathcal{E} \tag{3.8}$$

The variable y_{ij}^n is a boolean variable which is set to 1 if the edge carries the traffic for the nth connection and 0 otherwise. The single path routing approach to route flows is based on the weighted conflict graph (i.e., the path with least cost (least interference) are chosen for each user).

3.3 Joint Subcarrier and Power Allocation Under Time and QoS Constraints

3.3.1 Optimization Formulation

In OFDMA networks, the BS controls how subcarriers are allocated and to which links they are assigned. In this work we exploit spatial reuse and analyze the performance benefits of having such reuse. In order to ensure that links using the same subcarrier do not strongly interfere (spatial reuse), the subcarriers should be allocated to links which are far away from each other. Within the interference range of a node, n, $R_I^{max}(n)$, there are a set of nodes which we denote as the dominant interferers. Their proximity to n leads to a high probability that a transmission from any of them will result in interference at n. We denote the set of dominant interferers as $D_I(n)$. Note that $n \in D_I(n)$. The

set of links emanating from each node within $D_I(n)$ is called the interference link set, $L_I(n)$. Also all links emanating from n will also be in $L_I(n)$. In addition, we define the spatial reuse factor as λ_k which is the number of times each subcarrier is used within a relay enhanced cell (λ_k is different for each subcarrier). Note that $k \in K$, where K is the total number of subcarriers. Furthermore, we define the value λ_{max} to be the maximum number of times a subcarrier is allowed to be reused within the cell (i.e., each λ_k can not be greater than λ_{max}).

We aim to assign unique subcarriers to all links within the interference range of each node (i.e., links within $L_I(n)$ for all n). Outside of the interference range, reuse of subcarriers is allowed. The subcarrier assignment scheme is captured by the interference constraint given below.

Interference Constraint: Let (u, v) and (i, j) be two distinct links and let $\Psi(\cdot)$ denote the subcarrier assignment of a link. We define the interference constraint for subcarrier allocation as follows: For a given node n

$$\Psi(u,v) \neq \Psi(i,j), \quad \forall \quad (u,v) \in L_I(n) \quad \text{and} \quad (i,j) \in L_I(n) \quad \text{and} \quad (u,v) \neq (i,j) \quad (3.9)$$

The above constraint states that subcarriers assigned to links within the interference link set of each relay must be unique (each subcarrier is allocated only once within the interference link set). Fig. 3.2 shows an illustration of the interference constraint where links (u, v) and (i, j), both within $R_I(n)$, will be assigned different subcarriers.

Each link is allocated subcarriers from the subcarrier set K. To keep track of the available subcarriers in the interference link set of each node, we define the available subcarrier set for each link as follows. The available subcarrier set denoted as $\mathcal{A}(l)$ for link l at a particular time is the set of subcarriers which have not been allocated to any link in the interference link set of node n, $L_I(n)$.

We allocate subcarriers using the interference constraint above while jointly allocating

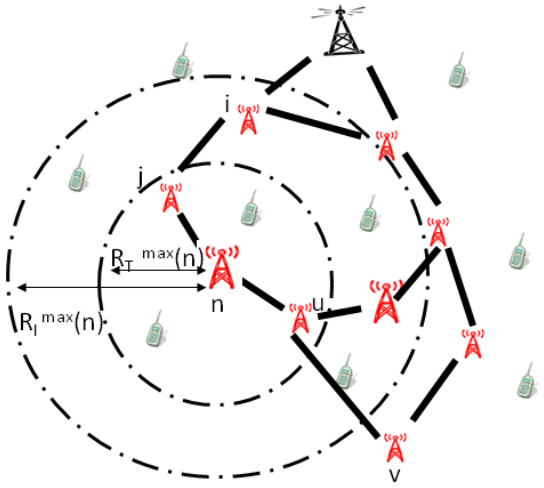


Figure 3.2: Illustration of the subcarrier allocation interference constraint

power to the nodes according to time and QoS constraints. The rate adaptive resource allocation technique under consideration in this paper jointly solves both problems [68]. The optimization problem can be formulated as follows

$$\text{maximize} \sum_{t=1}^{T} \sum_{n=1}^{\mathcal{V}} \sum_{k \in A_n(l)} log_2(1 + P_{n,k}(t)\gamma_{n,k}) \tag{3.10}$$

subject to

$$\sum_{t=1}^{T} \sum_{n=1}^{\mathcal{V}} \sum_{k=1}^{K} P_{n,k}(t) \leq P_{total} \tag{3.11}$$

$$P_{n,k}(t) \geq 0 \tag{3.12}$$

$$\sum_{t=1}^{T} \sum_{n=1}^{\mathcal{V}} \sum_{k=1}^{K} y_{n,k}(t) + \sum_{t=1}^{T} \sum_{m=1}^{\mathcal{V}} \sum_{k=1}^{K} y_{m,k}(t) \leq 1 \tag{3.13}$$

$$\frac{G_{n,k}P_{n,k}(t) + (1 - y_{n,k}(t))}{\xi} \geq \beta, \forall n \in \mathcal{V}, \forall t \in T, \forall k \in K \tag{3.14}$$

$$y_{n,k}(t) \leq \frac{P_{n,k}(t)G_{n,k}}{\eta B}, \forall n \in \mathcal{V}, \forall t \in T, \forall k \in K \tag{3.15}$$

$$A_1(l) \cup A_2(l) \cup ... \cup A_{\mathcal{V}}(l) = \{1, 2, ..., K\} \tag{3.16}$$

$$y_{n,k}(t) \in \{0,1\} \tag{3.17}$$

$$R_1 : R_2 : ... : R_{\mathcal{V}} = \alpha_1 : \alpha_2 : ... : \alpha_{\mathcal{V}} \tag{3.18}$$

where $\mathcal{V}$ is the total number of nodes (users and relays), K is the total number of subcarriers, P_{total} is the overall available power and $P_{n,k}(t)$ is the power allocated to the nth node on the kth subcarrier. This signal power is split across the different subcarriers that node n uses. $\gamma_{n,k} = \frac{|G_{n,k}|^2}{\frac{\eta W}{K}}$ is the channel gain to noise power ratio for the nth node on the kth subcarrier. $G_{n,k}$ is the channel gain for the nth node on the kth subcarrier, η is the noise power, and W is the overall available bandwidth. $A_n(l)$ is the set of all subcarriers allocated to the nth node. The rate of the nth node, R_n, is defined as $\sum_{k \in A_n(l)} log_2(1 + P_{n,k}(t)\gamma_{n,k})$ (as given in the objective function of Eq. 3.10). $\{\alpha_1, \alpha_2, ..., \alpha_{\mathcal{V}}\}$ is the set of predetermined constants to ensure proportional fairness amongst nodes.

Constraints in Eqs. 3.13-3.15 reflect the scheduling constraints. Because we use spatial reuse when assigning subcarriers, we must ensure that the transmissions on the same subcarrier do not interfere if scheduled in the same time slot. Therefore, we check if these transmissions contribute to the SINR and if so, schedule these transmissions in different time slots. We modify the SINR equation given in Eq. 2.1 to incorporate the effect of transmissions on the same subcarriers. The modified SINR is given as

$$SINR^t_{n,k} = \frac{G_{n,k}P_{n,k}(t)}{\eta + \sum_{m \neq n} X_m G_{m,k} P_{m,k}(t) \chi_{m,k}} \geq \beta \tag{3.19}$$

where X_m is a binary variable which denotes whether node m is transmitting or not. $\chi_{m,k}$, also a binary variable, denotes whether node m is transmitting on the same subcarrier k as node n. Note that those links that transmit on different subcarriers inherently do not interfere because of the orthogonality of OFDMA. For the scheduling constraints, we introduce a binary variable, $y_{n,k}(t)$, which is equal to 1 if a node n, scheduled in time slot t, transmits on subcarrier k and 0 otherwise. The constraint in Eq. 3.13 states that two adjacent links must be assigned different time slots while Eq. 3.14 expresses the required

SINR threshold that should be satisfied to have a successful transmission. The term $1-y_{n,k}(t)$ ensures that the SINR inequality is satisfied when node n does not transmit in time slot t. ξ denotes the denominator of Eq. 3.19. Eq. 3.15 is based on the assumption that all links in the network satisfy the SINR constraint when there are no concurrent transmissions.

3.3.2 Proposed Heuristic Solution

To solve the rate adaptive joint subcarrier and power allocation optimization problem presented in Section 3.3.1 we propose a heuristic solution. Each of the K subcarriers is to be allocated to at least one of the $\mathcal{V}$ nodes and the power allocated to each of the $\mathcal{V}$ nodes is to be optimized. This means that $\mathcal{V}+K$ parameters need to be optimized to achieve the optimal solution. Power allocation amongst subcarriers belonging to a particular node is achieved through waterfilling. According to [29], the optimization problem given in Eq. 3.10 can be simplified into one that has K optimization parameters by assuming equal power allocation to all subcarriers, i.e.,

$$P_{n,k} = \begin{cases} \frac{P_{total}}{K} & \text{if } k \in A_n(l) \\ 0 & \text{otherwise} \end{cases} \tag{3.20}$$

for all $k = 1, 2, ..., K$ and $n = 1, 2, ..., \mathcal{V}$. Since the power allocated to each subcarrier is fixed, optimization now involves assigning the K subcarriers to $\mathcal{V}$ nodes. In our proposed solution, optimization of the $K+\mathcal{V}$ parameters is carried out by alternating between subcarrier and power allocation. We use waterfilling for each node. When a subcarrier is allocated to a node, the power allocated to the node is incremented by $\frac{P_{total}}{K}$, i.e., the power allocated to each node is proportional to the number of subcarriers currently allocated to that node. The node's rate is also updated assuming that waterfilling is used. This updated rate information is used in the allocation of the remaining subcarriers. Thus, the gain from the waterfilling is seen in the subcarrier allocation stage by all the nodes

resulting in higher rates. Let T be the number of time slots and let T_n be the number of time slots assigned to the nth node. Let t_k^j be the time slot of subcarrier k of time index j.

The joint subcarrier and power allocation strategy is as follows

1. Initialize $A(l) = \{1, 2, 3, ..., K\}$, $T_n = \emptyset$, $\lambda_k = \emptyset$
2. $\forall\ n = 1$ to $\mathcal{V}$, $A_n(l) = \emptyset$, $P_n(t) = 0$
3. $\forall\ n = 1$ to $\mathcal{V}$,
   (a) $\gamma_n = max_k \gamma_{n,k}$, $\forall\ k \in A(l)$
   (b) $A_n(l) = A_n(l) \cup \{k\}$, $P_n(t) = P_n(t) + \frac{P_{total}}{K}$
   (c) $R_n = log_2(1 + P_n(t)\gamma(n))$
   (d) $A(l) = A(l) - \{k\}$
   (e) $\lambda_k + +$
   (f) Find a slot $t_k^j \in T$ so that the SINR is satisfied according to Eq. 3.19
   (g) $T_n \leftarrow T_n \cup \{t_k^j\}$
4. While $A(l) \neq \emptyset$,
   (a) Find i such that $\frac{R_i}{\alpha_i} \leq \frac{R_n}{\alpha_n}$
   (b) For the above i, find k such that $\gamma_{i,k} \geq \gamma_{i,j}$, $\forall (k, j) \in A(l)$
   (c) $A_i(l) = A_i(l) \cup \{k\}$, $P_i(t) = P_i(t) + \frac{P_{total}}{K}$
   (d) $A(l) = A(l) - \{k\}$
   (e) $\lambda_k + +$
   (f) Find a slot $t_k^j \in T$ so that the SINR is satisfied according to Eq. 3.19
   (g) $T_n \leftarrow T_n \cup \{t_k^j\}$

(h) $R_i = \sum_{k \in A_i(l)} log_2(1+P_{i,k}(t)\gamma_{i,k})$ where $P_{i,k}(t) = (\gamma - \frac{1}{\gamma_{i,k}})^+$ and $\sum_{n \in A_i(l)} P_{i,k}(t) = P_i(t)$

The $f(x) = (x)^+$ operator indicates that $f(x) = 0$ when $x < 0$ and $f(x) = x$ when $x \geq 0$. The algorithm described above uses the equation in 4(h) for the rate updates. The proposed algorithm requires waterfilling to be performed $K - \mathcal{V}$ times. In the simulations, given in Section 3.5.2, we use waterfilling in Step 4(h) after a subcarrier is allocated to a node. This is for the purpose of evaluating the performance of our proposed algorithm against existing algorithms.

3.4 Discussion of Computational Complexity

The interference based maximum multicommodity flow problem is solved based on an ϵ-approximation algorithm given in [80]. The running time of this algorithm is given as $\bigcirc(\epsilon^{-2}m^{-2})$, where m denotes the number of links in the network and ϵ is any error parameter greater than 0. The joint subcarrier and power allocation algorithm, RASP, has a complexity of $\bigcirc((NK + KlogK) * (K - \mathcal{V}))$ where K is the number of subcarriers and N is the number of nodes in the networks (users plus relays). $KN+KlogK$ is derived from [29] and $K - \mathcal{V}$ is the number of times waterfilling must be performed before all the subcarriers are allocated.

3.5 Performance Evaluation

3.5.1 Simulation Model and Performance Metrics

We consider a relay enhanced multihop cellular network, $\mathcal{G}$, in a $900m \times 900m$ region. Each user generates traffic and the flows are routed towards the base station. There is no downlink traffic generated. We use NS-2 to simulate the networks. The base station is located in the center of the network. Locations for the set of relay nodes that

form the mesh network are randomly generated. Locations for the user nodes are also randomly generated. We assume that the BS and relays have an infinite buffer, thus eliminating complications due to buffer overflow. The following numerical parameters are used in the simulations: System Bandwidth (W) = 1MHz, Number of subcarriers = 256 and 512, additive white gaussian noise (AWGN) (η) = -90bBW/Hz, Pathloss exponent (LOS/NLOS) = 2.35/3.76, P_{total} = 39 dBm, Frame length = 4ms, Time slot length = 0.1ms. As mentioned in Section 3.1, we use Rayleigh fading for the subcarriers. The maximum transmit power of each relay is 35dBm and the maximum transmit power of each user is 24dBm. Packets are scheduled using a first in first out (FIFO) priority scheme. We let $\alpha_1 : \alpha_2 : ... : \alpha_{\mathcal{V}} = 1 : 1 : ...$ so that the overall rate is maximized while trying to achieve equal rate for all nodes.

To evaluate the performance of our algorithms, we study the following performance metrics: 1) throughput generated by the MCFI; 2) end-to-end delay of the MCFI routing procedure; 3) effect of λ_{max} on subcarrier allocation; 4) throughput generated by the joint subcarrier and power allocation algorithm; and 5) power distribution versus varying number of nodes.

As benchmarks, we compare the MCFI algorithm with two interference aware routing procedures in the literature. First, the algorithm in [81] develops a routing metric where a node calculates the SINR to its neighboring links based on a 2-Hop interference Estimation AlgoRithm (2-HEAR). Second, the algorithm given in [82] uses a multicommodity flow approach to routing and uses the protocol model to capture the interference constraints. We denote this algorithm as MCF-Protocol in the simulation graphs. In addition, as benchmarks for comparison of our proposed rate adaptive joint subcarrier and power allocation algorithm, we compare with the two prominent rate adaptive allocation techniques given in [29] and [83]. In [29], power is allocated uniformly across all subcarriers used by a node. In [83] the power and subcarrier allocation problems are solved separately as individual problems rather than jointly as in this work.

3.5.2 Simulation Results and Discussion

We first evaluate the routing procedure of our MCFI formulation in terms of throughput. The throughput obtained by the MCFI is the overall normalized system throughput obtained under SINR interference constraints. The normalized throughput is determined as the absolute rate (in bits per second) divided by the channel capacity given by Shannon's capacity formula, $Blog_2(1 + SINR_{ij})$. Thus, the normalized throughput is impacted by the link SINR (interference) and subcarrier bandwidth. We run simulations on networks with 46 nodes (40 users, 6 relays, 85 links), 24 nodes (20 users, 4 relays nodes, 62 links) and 12 nodes (10 users, 2 relays, 44 links). Each network has 1 base station. The results are shown in Fig. 3.3 and are averaged over 20 simulations per network. The proposed MCFI algorithm achieves the highest possible throughput compared to the other algorithms. We can justify the better performance of our algorithm as follows: In both the 2-HEAR and MCF-Protocol algorithms, the routing paths are formed using incomplete interference information. In 2-HEAR the SINR calculated by each node only includes those nodes within a 2-hop range which means that even if interference beyond this range occurs, it is not captured in the routing metric. In the MCF-Protocol algorithm, interference is gauged using a distance based method (random interference range is used) which restricts the possibility that transmissions can occur even if they are close to each other as long as the signal strengths do not interfere. In our case, the MCFI algorithm quantifies the interference using a more refined interference range which may be less or more than the 2-hop range.

We next evaluate the ability of the our MCFI routing approach to decrease end-to-end delay (amount of time it takes to deliver packets from user to the BS). Based on the calculation of SINR at each receiver, the arrived packets are determined to be successfully accepted or dropped. For a given SINR value, two error modeling approaches are most commonly used in network simulations [84]: the SINR threshold (SINRT) based method and Packet Error Ratio (PER) based method. With the SINRT based method,

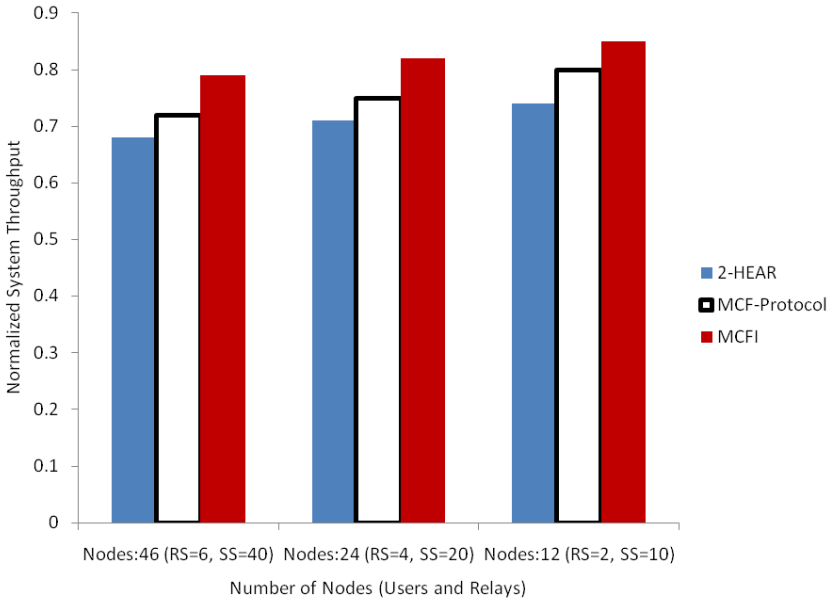


Figure 3.3: Throughput results of the MCFI algorithm compared to 2-HEAR and MCF-Protocol algorithms

packet error is determined by directly comparing the received SINR with the SINRT. With PER based method, the packet error decision is made probabilistically based on the PER, which can be yielded from the theoretical calculation, link layer simulation or experimental measurement. Generally, it is considered that the PER based method is simpler and more accurate than the SINRT method in a simulation setting; it is also readily available in NS-2. Thus, in our simulations we use the PER model to quantify the packet losses. Because of the fact that the MCFI algorithm captures interference more accurately than the other two algorithms, dropping of packets due to interference is limited. Therefore retransmission time is decreased, thereby improving end-to-end delay. The results are shown in Fig. 3.4 and Fig. 3.5 for networks with 46 nodes (40 users, 6 relays, 85 links) and 24 nodes (20 users, 4 relays nodes, 62 links), respectively[2]. The results are averaged over 20 simulations. As expected, when traffic load decreases,

[2]The solutions presented for MCFI and RASP are limited by the complexity of the individual algorithms and the SINR computation. In terms of scalability, our proposed solutions are adequate for moderately sized networks with 80-100 nodes.

the delay decreases for all three algorithms. However, the MCFI algorithm has the lowest end-to-end delay when compared to 2-HEAR and MCF-Protocol. We can conclude that the MCFI algorithm effectively incorporates interference constraints into the maximum multicommodity flow approach and thereby provides least interfering paths while maintaining a high throughput.

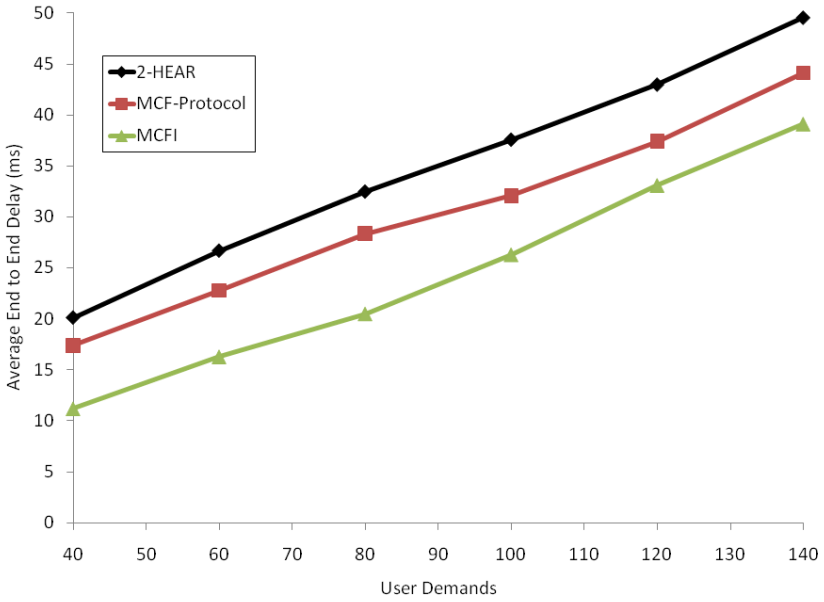


Figure 3.4: End-to-end delay comparison for networks with 46 nodes (6 relays and 40 users) using MCFI, MCF-Protocol and 2-Hear

Finally, we evaluate the effectiveness of the spatial reuse factor in the subcarrier allocation. Specifically, we evaluate the effect of the spatial reuse factor λ_{max} (maximum number of times a subcarrier can be used within a cell). To show how λ_{max} impacts the system performance, we show the total transmission rate for the flows in the network versus varying λ_{max} values. We run simulations using 256 and 512 subcarriers in networks with 50 nodes (46 user nodes, 4 relay nodes) and 100 nodes (90 user nodes, 10 relays). Note that each network has 1 base station. We use a 64-QAM modulation strategy. The results are shown in Figs. 3.6(a) and (b). The λ_{max} value ranges from 1 (no spatial reuse; all subcarriers used only once) to 10. From the results we see that moderate spatial reuse of subcarriers can considerably enhance the performance compared to the case where

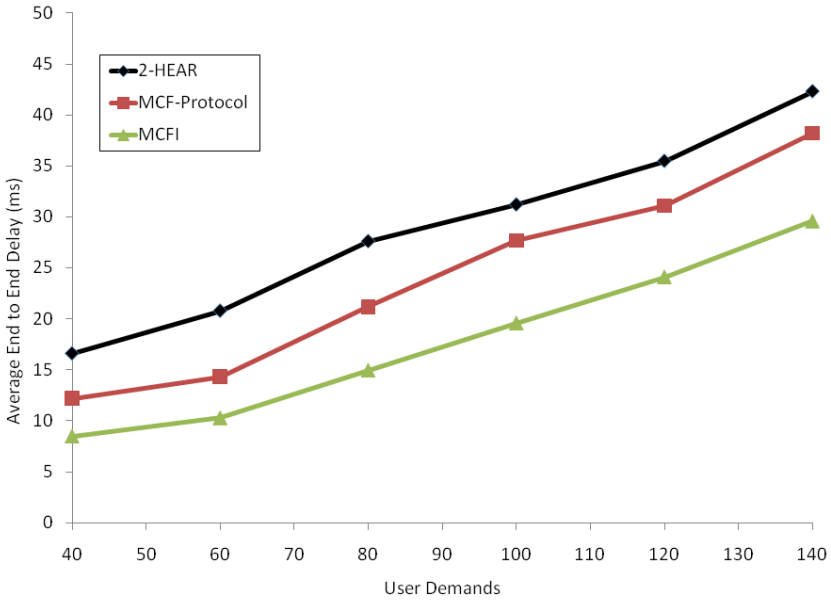
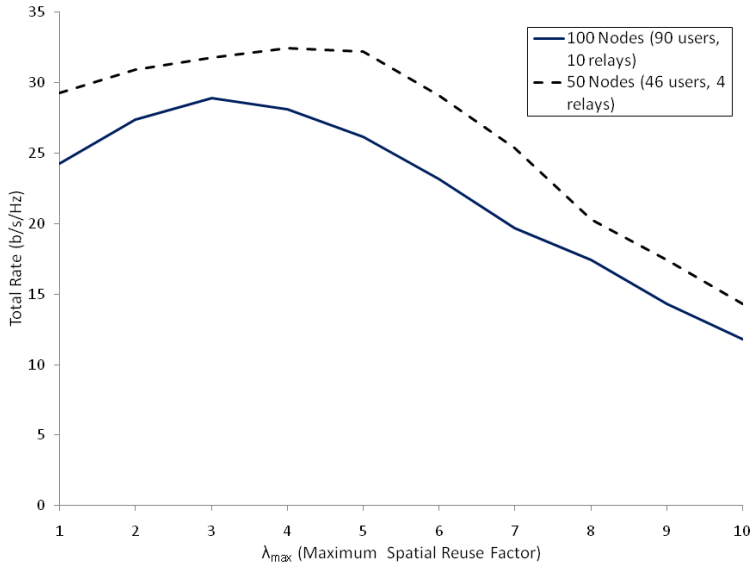


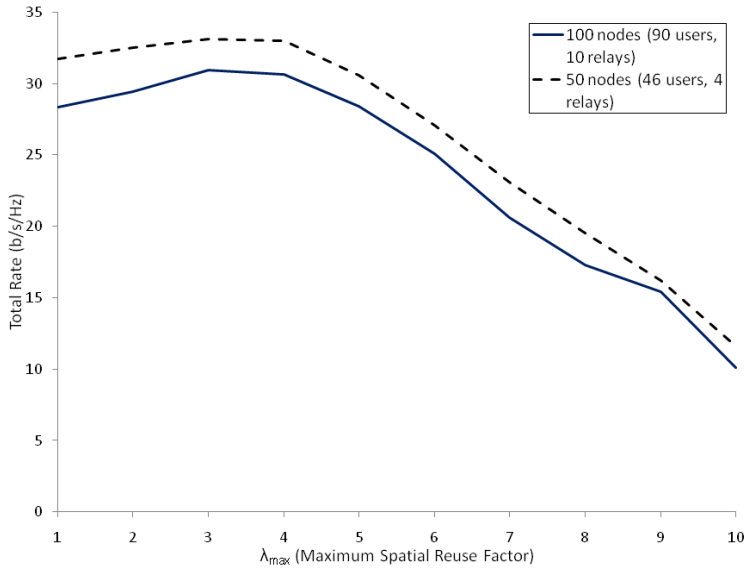
Figure 3.5: End-to-end delay comparison for networks with 24 nodes (4 relays and 20 users) using MCFI, MCF-Protocol and 2-Hear

no spatial reuse is used. In Fig. 3.6(a) we see that from λ_{max}=1 and $\lfloor\lambda_{max}\rfloor$=4, there is a 6.32% increase in performance for the 50 node case. Note that $\lfloor\lambda_{max}\rfloor$=4 is the maximum spatial reuse factor at which performance begins to decrease. In Fig. 3.6(b) there is a similar increase of 10% in performance for the 50 node case. The 50 node cases can handle more spatial reuse (i.e., the $\lfloor\lambda_{max}\rfloor$ for the 50 node cases before performance deteriorates is higher than for the 100 node cases) because the number of nodes and links is less, thus inherently they are less susceptible to the same level of interference as the 100 node networks. This is an indication that our subcarrier allocation strategy does mitigate interference while improving throughput by reusing subcarriers. On the other hand, once the λ_{max} value reaches a certain level, it becomes evident that there is drop in system performance indicating that spatial reuse is no longer a benefit (links begin to strongly interfere). Thus, an appropriate amount of spatial reuse in subcarrier assignment is tolerable.

We next evaluate the effectiveness of our joint subcarrier and power allocation algorithm in terms of throughput versus varying number of nodes. We compare our results



(a) 256 subcarriers



(b) 512 subcarriers

Figure 3.6: Effect of spatial reuse of the subcarrier allocation on the total rate

(referred to as RASP (Rate Adaptive joint allocation of Subcarriers and Power) in the simulation graphs) with the algorithms in [29] and [83], denoted as Rhee and Evans, respectively, in the simulation graphs. The results are shown in Fig. 3.7. We see that our algorithm performs better than that of other two approaches for the following rea-

sons: In [29], though proportional fairness is achieved, the frequency selective nature of a node's channel is ignored by allocating power uniformly across all subcarriers belonging to a particular node. The algorithm in [83] takes a two step approach to solve the subcarrier and power allocation problem rather than solve it jointly.

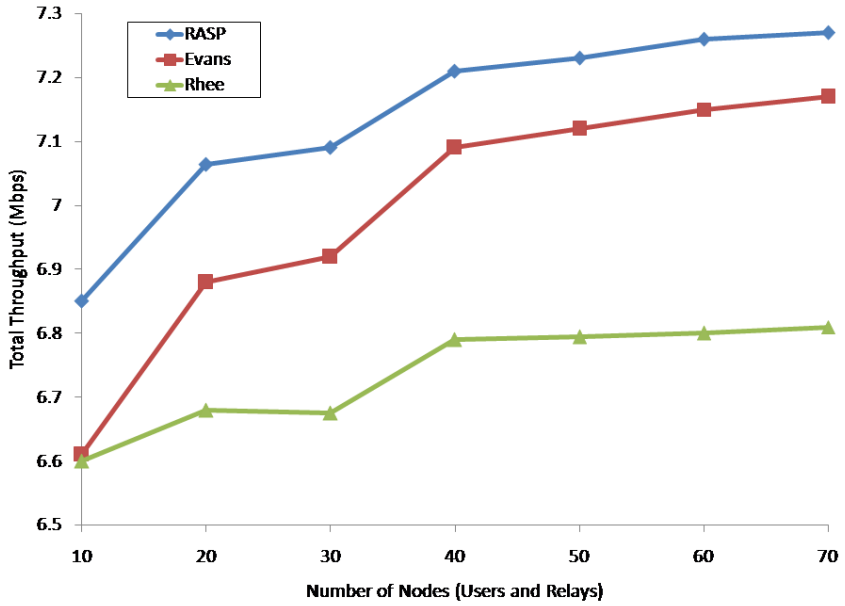


Figure 3.7: Comparison of total throughput versus the number of nodes for the rate adaptive joint allocation of subcarriers and power algorithm with relevant counterparts

Finally, we evaluate the power distributions for the case of varying nodes. We once again compare our proposed approach with those in [29] and [83]. The results are shown in Fig. 3.8 and are an average over 20 simulations. The power distributions across varying number of nodes is less using our approach than with approaches of Rhee and Evans. It can be seen that the performance of our approach and Evans are closer than that of Rhee particularly because with Rhee's approach, the power is uniformly allocated.

3.6 Chapter Summary

In this chapter, we developed a framework for interference aware rate adaptive resource allocation for relay enhanced wireless networks. This work consisted of two parts. The

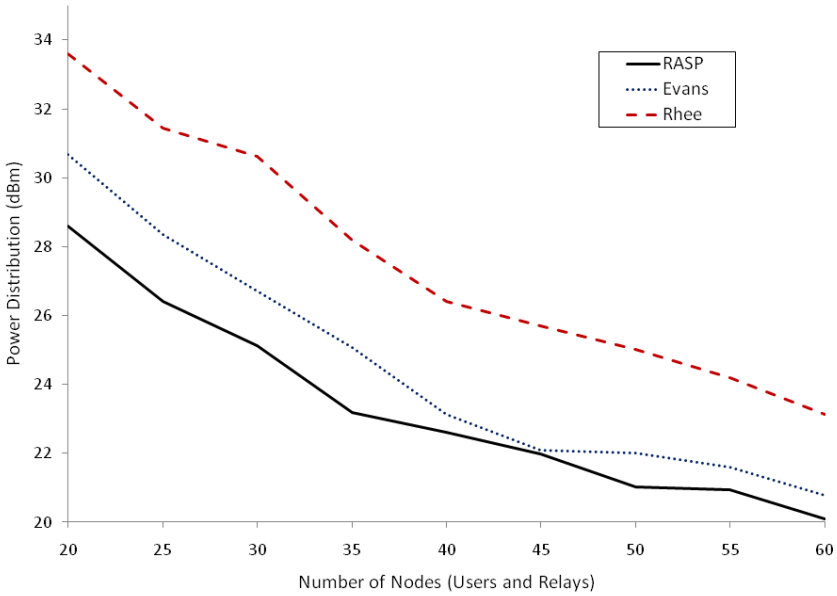


Figure 3.8: Comparison of power distribution over varying number of nodes

first dealt with an interference based maximum multicommodity flow (MCFI) optimization formulation in which a novel approach is developed to solve the traditional MCF problem (used in wired networks) in a wireless environment. Specifically, we employ weighted conflict graphs to quantify interference on the links of simultaneous transmissions. Furthermore, we integrate this interference information into our optimization formulation for the MCFI. Second, we proposed a rate adaptive joint subcarrier and power allocation algorithm considering SINR induced interference. Our subcarrier allocation procedure employs the concept of spatial reuse of geographically separated transmissions while simultaneously mitigating any resulting interference. We proposed an optimization formulation to jointly assign subcarriers to links while allocating power to nodes under time (scheduling) and QoS (rate) constraints. Due to computational complexity, the rate adaptive resource allocation optimization is solved using a sub-optimal heuristic. We showed that subcarrier allocation through spatial reuse improves throughput to a peak value after which performance degradation is observed. Thus, a tradeoff exists. Simulation results also show that our proposed rate adaptive subcarrier and power allocation

algorithm performs better than existing rate adaptive allocation algorithms counterparts in the literature [29, 83].

Chapter 4

Decoupled Optimization of Routing and Scheduling Using SINR Interference Constraints

Wireless routing policies that maximize the aggregate throughput is contingent upon efficient resource utilization. Optimizing the amount of traffic that is possible to route in a given system with several source-destination pairs can be achieved using maximum concurrent flows (MCF). Links in a wireless network can not be associated with a fixed capacity due to the unpredictable nature of the wireless channel. Therefore, link capacities must incorporate the effects of interference due to simultaneous transmissions.

The flow of traffic in the network determines the demand on individual links which in turn determines adequate link schedules. Concurrent transmissions increase system efficiency but at the same time can lead to erroneous reception at the receiver if SINR is too weak [85]. The spatial time division multiple access (STDMA) scheduling scheme achieves higher capacities by allowing TDMA time slots to be shared by simultaneous transmissions that are geographically separated (and therefore have less interference) such that the resulting interference is minimized [86]. The problem of STDMA scheduling has

been shown to be NP-complete [87]. The work presented in this chapter has been reported in [69].

4.1 Spatial TDMA and SINR Induced Interference

STDMA [86] allows TDMA time slots to be shared by simultaneous transmissions such that the resulting interference is minimized. Traditionally, scheduling techniques using spatial reuse have been designed using the protocol interference model and graph based techniques [88, 89]. However, these reuse schedules may result in interference in terms of SINR and therefore deterioration in network performance [90]. To fill this void, the seminal work of [87] has explicitly taken into account the SINR thresholds to construct minimum frame length schedules. However, the effect of routing decisions particularly in the presence of interference has been neglected. In addition, all the above mentioned works consider single channel networks.

In joint optimization of routing and scheduling [91, 92], the protocol interference model is used to determine interference constraints on link capacity. Because the protocol model depends on distance, interfering links are determined on the assumption that transmissions will interfere within some range. Using such a model, joint routing and scheduling optimization techniques define schedulable flows. A schedulable flow is one that can be scheduled interference free while achieving a flow assignment. Finding schedulable flow constraints is applicable in 802.11 based networks where RTS/CTS models determine interference [91, 92]. Since SINR is a more realistic indicator of interference, the physical interference model is preferred but much more difficult to implement. Therefore a decoupled optimization of routing and scheduling is preferred and we believe to be more practical in large scale networks.

We develop two optimization formulations, the first an interference aware maximum concurrent flow formulation for routing (MCF-ROPT) and a STDMA scheduling op-

timization that schedules flows in spatially reused time slots such that interference is mitigated (SM-TSS) [69].

4.2 Problem Preliminaries

This work considers the network architecture discussed in Chapter 2 and the OFDMA physical layer mechanism used in Chapter 3. We assume that each user is connected to at most one relay or directly to the BS and no connections exist between users. Note that relays do not inject traffic into the network. We denote G_{ij} to be the channel gain between transmitter i and receiver node j. We use a deterministic fading model, $G_{ij} = (d_{ij})^{-\kappa}$. Here d_{ij} is the distance between node i and node j, and κ is the path loss exponent. The channel gain of a link will depend on the subcarrier used. Let G_{ij}^c be the channel gain of link (i, j) on subcarrier c. It is assumed that each link knows the channel gains on all subcarriers by using some estimation method, i.e. link (i, j) knows G_{ij}^c for all c. As we are looking at a single cell scenario, this estimation is a valid assumption, given the relative size of the cell. Each link, (i, j) has a gain vector $\mathbf{G} = [G_{ij}^1, G_{ij}^2, \ldots, G_{ij}^C]$ which contains the gain values of all subcarriers for the link. Here, C is the number of available subcarriers.

4.3 Maximum Concurrent Flow Routing Optimization Using SINR (MCF-ROPT)

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a relay enhanced network where $|\mathcal{V}| = \mathcal{N} + \mathcal{R} + 1$ and $\mathcal{E}$ is the set of edges. On this network, we perform the subcarrier allocation algorithm discussed in Chapter 3, Section 3.3 in which subcarriers are spatially reused. Given this scenario, we compute the maximum achievable throughput using the maximum concurrent flow approach (MCF) [78]. The MCF is defined on network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with link capacities

$u_{ij} > 0, \forall (i,j) \in \mathcal{E}$, and $|\mathcal{N}|$ source-destination terminal pairs. The value $|\mathcal{N}|$ denotes the number of commodities in the network which in our case is represented by the number of users in the system. Each commodity i is associated with a certain amount of demand, d_i. In our case the demand is defined as the traffic that each user must route to the base station (destination). The problem is to find flows f_i from s_i to t_i that satisfy node conservation constraints and meet some objective function criteria so that the sum of flows on any edge does not exceed the capacity of the edge. The objective is to obtain the maximum possible fraction of demand that it is possible to route concurrently over the network without violating link capacity constraints. The maximum possible fraction, denoted as α, is defined as the maximum throughput of the network. Since each user has some demand to route, we can associate this as a commodity. Let u_{ij} be the capacity of link (i,j) and P_i be the set of paths from user terminal i to the base station. The variable $x(P)$ equals the amount of flow that is sent along path P. The MCF for our analysis is formulated as a linear program (LP). This formulation is modeled in similar terms to [80].

$$\text{maximize} \quad \alpha \tag{4.1}$$

$$\text{subject to}$$

$$\sum_{P:(i,j)\in P} x(P) \leq u_{ij}, \forall (i,j) \tag{4.2}$$

$$\sum_{P \in P_i} x(P) \geq \alpha d_i, \forall i \tag{4.3}$$

$$x(P) \geq 0, \forall P \tag{4.4}$$

The constraint in Eq. 4.2 specifies that the total flow crossing each link (i,j) must be less than or equal to the capacity of the link. Eq. 4.3 guarantees that the same percentage of demand is served for all the commodities. In our model, the capacity of the link is

given by Shannon's capacity. The link capacity is dependent on the subcarrier that a link is assigned to. Eq. 4.5 gives the capacity of a link (i, j) assigned to subcarrier c as

$$u_{ij} = Blog_2(1 + SINR_{ij}^c) \tag{4.5}$$

where B is the bandwidth of each subcarrier. The $SINR_{ij}^c$ is defined as

$$SINR_{ij}^c = \frac{G_{ij}^c P_i^c}{\eta + [\sum_{k \neq i,j} X_k G_{kj} P_k^c \chi_k^c]} \geq \beta \tag{4.6}$$

G_{ij}^c is the channel gain of link (i, j) on subcarrier c, P_i^c is the transmit power of node i on subcarrier c, G_{kj}^c is the channel gain from a simultaneously transmitting node k to the receiver node j on subcarrier c, and η is the noise power. In addition, the binary variable X_k denotes whether node k is transmitting or not on subcarrier c. Eq. 4.6 allows us to check whether those links transmitting on the same contribute to the SINR. Note that those that transmit on the same subcarrier inherently do not interfere due to the orthogonality of the subcarriers in OFDMA systems. Also, we look at only those nodes that are simultaneously transmitting rather than assuming that all nodes transmit at the same time as it is traditionally done.

In [78], the dual of the MCF is given using distance functions.The dual of the MCF linear program is modeled as follows:

$$\text{minimize} \quad \sum_{(i,j)} u_{ij} l_{ij} \tag{4.7}$$

subject to

$$\sum_{(i,j) \in P} l_{ij} \geq z_i, \forall i, \forall P \in P_i \tag{4.8}$$

$$\sum_{1 \leq i \leq \mathcal{N}} d_i z_i \geq 1 \tag{4.9}$$

$$l_{ij} \geq 0, \forall (i,j) \tag{4.10}$$

$$\forall i : z_i \geq 0 \tag{4.11}$$

The dual LP formulation is solved using an ϵ-approximation scheme given in [80]. The algorithm of [80] proceeds in phases. In each phase, there are $|\mathcal{N}|$ iterations. In iteration i, the objective is to route d_i units of flow from user i to the BS.

Although the ϵ-approximation algorithm provides a framework to solve a MCF linear program, it can not be applied directly as is described in [80] because it has not been tailored to consider interference constraints at each iteration. Rather, it is a straightforward approach to solve MCF in an ideal network. In order to ensure interference is recognized each time a path is found, we implement the ϵ-approximation algorithm in the following manner: At each iteration, i, links are chosen to form a path to the BS. When selecting these links, we must ensure that they satisfy the SINR requirement which affects the capacity, u_{ij}, given in Eq. 4.5. It is possible that the choice of links in an iteration will interfere with links already chosen in the previous iterations (iterations 1 to $i-1$). Therefore, while calculating the SINR value of the new links of iteration i (using Eq. 4.6), we must guarantee that the SINR values of the previous links are not violated. If a link in a new path during iteration i interferes with that of a link in a previously established path (i.e., SINR drops below β), then that link is not chosen and a new link is found to add to the path of iteration i. In this manner, we are able to incorporate within the MCF computation the signal quality of the links to render paths with least interference thereby maximizing the achievable network throughput.

To solve this optimization problem, we use an ϵ-approximation algorithm derived from [80] with a complexity of $\bigcirc(\epsilon^{-2}m(n+m))$ where m is the number of links in the network and n is number of nodes (users plus relay nodes) in the network.

4.4 STDMA Multicarrier Traffic Sensitive Scheduling

The MCF-ROPT routing approach (given in Section 4.3 obtains the total flow on each link for all connections, such that maximum total traffic is routed. In the scheduling problem, SM-TSS, we want to schedule transmissions such that interference is mitigated and the total maximum achievable throughput obtained using MCF-ROPT is maintained. For a link assignment using SM-TSS, a feasible interference free schedule will be available if the following constraints are satisfied: 1) adjacent links are assigned different time slots (because a node can not do more than one thing at a time (i.e., receive from two different sources)) regardless of whether they transmit on the same or different subcarriers; 2) a time slot can be assigned to a link only if the SINR for the link is satisfied; 3) two links transmitting on the same subcarrier can be scheduled in the same time slot if and only if the SINR values for the receivers are satisfied. The number of time slots in a frame changes depending on the number of traffic flows present.

We assume that each unit of flow requires a time slot for transmission. In other words, a link (i, j) requires F_{ij} scheduled time slots, where F_{ij} denotes the total units of flow on a link (i, j). We also assume that an initial frame length, T, is given. To ensure feasibility of the scheduling, we assume that one time slot is needed to schedule each unit of flow, given no reuse is incorporated. Therefore, we select T to be the sum of all the link flows.

We define two variables in our scheduling optimization formulation:

$$u_t = \begin{cases} 1 & \text{if time slot } t \text{ is used} \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ijt}^{c} = \begin{cases} 1 & \text{if time slot } t \text{ is assigned to link } (i,j) \\ & \text{transmitting on subcarrier } c \\ 0 & \text{otherwise} \end{cases}$$

We run the SM-TSS on network $\mathcal{G}$. Eqs. 4.12-4.20 show the optimization formulation for our scheduling problem.

$$\text{minimize} \quad \sum_{t \in T} u_t \tag{4.12}$$

subject to

$$\sum_{t \in T} y_{ijt}^{c} \geq F_{ij}, \quad \forall (i,j) \in \mathcal{E}, \ \forall c \in C \tag{4.13}$$

$$y_{ijt}^{c} \leq u_t, \quad \forall (i,j) \in \mathcal{E}, \ \forall t \in T, \ \forall c \in C \tag{4.14}$$

$$\sum_{(i,j) \in \mathcal{E}} y_{ijt}^{c} + \sum_{(k,i) \in \mathcal{E}} y_{kit}^{c} \leq 1, \quad \forall t \in T, \ \forall c \in C \tag{4.15}$$

$$\frac{G_{ij} P_i + (1 - y_{ijt}^{c})}{\xi} \geq \beta, \quad \forall (i,j) \in \mathcal{E}, \ \forall t \in T, \ \forall c \in C \tag{4.16}$$

$$y_{ijt}^{c} \leq \frac{P_i G_{ij}}{N_j \beta}, \quad \forall (i,j) \in \mathcal{E}, \ \forall t \in T, \ \forall c \in C \tag{4.17}$$

$$u_t \leq u_{t-1}, \quad \forall t \in T \tag{4.18}$$

$$u_t \in \{0,1\}, \quad \forall t \in T \tag{4.19}$$

$$y_{ijt}^{c} \in \{0,1\}, \quad \forall (i,j) \in \mathcal{E}, \ \forall t \in T, \ \forall c \in C \tag{4.20}$$

The objective function minimizes the number of time slots to be used. The constraint in Eq. 4.13 ensures that every link is assigned time slots at least equal to the amount of flow on the link. Eq. 4.14 is the coupling constraint between the variables (the time slots assigned to a link (i,j) must be less than or equal to the total number of time slots in the frame). Eq. 4.15 states that two adjacent links must be assigned different time slots.

The constraint in Eq. 4.16 expresses the required SINR threshold that should be satisfied to have a successful transmission. The term $1 - y_{ijt}^{c}$ ensures that the SINR inequality is satisfied when link (i, j) does not transmit in time slot t. ξ denotes the denominator of Eq. 4.6. Eq. 4.17 is based on the assumption that all links in $\mathcal{E}$ satisfy the SINR constraint when there are no concurrent transmissions. Finally, Eq. 4.18 implies that slot t is assigned only if slot $t-1$ has been assigned. In order to solve the STDMA scheduling problem in a manner that is computationally tractable, we use the set covering approach to develop an integer optimal solution which can then be further decomposed using a column generation approach which provides a relaxation of the LP problem formed by the set covering.

The set covering formulation is based on the concept of a transmission group, which represents a group of links that can be in simultaneous transmission. Let ζ_Z denote the set of all transmission groups of links. We introduce the following integer variables:

1. x_z is the number of time slots assigned to transmission group z.

2. $s_{ij}^{z,c} = 1$ if group z contains link (i, j) transmitting on a subcarrier c or 0 otherwise.

Since we know the traffic load on each link, F_{ij}, from the MCF-ROPT analysis, we formulate the set covering formulation for the scheduling problem as follows

$$\text{minimize} \quad \sum_{z \in \zeta_Z} x_z \tag{4.21}$$

subject to

$$\sum_{z \in \zeta_Z} s_{ij}^{z,c} x_z \geq F_{ij}, \quad \forall (i, j) \in \mathcal{E}, \ \forall c \in C \tag{4.22}$$

$$x_z \in \{0, 1\}, \quad \forall z \in \zeta_Z \tag{4.23}$$

The objective function is to minimize the total number of assigned time slots. The constraint in Eq. 4.22 ensures that each link can be assigned enough time slots to support

the transmission traffic load and Eq. 4.23 indicates that x_z is an integer variable. The complexity of the set covering formulation lies mainly in the cardinality of the set ζ_z. For networks of realistic size, there are many numbers of transmission groups. Thus, we use the column generation technique to solve the set covering formulation.

The column generation approach is a successful decomposition technique to solve large-scale LP problems. In the decomposition scheme for the set covering formulation, the original LP is decomposed into two parts: a master problem and a subproblem. The master problem contains a subset of columns and the subproblem is solved to identify whether the master problem needs a new column or not. If the master problem has to be enlarged, one or several columns are added to the master problem, which is then re-optimized, and the procedure repeats until it is large enough to find an optimal solution of the original LP.

To use the column generation approach, we first reformulate the formulation (given in Eqs. 4.21-4.23), where x_z is replaced by y_z to represent the proportion of time slots that is assigned to transmission group z ($y_z = \frac{x_z}{T}$). The LP relaxation of the set covering formulation is given as

$$\text{minimize} \quad \sum_{z \in \zeta_Z} y_z \tag{4.24}$$

subject to

$$\sum_{z \in \zeta_Z} s_{ij}^{z,c} y_z \geq F_{ij}, \quad \forall (i,j) \in \mathcal{E}, \ \forall c \in C \tag{4.25}$$

$$0 \leq y_z \leq 1, \quad \forall z \in \zeta_Z \tag{4.26}$$

Using the formulation of Eqs. 4.24-4.26, we decompose the SM-TSS problem using column generation. We first look at the master problem of the column generation approach. We consider a subset of ζ_z, denoted ζ'_z. To ensure feasibility of the master problem, we let ζ'_z be the set of links to be scheduled in a frame derived by a pure TDMA scheme, meaning each transmission group in ζ'_z contains only 1 link. This yields the

master problem shown below

$$\text{minimize} \quad \sum_{z \in \zeta'_Z} y_z \tag{4.27}$$

subject to

$$\sum_{z \in \zeta'_Z} s_{ij}^{z,c} y_z \geq F_{ij}, \quad \forall (i,j) \in \mathcal{E}, \ \forall c \in C \tag{4.28}$$

$$0 \leq y_z \leq 1, \quad \forall z \in \zeta'_Z \tag{4.29}$$

When the master problem is solved, we determine whether ζ'_z is sufficiently large to find an optimal solution or not. This is equivalent to examining whether there exists any element $z \in \zeta_z$ for which the corresponding variable y_z has a positive reduced cost. Using the LP-dual theory described in [93], the reduced cost of variable y_z is:

$$RC_z = 1 - \sum_{(i,j) \in \mathcal{E}} \gamma_{ij} s_{ij}^{z,c} \tag{4.30}$$

where γ_{ij}, $\forall (i,j) \in \mathcal{E}$ are the optimal dual variables to the constraint in Eq. 4.28. The subproblem should be solved if and only if the minimum of Eq. 4.30 is negative.

We formulate the subproblem using the following variable: $s_{ij}^c = 1$ if link (i,j) is included in the transmission group and transmits on a subcarrier c or 0 otherwise. The subproblem for the column generation is formulated as follows:

$$\text{maximize} \quad \sum_{(i,j) \in \mathcal{E}} \gamma_{ij} s_{ij}^c \tag{4.31}$$

subject to

$$\sum_{(i,j) \in \mathcal{E}} s_{ij}^c + \sum_{(k,i) \in \mathcal{E}} s_{ki}^c \leq 1, \quad \forall c \in C \tag{4.32}$$

$$\frac{G_{ij} P_i + (1 - s_{ij}^c)}{\xi} \geq \beta, \quad \forall (i,j) \in \mathcal{E}, \ \forall c \in C \tag{4.33}$$

$$s_{ij}^{c} \leq \frac{P_i G_{ij}}{\eta\beta}, \quad \forall (i,j) \in \mathcal{E}, \ \forall c \in C \tag{4.34}$$

If the solution to the subproblem results in a non-positive reduced cost, the optimal LP-value to the master problem is found. Otherwise, the master problem is re-optimized with a new column added to ζ'_Z and the procedure continues until we get the upper bound of the integer optimum of the scheduling problem. The column generation analysis provides us with a schedule that has a minimum number of time slots with the link flows scheduled in a non-interfering manner. Therefore, we obtain a feasible set of transmission links for each time slot.

4.5 Performance Evaluation

4.5.1 Simulation Model and Performance Metrics

We consider a relay enhanced multihop cellular network, $\mathcal{G}$, in a $900m \times 900m$ region. Each user generates traffic and the flows are routed towards the base station. There is no downlink traffic generated. We use NS-2 to simulate the networks. The base station is located in the center of the network. Locations for the set of relay nodes that form the mesh network are randomly generated. Locations for the user nodes are also randomly generated. We assume that the BS and relays have an infinite buffer, thus eliminating complications due to buffer overflow. The AWGN factor for link (i,j) using subcarrier c is 10^{-11}W and the pathloss exponent, κ, is 3. We also assume that the channel gains are known in advance or estimated accurately (i.e., via pilot symbols).

Analysis of the SM-TSS problem is done by using three test networks consisting of 20, 40 and 60 nodes. Each network has 256 and 512 subcarriers. The master and sub-problems of the SM-TSS column generation are solved using CPLEX branch and bound enumeration methodology. We run the column generation on the network $\mathcal{G}$. The column generation allows us to find the number of time slots needed to achieve the maximum

throughput (obtained from MCF-ROPT) such that the length T is the smallest possible to accommodate the link traffic (flow on each link). We assume a first in first out (FIFO) scheduling priority scheme.

Note that although the MCF-ROPT and SM-TSS are optimized separately, the flow results obtained by MCF-ROPT are used in the SM-TSS to determine the number of time slots required using spatial reuse. Thus, the SM-TSS results achieved are based on the effective throughput obtained through the MCF-ROPT algorithm.

4.5.2 Simulation Results and Discussion

We first show that the number of time slots obtained from the SM-TSS to schedule flows determined from the MCF-ROPT is less than the initial T time slots. Recall that T is equal to the number of units of flow assigned on all links assuming no spatial reuse of time slots and that each unit of flow needs one time slot for transmission. The results are given in Table 4.1. The results show that the SM-TSS effectively uses spatial reuse and interference mitigation to reduce the number of time slots to achieve the flow assignments.

Table 4.1: Time slots obtained using SM-TSS versus initial T for MCF-ROPT flow assignments

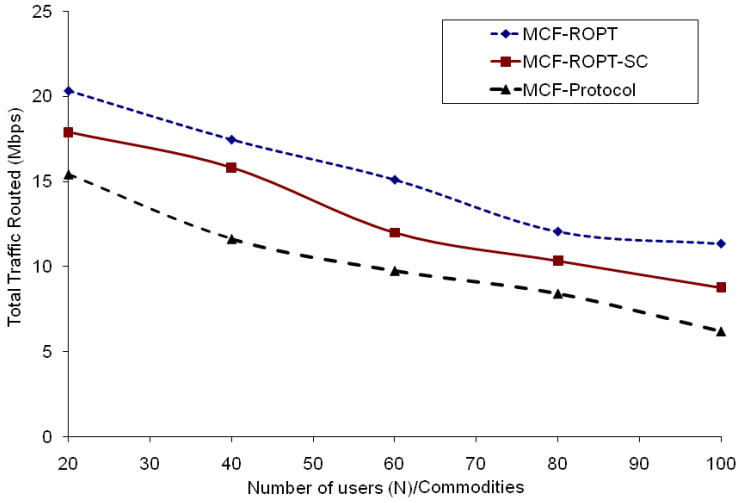
Nodes	SM-TSS Time Slots	T Time Slots
20 (5 RS, 15 MS)	127	163
40 (15 RS, 25 MS)	308	343
60 (20 RS, 40 MS)	812	988

We next evaluate the throughput obtained using the MCF-ROPT. The MCF-ROPT

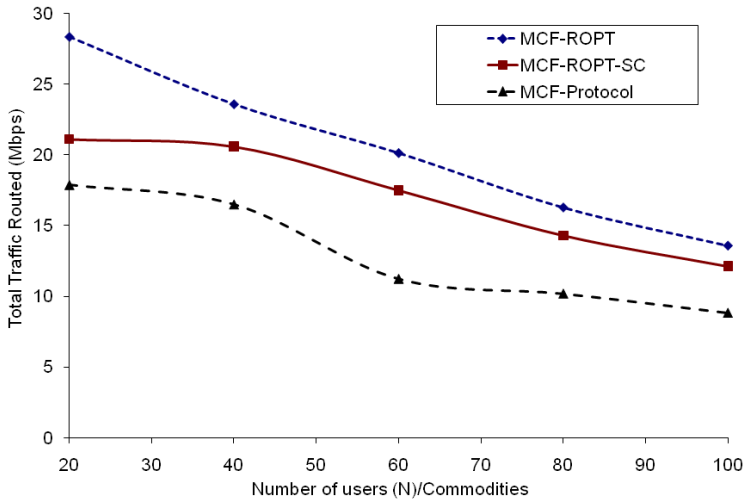
algorithm uses the physical interference model to quantify interference under multiple carriers. We compare our approach to two different protocols: 1) we implement a maximum concurrent flow optimization given in [82] (denoted as MCF-Protocol) where the interference constraints on the links are captured using the protocol interference model. This is then used to obtain the link capacities and thereby overall throughput. We implement the approach given in [82] with multiple subcarriers; and 2) we implement the MCF-ROPT under single channel conditions. This is known as MCF-ROPT-SC in the simulation graphs. We run simulations on networks generated by NS-2 with varying relay and user nodes (commodities) with 256 and 512 subcarriers. Traffic is Poisson generated, with each packet of size 500 bytes. The throughput results obtained are shown in Figs. 4.1(a) and 4.1(b), for 256 and 512 subcarriers respectively. Our MCF-ROPT approach provides the best throughput in comparison to the other two approaches due to better interference mitigation and efficient subcarrier reuse.

4.6 Chapter Summary

To alleviate the need and complexity of having joint interference constraints between routing and scheduling to determine schedulable flows, in this chapter, we developed a decoupled optimization approach for joint routing and scheduling under SINR interference constraints. We solved a maximum concurrent flow routing optimization problem by modifying the ϵ-approximation algorithm used for non-interference based multicommodity flow problems (i.e., in the wired environment). Furthermore, we used Spatial TDMA (STDMA) to allocate time slots to simultaneous transmissions. To solve the scheduling problem, we used the column generation technique to obtain the number of time slots required. We have shown that the proposed solution improves throughput and reduces the number of time slots required to route and schedule concurrent traffic.



(a) 256 subcarriers



(b) 512 subcarriers

Figure 4.1: MCF-ROPT compared to MCF using protocol interference model (MCF-Protocol) and MCF-ROPT under single channel conditions (MCF-ROPT-SC): (a) Networks with 256 subcarriers; (b) Networks with 512 subcarriers

Chapter 5

Multipath Routing and Max-Min Fair QoS Provisioning Under Interference Constraints

A fundamental issue of resource allocation is the fair distribution of bandwidth to competing users while simultaneously ensuring that no user is starved (i.e., denied necessary resources) and increasing throughput. This concept of fairness in resource allocation is a fundamental QoS policy. For this reason the max-min fairness (MMF) model has been extensively used in the literature to model the fair allocation of network resources. The classic MMF problem was originally defined for wired networks in order to allocate bandwidth to a set of given routes [94]. Research on MMF routing in the wired environment can be split into two categories: nonsplittable and splittable (multipath). In the nonsplittable case [94, 95], a MMF distribution of resources (bandwidth) to connections is done for fixed single path routing. In the splittable (multipath) MMF routing case, the traffic demands are allowed to be split among multiple flows (paths) [96, 97, 98, 99]. Multipath routing (MPR) has long been recognized as an effective strategy to achieve load balancing and increase reliability. It has been shown in [98] that multipath (split-

table) demand routing is a linear relaxation of the nonsplittable case, thus rendering the problem computationally tractable. To improve the transmission reliability and increase the probability of network survivability, the multiple paths can be selected to be link- or node-disjoint. In this case, the MPR approach is referred to as disjoint multipath routing (DMPR).

An important feature of multipath routing is the ability to provide QoS in terms of fair bandwidth allocation. Fairness based routing protocols that use the max-min model have been recently proposed in the literature [100, 101, 102, 103, 104]. All these works focus on the lexicographic (node ordering) optimization of routing for fair bandwidth allocation. These solutions can lead to high throughput solutions with guaranteed max-min bandwidth allocation value. However, they are formulated for ideal scenarios. Specifically, the inherent influence of interference has not been taken into account.

In this work, we develop a multipath routing scheme for fair resource allocation using a novel isotonic routing metric that is cognizant of interference. We also develop an interference aware lexicographically fair resource allocation optimization formulation using max-min fairness to allocate bandwidth to the routing paths. The work presented in this chapter has been published in [70], [71] and [72].

5.1 Interference Based Routing Metrics

Providing fault tolerance and QoS provisioning in the presence of interference are major issues that must be studied jointly in wireless networks in order to get a realistic sense of network performance. Developing routing metrics has long been the central focus of network layer protocol design. To compute paths using an interference aware routing metric is essentially equivalent to computing minimum weight (shortest) paths where the link weight is generated by the routing metric. In order to efficiently compute minimum weight paths using algorithms such as Dijkstra's shortest path or Bellman-Ford, the

routing metric must be *isotonic*. The isotonic property essentially means that a routing metric should ensure that the order of the weights of two paths are preserved if they are appended by a common third path. In addition, isotonicity ensures loop free routing. If a routing metric is not isotonic, only algorithms with exponential complexity can calculate minimum weight paths, which is not tractable for networks of even moderate size [105].

The two most prominent metrics are Expected Transmission Count (ETX) [106] and Expected Transmission Time (ETT) [107]. ETX is defined as the expected number of MAC layer transmissions needed to successfully deliver a packet through a wireless link. ETT improves upon ETX by considering the differences in transmission rates. Although both metrics are isotonic, neither considers interference. The earliest metric to consider interference is Weighted Cumulative ETT (WCETT) [107]. This metric essentially captures intra-flow interference by reducing the number of nodes on a path of a flow that transmit on the same channel; it gives low weight to paths that have more diversified channel assignments. However, WCETT does not capture inter-flow interference and is not isotonic which prevents the use of an efficient loop free routing algorithm to compute minimum weight paths. The Metric for Interference and Channel switching (MIC) [105] improves WCETT by capturing inter-flow interference and overcomes the non-isotonicity problem. However, MIC does not measure interference dynamically, meaning that changes to interference level over time due to signal strength and traffic load may not be captured accurately. The Interference AWARE (iAWARE) routing metric [46] computes paths with lower inter-flow and intra-flow interference than MIC and WCETT. It uses SNR and SINR to continuously monitor neighboring interference variations. Yet, iAWARE is not isotonic. Recently, improvements to the ETX and ETT metrics such as Interferer Neighbor Count (INX) were proposed in [47]. Similar to MIC, INX takes into account interference through the number of links that can interfere on a link l. This metric performs better only in low traffic load conditions, and therefore load balancing is not completely resolved.

According to the main requirements of interference, load awareness and isotonicity, existing routing metrics address only some specific requirements. For this reason, in this work, a new routing metric is proposed in order to simultaneously address all of these aspects.

5.2 Problem Preliminaries

In this work we develop an isotonic routing metric, RI^3M which is then used to find link disjoint paths from each user to the base station. In addition, we optimize bandwidth allocation under interference constraints using the MMF model in which the lexicographically largest bandwidth allocation vector is determined.

Our network architecture, discussed in Chapter 2, consists of a mesh network of relays and users connected to at least two relay nodes which inherently provides for multiple paths to exist to the base station. This topology setup ensures that the network is at least 2-link connected (i.e., each node has at least two link disjoint connections to other nodes). In DMPR schemes, the disjointedness property ensures that when k multiple paths are constructed, no set of $k-1$ link failures can disconnect all the paths. Through Menger's Theorem [108] it has been shown that for two distinct nodes x and y, the minimum number of edges whose removal disconnects x and y is equal to the maximum number of pairwise link disjoint paths from x to y. Thus, in our case, 2-connectivity is a necessary and sufficient condition to find a solution for two disjoint paths for each user node to the base station. 2-connectivity in wireless networks has been studied in [44, 45][1].

The system model used in this problem is based on CSMA/CA relay enhanced cellular networks [74]. In order to accurately mimic the behavior of CSMA/CA networks, the interference model that is implemented in this work is an enhancement of the interfer-

[1]Notice that maintaining 2-connectivity is a necessary condition for finding two disjoint paths from each user to the base station. Guaranteeing 2-connectivity is feasible in a static wireless environment as considered in this work. However, in the presence of mobility, 2-connectivity of the network can not be ensured due to time varying changes in the topology. Thus, this constraint and the solutions obtained herewith are limited to static wireless networks.

ence model described in Chapter 2, Section 2.2. Specifically, in addition to the "reality check" method which defines in practical terms the interference and transmission ranges of each node based on the SINR threshold, we add the following requirements to the interference model. Since link layer availability is required for CSMA/CA, an ACK packet is generated by each receiver for every data packet it receives. Due to carrier sensing and RTS/CTS/ACK exchanges, a transmission along link $e = (u, v)$ (in either direction) blocks all simultaneous transmissions within the interference ranges of u and v. In the physical interference model, successful reception of a packet sent by node u to node v depends on the SINR at v. To be coherent with the link-layer availability, we extend the physical interference model as follows. We assume that a packet sent by node u is correctly received by node v if and only if the packet is successfully received by v, and the ACK sent by node v is correctly received by node u. Furthermore, for a transmission from node x to node y that is concurrent with the packet on (u, v), we account for the interference both from node $x's$ data packet and from node $y's$ ACK. Although only one of x and y transmits at a time, their data and ACK packets could both overlap with either the data packet or the ACK along (u, v). Thus, we choose the maximum of the interferences from x and y when calculating the total interferences at u and v. Note that which of the two (x or y) contributes the maximum interference could be different at u and v. Thus, a packet sent along link (u, v) (in either direction) is correctly received if and only if:

$$SINR_{uv} = \frac{P_v(u)}{\eta + \sum_{(x,y)\in\mathcal{E}'} max(P_v(x), P_v(y))} \geq \beta \tag{5.1}$$

and

$$SINR_{vu} = \frac{P_u(v)}{\eta + \sum_{(x,y)\in\mathcal{E}'} max(P_u(x), P_u(y))} \geq \beta \tag{5.2}$$

where $\mathcal{E}'$ contains all links that have simultaneous transmissions concurrent with the one on (u, v) and $P_v(u)$ is the received power at node v from the transmitted signal by node

u.

Notice that optimization techniques to find an efficient algorithm that determines the collision domain and backoff times in CSMA/CA networks for each node based on the interference range has been studied [109]. The authors propose closed form expressions for the mean backoff time in terms of path flow variables, making it possible to optimize the network based on multipath routing. However, their approach is analytically complex. In addition, since the focus of this work is to incorporate the physical layer interference into the protocol model, determining the optimal collision domain and wait periods are not relevant.

5.3 Isotonicity

As mentioned earlier, isotonicity reflects the ability of a routing metric to compute minimum weight, loop free paths. Assume that for any path a, its weight, denoted as $W(a)$, is defined by a routing metric which is a function of a. Denoting the concatenation of two paths, a and b, by $a \oplus b$, isotonicity can be defined as follows:

Definition 5.3.1 ***Isotonicity:*** *A routing metric, $W(\cdot)$, is isotonic if $W(a) \leq W(b)$ implies that both $W(a \oplus c) \leq W(b \oplus c)$ and $W(c' \oplus a) \leq W(c' \oplus b)$, for all a, b, c, c'.*

Fig. 5.1 illustrates the isotonicity property. In [105] it was shown that isotonicity is a necessary and sufficient condition for both the Bellman-Ford and Dijkstra's algorithms to find minimum weight paths that are loop free. Therefore if a routing metric can be proven to be isotonic, any variation of a shortest path algorithm can be used to route packets in a wireless network.

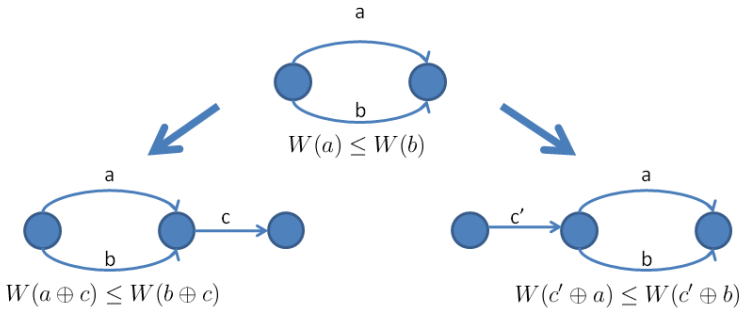


Figure 5.1: Example of the isotonic property

5.4 Routing with Inter-Flow and Intra-Flow Interference Metric (RI^3M)

5.4.1 Problem Formulation

The RI^3M interference routing metric takes into consideration the following three factors: inter-flow interference, intra-flow interference and traffic load. Inter-flow interference generally results in bandwidth starvation for some nodes since a flow contends for bandwidth along its own path and its neighboring area. To prevent such starvation, the routing metric must balance the traffic load along the path of the flow and reduce the inter-flow interference imposed in the neighboring area. RI^3M consists of two components. The first component, IL, deals with inter-flow interference and load awareness. The second component, channel switching cost, CSC, captures intra-flow interference. We now formalize our routing metric. Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be an undirected, 2-connected network, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of links. Let p be a path from a user node to the BS. We define RI^3M as follows:

$$\sum_{\forall(i,j)\in p} IL_{ij} + \sum_{\forall i \in p} CSC_i \tag{5.3}$$

where node i represents a node on path p and link (i, j) represents a link on the path p.

IL_{ij} Component

The IL_{ij} component is intended to capture information about the inter-flow interference and traffic load simultaneously. It consists of two separate subcomponents. To capture the inter-flow interference, we use the concept of the interference ratio (IR) [46], which is based on the physical interference model. The IR depicts the interference based on the ratio between SNR and SINR. The IR captures interference by monitoring the signal strength values. When there is no interference (i.e., no interfering neighbors or no traffic generated by interfering neighbors), the SINR of link (i, j) is independent of the inter-flow interference and the quality of the link is determined by the intra-flow interference component. Eq. 5.4 shows the IR ratio ($0 < IR_{ij} \leq 1$).

$$IR_{ij} = \frac{SINR_{ij} + SINR_{ji}}{SNR_{ij}} \tag{5.4}$$

where SNR_{ij} is given by $\frac{P_j(i)}{\eta}$ and the SINR in the numerator is the sum of the SINR values given in Eqs. 5.1 and 5.2.

To estimate the traffic load on a wireless relay node, a typical approach is to measure the traffic volume going through the corresponding node in terms of byte rate or packet rate. Unfortunately, this approach is unable to give an accurate estimate of the usage of the radio channel at which the node operates because the total capacity of the network is not fixed and depends on many factors, such as the physical transmission rate of each relay node, frame size, number of retransmissions, interference, etc. Simply counting the bytes or even packets going through a relay node fails to take into account these factors. In light of these limitations, [110] adopts an alternative approach to estimate the traffic load, which is based on the percentage of channel time of the relay node that is consumed for frame transmission.

To measure the traffic load, we use the concept of Channel Busy Time (CBT). A radio channel's time consists of a series of interleaved busy periods and idle periods. A busy period is a time period in which one node attempts to transmit frames while other nodes hold off their transmission. An idle period is a time period in which every node considers the radio medium available for access. Using the CBT, it is possible to estimate the traffic load (channel utilization) on each link. The CBT calculation is the percentage of time that a channel is busy (transmitting). In order to compute this time, we first define the different states that a node can be assigned:

- Success: This state refers to the case where a node has successfully received the acknowledgment of the packet it has sent.
- Backoff: Even though a node has some data to transmit and the medium is free, there is a random waiting period (during which the wireless medium has to remain idle) before it starts sending its data.
- Wait: If there are ongoing transmissions within the interference range of the node which causes the SINR threshold to drop below β, it has to wait until the ongoing communications are completed before starting its own data.
- Collision: In this state, a node which has sent a packet never receives an acknowledgement for this packet.

Let $T_{success}$, $T_{backoff}$, T_{wait} and $T_{collision}$ be the time spent respectively in the states Success, Backoff, Wait and Collision. The idle time (i.e., time where there is no data to keep the channel busy), T_{idle}, considers backoff times, collision times and the waiting times. Thus the percentage of time the channel spends idle is defined as

$$T_{idle} = \frac{T_{backoff} + T_{collision} + T_{wait}}{T_{backoff} + T_{collision} + T_{wait} + T_{success}} \tag{5.5}$$

Let us denote the denominator of Eq. 5.5 as the total time, T_{total}. Then the CBT for a

link (i, j) is defined in Eq. 5.6.

$$CBT_{ij} = \frac{T_{total} - T_{idle}}{T_{total}} \tag{5.6}$$

The CBT is used as a smoothing function, weighted over IR_{ij}. Using the IR_{ij} and CBT_{ij} subcomponents, IL_{ij} is defined as follows

$$IL_{ij} = (1 - IR_{ij}) * CBT_{ij} \tag{5.7}$$

where $0 \leq IR_{ij} \leq 1$ and $0 \leq CBT_{ij} \leq 1$.

CSC Component

To reduce the intra-flow interference, the RI^3M routing metric uses the CSC component. CSC, originally defined in [111], designates paths with consecutive links using the same channel with higher weight than paths that alternate their channel assignments. This allows paths with more diversified channel assignments to be favored in the routing process. Intra-flow interference can occur between successive nodes on a path; however depending on the interference range, it can also occur between nodes further away along the path. In this case, it is necessary to consider the channel assignments at more hops in order to choose an effective path that reduces intra-flow interference. To eliminate the intra-flow interference between node i and its previous hop, $prev(i)$, node i must transmit to the next hop, $next(i)$ using a different channel from the one it uses to receive from $prev(i)$. CSC denotes $CH(i)$ as the channel that node i transmits on to $next(i)$. The CSC of node i for intra-flow interference reduction of successive nodes is given as

$$CSC_i = \begin{cases} w_1 & \text{if} \quad CH(prev(i)) \neq CH(i) \\ w_2 & \text{if} \quad CH(prev(i)) = CH(i) \end{cases} \tag{5.8}$$

where $w_2 > w_1 \geq 0$ to ensure that a higher cost is imposed for those nodes that transmit on the same channel consecutively. In order to capture intra-flow interference between two nodes that are two hops away, node i interferes with both nodes $prev(i)$ and $prev^2(i)$ where $prev^2(i)$ is the node that is the two hop precedent of i. According to [111], the multihop extension of the CSC equation of Eq. 5.8 is

$$CSC_i = \begin{cases} w_2 & \text{if } CH(prev^2(i)) \neq CH(i) = CH(prev(i)) \\ w_3 & \text{if } CH(prev^2(i)) = CH(i) \neq CH(prev(i)) \\ w_2 + w_3 & \text{if } CH(prev^2(i)) = CH(i) = CH(prev(i)) \\ w_1 & \text{, otherwise} \end{cases} \tag{5.9}$$

where w_3 captures the intra-flow interference between nodes $prev^2(i)$ and i and w_2 captures the intra-flow interference between nodes $prev(i)$ and i. The weight w_3 must be strictly less than the weight w_2 because since the further away that two nodes are, the less interference exists between them. We consider intra-flow interference up to the limit of a node's interference range which is typically within a 3 hop range.

5.4.2 Virtual Network Decomposition to Illustrate Isotonicity

The RI^3M routing metric is not isotonic if used directly. We can see this in the example network given in Fig. 5.2. In the example, a link is represented by three parameters: starting node of the link, ending node of the link and the channel the link transmits on. If we assume that link $(A, B, 1)$ has a smaller RI^3M value than link $(A, B, 2)$, the weights of paths $(A, B, 1)$ and $(A, B, 2)$ satisfy: $RI^3M(A, B, 1) < RI^3M(A, B, 2)$. However, adding path $(B, C, 1)$ to path $(A, B, 1)$ introduces a higher cost than adding $(B, C, 1)$ to $(A, B, 2)$ because of the reuse of channel 1 on path $(A, B, 1) \oplus (B, C, 1)$. Thus, $RI^3M((A, B, 1) \oplus (B, C, 1)) > RI^3M((A, B, 2) \oplus (B, C, 1))$, which does not satisfy the definition of isotonicity as given in Section 5.3.

To make RI^3M an isotonic routing metric, we use a decomposition technique that

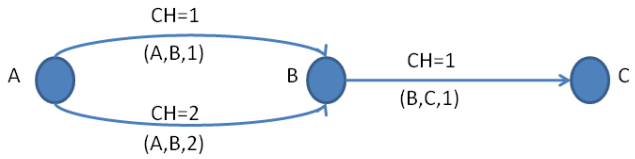


Figure 5.2: Example to prove non-isotonicity of RI^3M routing metric

creates a virtual network from the real network and decomposes RI^3M into isotonic link weight assignments on the virtual network. First introduced in [111] to prove the isotonicity of the MIC routing metric, the decomposition of RI^3M is based on the fact that the non-isotonic behavior of RI^3M is caused by the different increments of path weights due to the addition of a link on a path. Whether a cost increment will be different by adding a link is only related to the channel assignment of the previous link on the path. Since the possible assignments of channels for the predecessor link are limited, we introduce several virtual nodes to represent these possible channel assignments. Namely, for every channel c that a node $X's$ radios are configured to, two virtual nodes, $X_i(c)$ and $X_e(c)$ are introduced. $X_i(c)$ represents that node $prev(X)$ transmits to X on channel c. $X_e(c)$ indicates that node X transmits to its next hop, $next(X)$, on channel c. The subscript i stands for ingress and the subscript e stands for egress. In addition, two additional virtual nodes are introduced, $X-$ and $X+$, which represent the start and end nodes of a flow (i.e., $X-$ is used as the virtual destination node for flows destined to node X and $X+$ is used as the virtual source node for flows starting at node X). Hence, $X+$ has a link weight with 0 pointing to each egress node and $X-$ has a link weight 0 with each ingress virtual node of X.

Links from the ingress virtual nodes to the egress virtual nodes at node X are

added and the weights of these links are assigned to capture different CSC costs. Link $(X_i(c), Xe(c))$ represents that node X does not change channels while forwarding packets and hence weight w_2 is assigned to this link. Similarly, weight w_1 is assigned to link $(X_i(c), X_e(c1))$, where $c \neq c1$, to represent the low cost of changing channels while forwarding packets. Links between the virtual nodes belonging to different real nodes are used to capture the IL weight. Fig. 5.3 shows the virtual decomposition of Fig. 5.2.

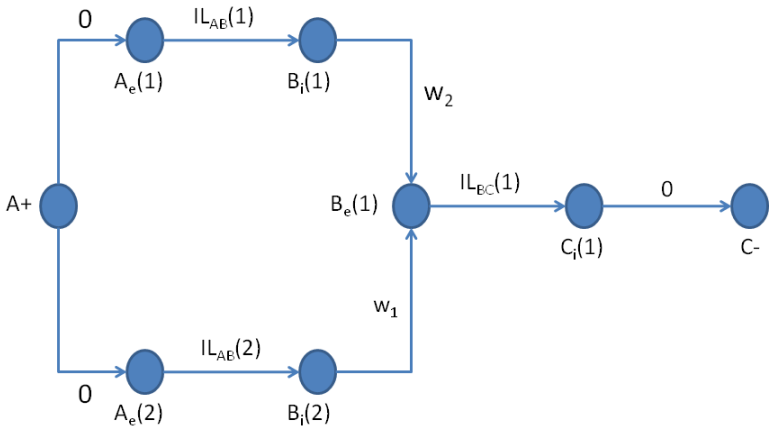


Figure 5.3: Decomposition of the network in Fig. 5.2 into a virtual network to prove RI^3M isotonicity

By building the virtual network from a real network, RI^3M is essentially decomposed in the real network into weight assignments to the links between virtual nodes. This is because the RI^3M weight of a real path in a real network can be reconstructed by aggregating all of the weights of the virtual links on the corresponding virtual path. The IL part of RI^3M is reflected in the weight of the links between virtual nodes in different real nodes. The CSC costs are captured by routing through different virtual links inside real nodes. Table 5.1 illustrates the real network mapping into the virtual network.

Now that RI^3M has been shown to be isotonic using a virtual network decomposi-

Table 5.1: Real network mapping to the virtual network

Real Path	Virtual Path	RI³M
$(A, B, 1) \oplus (B, C, 1)$	$A_e(1) \rightarrow B_i(1) \rightarrow B_e(1) \rightarrow C_i(1)$	$IL_{AB}(1) + IL_{BC}(1) + w_2$
$(A, B, 2) \oplus (B, C, 1)$	$A_e(2) \rightarrow B_i(2) \rightarrow B_e(1) \rightarrow C_i(1)$	$IL_{AB}(2) + IL_{BC}(1) + w_1$

tion, it can be used with any shortest path algorithm to find least interfering (minimum weight) paths. The problem of finding two link disjoint paths (primary and backup) of minimum total weight across a network has been dealt with efficiently by Suurballe's algorithm [112]. The algorithm developed by Suurballe has become the reference algorithm for finding link disjoint paths in wireless networks. Suurballe's algorithm always finds two link disjoint paths from a source node to the destination, as long as the paths exist in the network, assuring the total weight of both paths is the minimum among all pairs of paths in the network. We run Suurballe's algorithm on the virtual network, $\mathcal{G}_v(\mathcal{V}_v, \mathcal{E}_v)$, where $\mathcal{V}_v$ and $\mathcal{E}_v$ are the nodes and links of the virtual network, respectively. The link weights are determined by the values of the RI^3M routing metric. The steps of the algorithm are given in Fig. 5.4.

Suurballe's Link Disjoint Routing Algorithm Using RI^3M Link Weights

Inputs: Virtual connectivity graph, $\mathcal{G}_v(\mathcal{V}_v, \mathcal{E}_v)$ with RI^3M link weights

Output: Two link disjoint paths for each source node to the BS.

Step1: Run Dijkstra's shortest path algorithm to find the shortest primary path P between a source node and the BS.

Step2: Reverse the direction of P's links and invert their weights.

Step3: Run Dijkstra's shortest path algorithm again to find the shortest backup path Q from source to BS in the modified graph with undirected edges and directed arcs.

Step4: Transform the modified graph back to the original one and erase all overlapping edges on paths P and Q and all the edges which do not belong to any of these two paths. The remaining edges form the required shortest pair of paths.

Figure 5.4: Overall steps of Suurballe's link disjoint routing algorithm using RI^3M

5.5 Lexicographic MMF Multipath Flow (LMX:M^3F) Routing Algorithm with Interference Constraints

5.5.1 Problem Formulation and Definitions

In this section we model the MMF bandwidth allocation problem as a multi-commodity flow (MCF) problem. The MCF problem is a network flow problem where multiple commodities (demands) flow through the network (as used in Chapter 3 and 4). We consider the case that each demand has two candidate paths (where the paths are determined by using RI^3M). Thus the flows realizing each demand volume is split among the allowable paths. In the remainder of this paper we will denote vectors with bold letters and an arrow overhead. We will denote optimal vectors as regular vectors except with an additional star (*).

Definition 5.5.1 ***Multicommodity Flow:*** *Given a set $\mathcal{D}$ of demands, let $\delta_{edp}x_{dp} \geq 0$ be the flow allocated to path p of commodity (demand) d, $d \in \mathcal{D}$ on link $e \in \mathcal{E}$, where δ_{edp}*

is a binary variable that denotes whether link e belongs to path p or not. $\vec{\boldsymbol{X}}=(\vec{\boldsymbol{X}}_d : d \in \mathcal{D})$ *is a feasible multicommodity flow if* $\sum_{d\in\mathcal{D}} \sum_{p\in\mathcal{P}_d} \delta_{edp} x_{dp} \leq C_e$.

The capacity of link $e \in \mathcal{E}$ is denoted C_e and is mathematically expressed as

$$C_e = log_2(1 + SINR_e) \geq \beta \tag{5.10}$$

where $SINR_e$ is given in Eq. 2.1.

Given a network $\mathcal{G}$, our objective is to determine an MMF (Max-Min-Fair) bandwidth allocation vector under interference constraints where the allocation vector is lexicographically the largest possible.

Definition 5.5.2 *A n-vector* $\vec{\boldsymbol{x}} = (x_1, x_2, ..., x_n)$ *sorted in non-decreasing order* $(x_1 \leq x_2 \leq ... \leq x_n)$ *is* ***lexicographically greater*** *than another n-vector* $y = (y_1, y_2, ..., y_n)$ *sorted in non-decreasing order* $(y_1 \leq y_2 \leq ... \leq y_n)$ *if an index* k, $1 \leq k \leq n-1$ *exists, such that* $x_i = y_i$ *for* $i = 1, 2, ..., k$ *and* $x_{k+1} > y_{k+1}$.

We will discuss how our lexicographic bandwidth allocation problem is formulated using the interference aware routing metric developed in Section 5.4.

5.5.2 LMX:M^3F Algorithm

Given the network $\mathcal{G}$, paths for routing the traffic flow are found by using the routing metric given in Section 5.4 and running Suurballe's multipath routing algorithm. Given these paths, we provide the formulation of the lexicographically largest allocation vector using MMF considering interference constraints and the subsequent methodology used to solve it. The LMX:M^3F formulation is given in Eqs. 5.11-5.14 (referred to as Problem A in the remainder of the thesis) and follows a multicommodity flow approach.

LMX:M^3F: Problem A

Objective: Find total bandwidth allocation vector, $\vec{\mathbf{X}}$, such that it is lexicographically

maximal among all total bandwidth allocation vectors.

$$\text{lexicographically maximize } \vec{\mathbf{X}} \tag{5.11}$$

subject to

$$\sum_{p \in \mathcal{P}_d} x_{dp} = X_d, \forall d \in \mathcal{D} \tag{5.12}$$

$$\sum_{d \in \mathcal{D}} \sum_{p \in \mathcal{P}_d} \delta_{edp} x_{dp} \leq C_e, \forall e \in \mathcal{E} \tag{5.13}$$

$$x_{dp} \geq 0 \tag{5.14}$$

where $\mathcal{P}_d$ is the set of paths for demand d, x_{dp} is the flow (bandwidth) allocated to path p of demand d, and X_d is the total flow (bandwidth) allocated to demand d, $\vec{\mathbf{X}} = (X_1, X_2, ..., X_D)$.

In order to find the MMF allocation vector for the corresponding paths, we define the *demand satisfaction vector*, $\vec{\mathbf{t}}$. Let $\gamma_d \geq 0$ be the flow value of x_{dp}, and $\zeta^+(v)$ and $\zeta^-(v)$ be the outgoing and incoming links to node v, respectively. The law of flow conservation states that

$$\sum_{e \in \zeta^+(v)} x_{dp} - \sum_{e \in \zeta^-(v)} x_{dp} = \begin{cases} \gamma_d & \text{if } v = BS \\ -\gamma_d & \text{if } v = source \\ 0 & \text{, otherwise} \end{cases} \tag{5.15}$$

A feasible multicommodity flow, $\vec{\mathbf{X}}$, with $\gamma_d \geq h_d$, $d \in \mathcal{D}$, defines an admissible flow (bandwidth), where h_d is the amount of demand to be routed. Assume $\vec{\mathbf{X}}$ is feasible and also consider a vector $\vec{\mathbf{t}} = (t_d \geq 0 : d \in \mathcal{D})$ such that $\gamma_d = t_d h_d$ in Eq. 5.15. If $t_d \geq 1$ for all $d \in \mathcal{D}$, then the flow is admissible (i.e., it fulfills the demand requirement h_d, $d \in \mathcal{D}$). Thus $\vec{\mathbf{t}}$ is denoted as the demand satisfaction vector for routing vector $\vec{\mathbf{X}}$. Specifically, the physical meaning of the value t is the amount that is added to saturate/satisfy x_{dp}.

We solve for t using the optimization formulation given in Eqs. 6.5-5.20 (referred to as Problem B in the remainder of the paper).

Problem B

$$\text{maximize} \quad t \tag{5.16}$$

subject to

$$X_d = \sum_{p \in \mathcal{P}_d} x_{dp}, \forall d \in \mathcal{D} \tag{5.17}$$

$$t - X_d \leq 0, \forall d \in \mathcal{D} \tag{5.18}$$

$$\sum_{d \in \mathcal{D}} \sum_{p \in \mathcal{P}_d} \delta_{edp} x_{dp} \leq C_e, \forall e \in \mathcal{E} \tag{5.19}$$

$$x_{dp} \geq 0 \tag{5.20}$$

The objective function in Eq. 5.16 and the constraint in Eq. 5.18 are equivalent to the ultimate objective to be achieved, given in Eq. 5.21.

$$\max \quad \min X_d : d \in \mathcal{D} \tag{5.21}$$

Problem A can be solved by computing consecutively the value of the demand satisfaction vector of Problem B. Primarily, the idea is that first the lowest value among the components of $\vec{\mathbf{t}}$ has to be maximized before the second lowest value is maximized. In order to ensure that the demands are satisfied, we have to check which total demand allocations, X_d, can be further increased. A demand d whose satisfaction value t_d can not be further increased is called blocking [113]. To check the satisfaction of a demand, the linear program (LP) shown in Eqs. 5.22-5.26, referred to as Problem C, is solved for each demand, d

Problem C

$$\text{maximize} \quad X_d \tag{5.22}$$

subject to

$$X_{d'} = \sum_{p \in \mathcal{P}_{d'}} x_{d'p}, \forall d' \in \mathcal{D} \tag{5.23}$$

$$t_{d'} - X_{d'} \leq 0, \forall d' \in \mathcal{D} \tag{5.24}$$

$$\sum_{d' \in \mathcal{D}} \sum_{p \in \mathcal{P}_{d'}} \delta_{ed'p} x_{d'p} \leq C_e, \forall e \in \mathcal{E} \tag{5.25}$$

$$x_{d'p} \geq 0 \tag{5.26}$$

where $t_{d'}$ are constants. To put Problem C in perspective, let t^* be the optimal solution of the LP. A demand is non-blocking (can be further increased) if the optimal X_d value, X_d^*, is strictly greater than t^* (i.e., $X_d^* > t^*$).

The components of Problem B and Problem C are used in conjunction to solve the original LMX:M^3F (Problem A) problem. The algorithm for solving LMX:M^3F is given in Fig. 5.5.

5.6 Performance Evaluation

5.6.1 Simulation Model and Performance Metrics

In this section we evaluate the performance of the RI^3M routing metric and the LMX:M^3F algorithm via simulations. We consider a 2-connected cellular network, $\mathcal{G}$, in a $900m \times 900m$ region where all nodes are stationary. Each user generates traffic and the flows are routed to and from the base station. We use NS-2 to simulate the networks and use CPLEX to solve the optimization formulation for LMX:M^3F. The base station is located in the center of the network. Locations for the set of relay nodes that form the mesh network are randomly generated. Locations for the user nodes are also randomly generated. We assume that the BS and relays have an infinite buffer, thus eliminating complications due to buffer overflow. We also assume that channels have been assigned

LMX:M^3F Algorithm

Step1: Solve Problem B. Let $(t^*, \vec{\mathbf{x}}^*, \vec{\mathbf{X}}^*)$ be the optimal solution of Problem A. Initialize: $k := 0$ (number of iterations), $Z_0 := \emptyset$ (set of demands that are blocking/saturated) $Z_1 = \{1, 2, ..., \mathcal{D}\}$, and $t_d := t^*$ for each $d \in Z_1$.

Step2: $k := k + 1$. Consider each demand, $d \in Z_1$, one by one to check whether the total allocated bandwidth X_d^* can be increased more than t^* without decreasing the already found maximal allocations t'_d for all other demands, d'. To check the demands, solve Problem C. If there are no blocking demands in Z_1, go to **Step3**. Otherwise for blocking demand d, add d to set Z_0 and delete it from set Z_1, $Z_0 := Z_0 \cup \{d\}$, $Z_1 := Z_1 \setminus \{d\}$. If $Z_1 = \emptyset$, STOP. Then $\vec{\mathbf{X}}^* = (X_1^*, X_2^*, ..., X_D^*) = (t_1, t_2, ..., t_d)$ is the solution of Problem A.

Step3: To improve the current best bandwidth allocation, solve the following LP (**Problem D**).

maximize t
subject to
$X_d = \sum_{p \in \mathcal{P}_d} x_{dp}, \forall d \in Z_1$
$t - X_d \leq 0, \forall d \in Z_0$
$\sum_{d \in \mathcal{D}} \sum_{p \in \mathcal{P}_d} \delta_{edp} x_{dp} \leq C_e, \forall e \in \mathcal{E}$
$x_{dp} \geq 0$

Let $(t^*, \vec{\mathbf{x}}^*, \vec{\mathbf{X}}^*)$ be the optimal solution of Problem D. Put $t_d := t^*$ for each $d \in Z_1$. Go to **Step2**.

Figure 5.5: Summary of the steps for the LMX:M^3F algorithm

using a generic link coloring scheme [114]. The simulation parameters used are as follows: System Bandwidth (W) = 1MHz, AWGN Noise (η) = -90bBW/Hz; Maximum transmission power: Relay (35dBm), User (24dBm) (note that the power levels of the nodes are such that it is sufficient to allow nodes connect to at least two of its neighbors, ensuring 2-connectivity); PHY Specification: 802.11; Number of channels per radio: 12; Antenna: Omnidirectional. To evaluate the performance of RI^3M, we study the following performance metrics: 1) end-to-end delay (amount of time it takes to deliver packets from the user node to the BS); 2) flow throughput; and 3) packet loss ratio. We simulate 20 runs for each set of data and show the average results. To evaluate the performance of

LMX:M^3F, we adopt the following performance metrics: 1) Bandwidth Blocking Ratio (BBR): BBR represents the percentage of the amount of blocked traffic over the amount of bandwidth requirements of all traffic requests (connection requests) during the entire simulation period; 2) Total Bandwidth Usage: This measurement helps us examine whether our LMX:M^3F algorithm can save more network resources (use less) than other established MMF routing algorithms that incorporate interference; and 3) Link Load: Measurement that indicates the traffic load on each link due to different routing approaches. Note that the performance evaluation of LMX:M^3F is based upon the paths determined from using the RI^3M routing metric.

As benchmarks for evaluating the effectiveness of our proposed metric, our comparison is with 5 other routing metrics in the literature, specifically, ETX [106], ETT [107], MIC [105], iAWARE [46] and INX [47]. Each metric is used with Suurballe's disjoint multipath routing algorithm. We also compare our proposed approach with two disjoint multipath routing algorithms. First, the algorithm developed in [81] develops a routing metric where a node calculates the SINR to its neighboring links based on a 2-Hop interference estimation algorithm (2-HEAR). Second, the algorithm developed in [50] provides an interference minimized multipath routing (I2MR) algorithm that increases throughput by discovering zone disjoint paths using the concept of path correlation. As benchmarks for evaluating the effectiveness of our bandwidth allocation algorithm, we compare LMX:M^3F to two MMF bandwidth allocation algorithms that consider interference when allocating bandwidth. First, the algorithm developed in [15] is an interference based routing and bandwidth allocation algorithm, known as MICB. The protocol model is used to create an auxiliary graph such that the maximum interference level within the network does not exceed a maximum value. Second, the algorithm described in [42] quantifies interference through the creation of contention graphs where interfering flows are captured in multihop wireless networks. We refer to this algorithm as MMCFContGr. We modify the implementations of these algorithms so that multiple

paths are considered.

5.6.2 Simulation Results and Discussion

We first evaluate RI^3M in terms of end-to-end delay. We use the end-to-end user demand delivery delay as a metric to evaluate the impact of the interference quantification method of RI^3M in comparison to the existing routing metrics and the two established disjoint multipath routing algorithms. To measure the end-to-end delay, the transmitting rate of the user and relay nodes are set to 4.5Mbps [115]. All routing flows are CBR flows with 512 byte packets. To model the packet dropping error, for a given SINR value, we use the packet error ratio (PER) [84], which is readily available in NS-2.

Performance Evaluation of RI^3M

We first compare RI^3M with the existing routing metrics. We simulate networks with 99 nodes (6 relays, 93 user nodes). All networks have one base station. Fig. 5.6 shows the average end-to-end delay values of RI^3M versus the other routing metrics, measured against varying demands (traffic load). We see that the proposed RI^3M achieves lowest delay in comparison to the other metrics, particularly as demands increase. It can be said that RI^3M quantifies interference more accurately because it considers the influence of inter-flow and intra-flow interference which allows us to avoid paths with high interference, thereby reducing the time taken to deliver a packet. INX performs most closely to our algorithm since it quantifies interference through the number of links that interfere with another link l. The remaining metrics perform somewhat similarly because most of them are derived from one another (as discussed in Section 5.1) and therefore despite small implementation differences, there is no overarching performance improvement among the remaining metrics (as can be seen from Fig. 5.6). The delay value under all the metrics (including RI^3M) increases as demands increase, which intuitively is correct.

In Fig. 5.7, the average end-to-end values for RI^3M with Suurballe's algorithm, re-

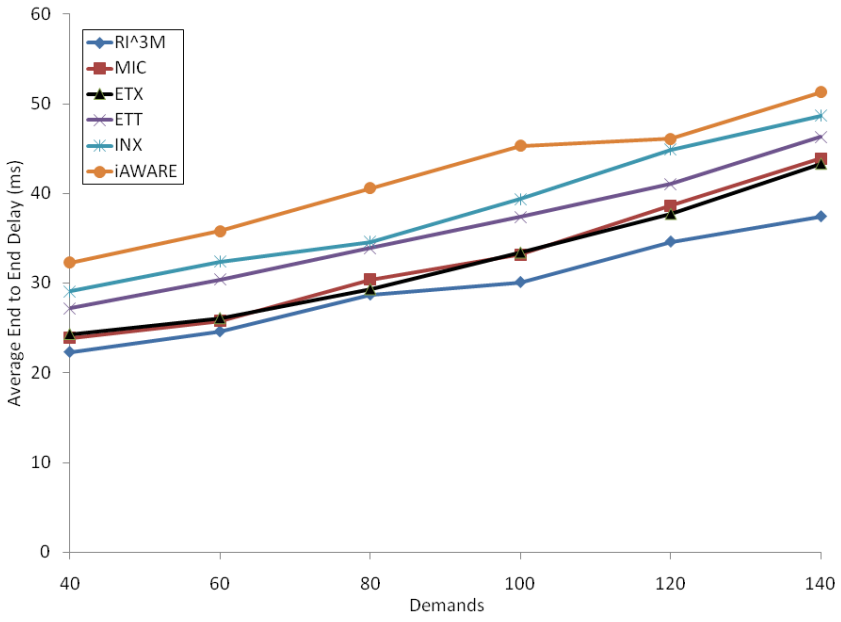


Figure 5.6: Average end-to-end delay values for RI^3M compared to prominent routing metrics in the literature

ferred to as SRA-RI^3M in the simulation graphs, is compared to the two above mentioned disjoint multipath routing algorithms. They are referred to as 2-HEAR and I2MR in the simulation graphs. The SRA-RI^3M achieves the lowest end-to-end delay compared to the other algorithms. We can justify the better performance of our results as follows: In both 2-HEAR and I2MR, the paths are formed using incomplete interference information. In 2-HEAR the SINR calculated by each node only includes those nodes within a 2-hop range which means that even if interference beyond this range occurs, it is not captured in the routing metric (inter-flow and intra-flow interference not fully accounted for). If the interference level is high beyond the 2-hop range, then any paths built may not be successful as interference may cause a drop in packets and a retransmission is required. This obviously incurs delay. A similar argument can be used with the I2MR algorithm. In our case, RI^3M quantifies the interference from both within flows and in the neighboring area.

Next, we show the average packet loss incurred from the various routing metrics

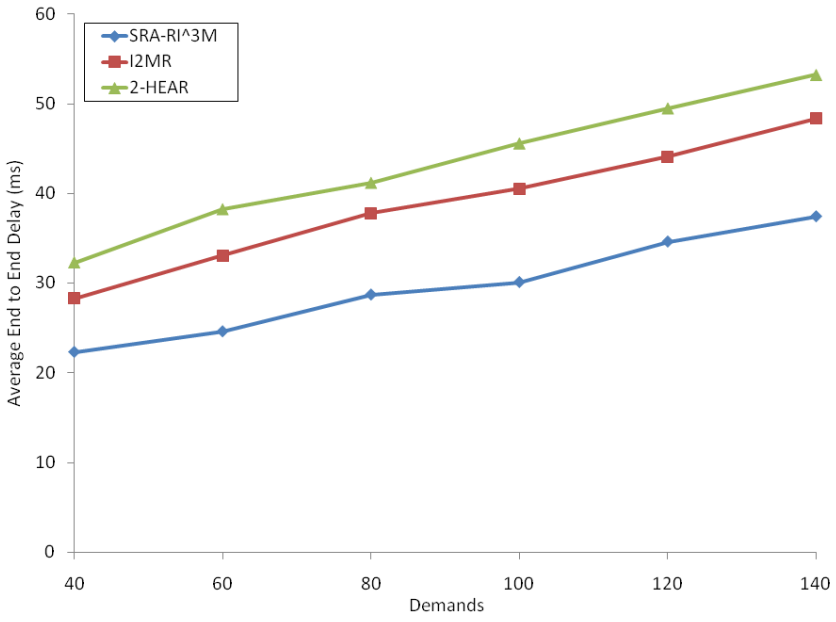


Figure 5.7: Comparison of average end-to-end delay for Suurballe's disjoint multipath routing algorithm using RI^3M (SRA-RI^3M) and two established disjoint multipath routing algorithms, I2MR and 2-HEAR

and the average flow throughput when each metric is used. Fig. 5.8 shows the packet loss ratio and Fig. 5.9 shows the average flow throughput. It can be seen that MIC and iAWARE have the lowest throughput and highest packet loss ratio at low traffic demands in comparison to the other metrics. ETX and INX have better throughput and loss ratios with low loads, but their performance decreases with high traffic demands. In Fig. 5.8 the ETT metric exhibits unstable behavior primarily because it overestimates link quality by inaccurately probing the channel. Moreover, the results show that ETT does not depend on the traffic load. Although MIC and iAWARE partially rely on ETT, these metrics employ normalization functions to smoothen ETT values and therefore become more stable. This explains the unpredictability of the results for the three metrics, ETT, MIC, and iAWARE. The remaining metrics perform intuitively as they should with greater packet loss as demands increase. The ETT, MIC and iAWARE routing metrics behave in a similar unpredictable manner for the throughput results given in Fig. 5.9 for the

same reason given above. Overall, RI^3M is able to achieve higher throughput and lower loss ratio than the remaining metrics over the varying traffic demands shown.

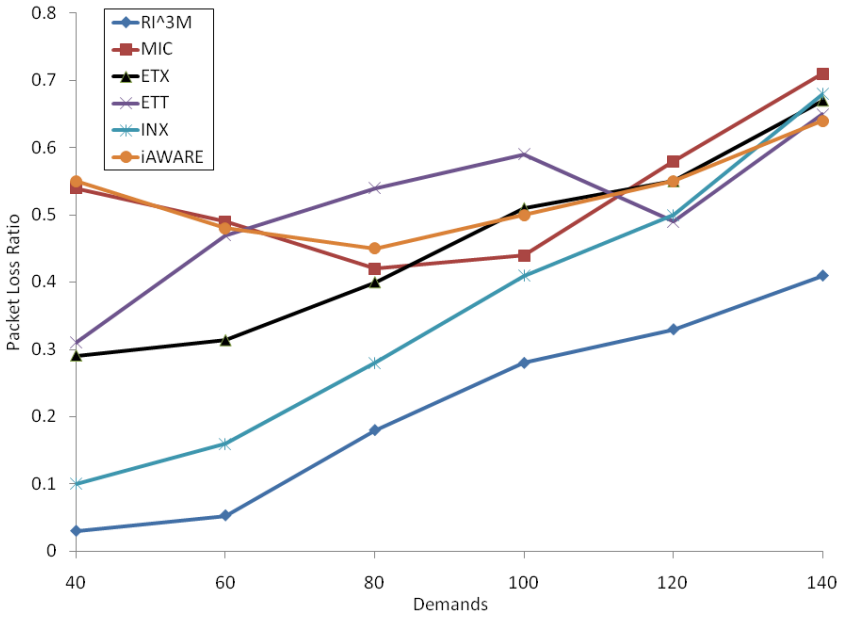


Figure 5.8: Comparison of packet loss ratio when using RI^3M versus prominent routing metrics in the literature

Performance Evaluation of LMX:M^3F

For the LMX:M^3F algorithm, we first evaluate it in terms of BBR. We compare it with MICB [15] and MMFContGr [42], respectively, as shown in the simulation graphs. We run all three algorithms on networks with different densities. Figs. 5.10 and 5.11 show the BBR results from the simulated networks with 46 (6 relays and 40 users) and 24 (4 relays and 20 users) nodes, respectively (each network has 1 base station). It can be seen that our LMX:M^3F algorithm performs the best in most cases. The blocking ratio increases no matter which algorithm is used because of heavier traffic load. The average blocking ratio difference between our solution and that of MICB and MMFContGr is 16% and 13%, respectively for network of size 46 nodes. Similarly the average difference between our algorithm and MICB and MMFContGr for network of size 24 nodes is 18%

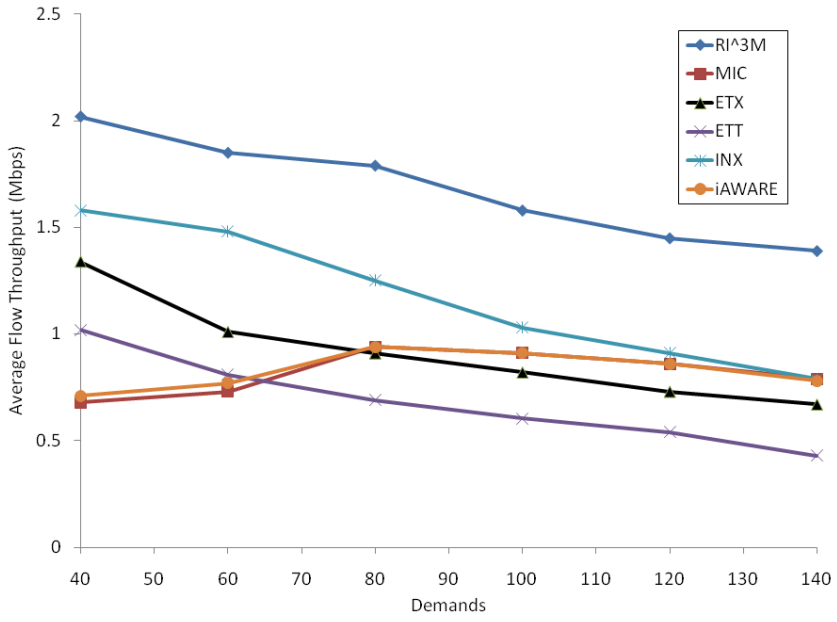


Figure 5.9: Average flow throughput generated by RI^3M versus prominent routing metrics in the literature

and 32%, respectively. Essentially the BBR indicates if a connection request for traffic is blocked. If traffic is blocked it means that there is less bandwidth on a link than there should be to accommodate the offered traffic. For best performance, the BBR should be kept as low as possible. Given the BBR results in Figs. 5.10 and 5.11, the BBR of LMX:M^3F is lower than that of the MICB and MMFContGr algorithms. Therefore, we can claim that the network performance improves under our proposed algorithm.

Next we show the real time network resource usage for all the three algorithms. Fig. 5.12 and Fig. 5.13 show the results of the bandwidth usage for the three algorithms for networks with 46 nodes and 24 nodes, respectively. As expected, LMX:M^3F uses the least amount of bandwidth for varying demands. In the case of 46 nodes, on average the bandwidth usage of LMX:M^3F compared to MICB and MMFContGr is 11% and 14% less, respectively. The bandwidth usage of the LMX:M^3F for the case of 24 nodes is on average 2% and 6% less than for the other two algorithms. The bandwidth usage shown in Fig. 5.12 shows that the LMX:M^3F algorithm clearly uses less bandwidth than

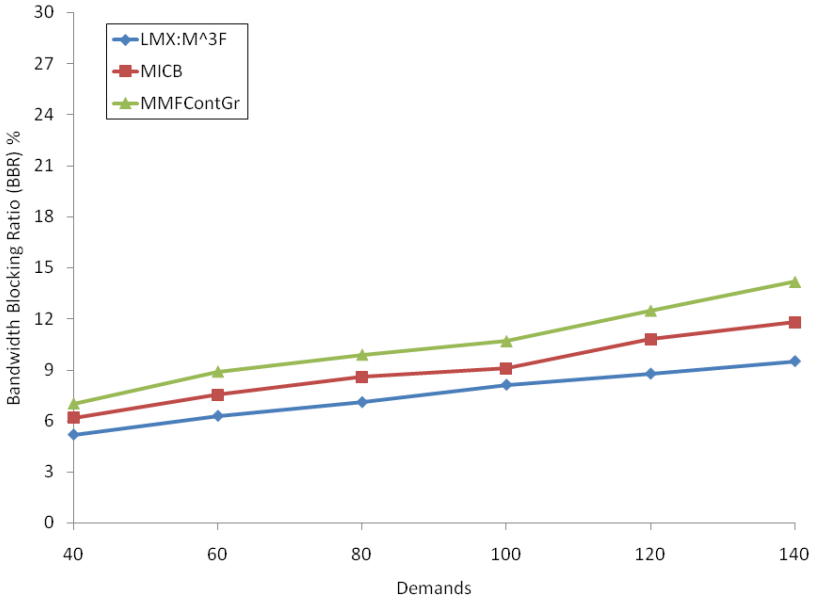


Figure 5.10: BBR comparison for networks with 46 nodes (6 relays and 40 users)

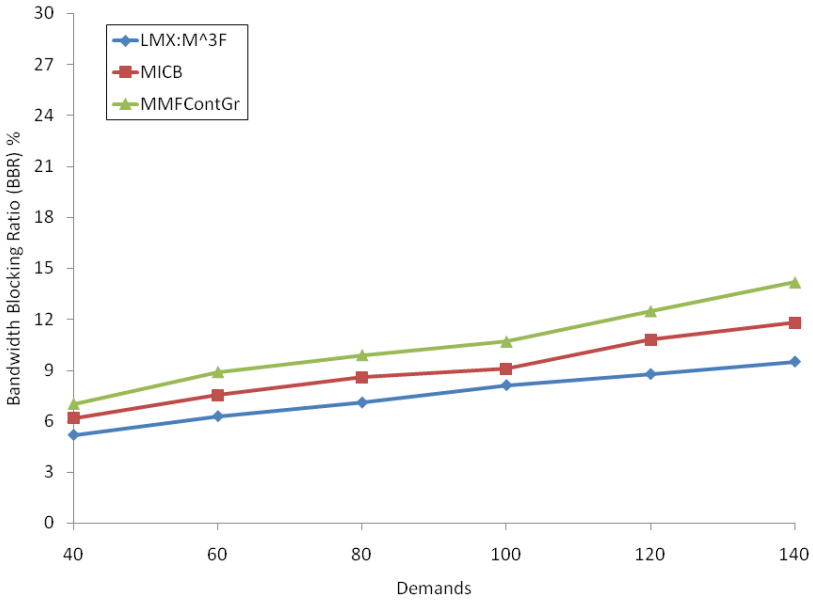


Figure 5.11: BBR comparison for networks with 24 nodes (4 relays and 20 users)

the other two approaches. However, there is less clarity in the case of Fig. 5.13 (24 nodes) because the density of the network is less. Therefore, there is not a great deal of difference between the performances of the individual algorithms even though we are simulating against the same number of varying demands. The conclusion is that our approach is more effective in network resource usage in higher density networks. Given that BWA networks are generally used in dense urban settings, our approach fits the application. However, the LMX:M^3F algorithm is time consuming to solve for very large networks with thousands of demands because each demand must be checked for bandwidth satisfaction (see Problem C). Thus, our algorithm is limited to a certain extent because of scalability.

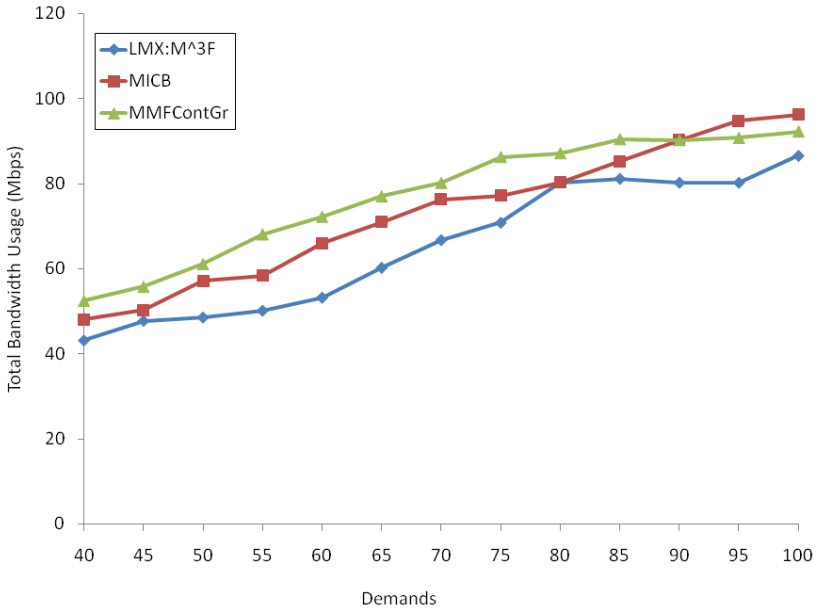


Figure 5.12: Comparison of total bandwidth usage for networks with 46 nodes (6 relays and 40 users)

Lastly, we look at the impact that our algorithm has on the load balancing of the network across various links. We compare the LMX:M^3F algorithm with that of an unbalanced routing scheme (no fairness incorporated) and a traditional max-min fair routing approach, which minimizes the load of only the maximally loaded link in the

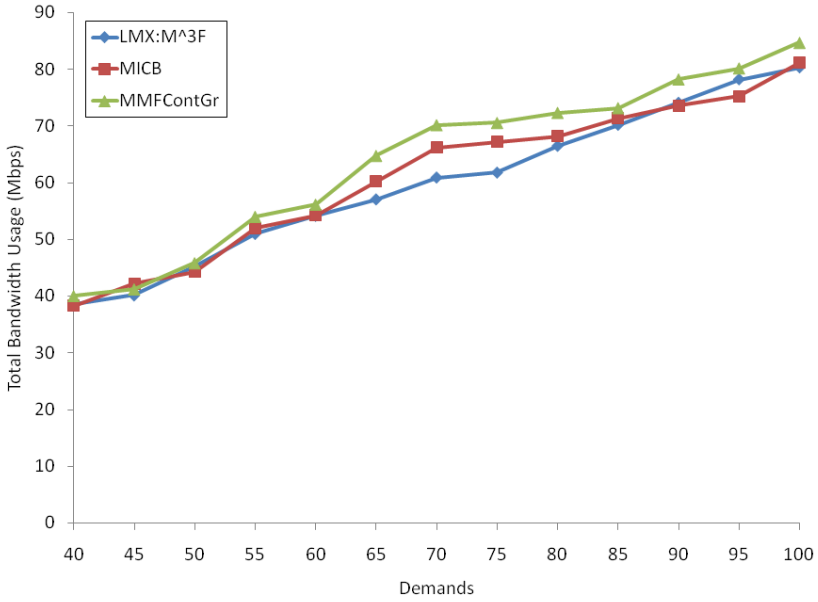


Figure 5.13: Comparison of total bandwidth usage for networks with 24 nodes (4 relays and 20 users)

network (does not look for the lexicographically highest). The results are shown in Figs. 5.14 and 5.15 where we simulate networks with 10 (2 relays, 8 users) and 15 (2 relays, 13 users) nodes (each network has 1 base station), respectively. The link number represents each individually numbered link in the network. Thus, Figs. 5.14 and 5.15 show the link load for each individual link. We see that the unbalanced routing scheme has some links with 100% utilization. When the traditional max-min routing approach is used, the link load utilization is better but there are still some links that are nearly 90% loaded. Our lexicographic bandwidth allocation algorithm performs an optimization of all the links and presents a better load balance of the traffic load as can been in the results. We observe that the LMX:M^3F algorithm generally results in approximately 75% of the links having the same load. We also see that the maximum load of any link is less than 1. This allows for spare capacity to exist on the link so that a proportionate increase in demands can be tolerated.

5.7 Chapter Summary

In this chapter, we proposed a multipath routing scheme for fair resource allocation under interference constraints. To enhance service availability and fault tolerance, we first developed a novel isotonic routing metric, RI^3M, to find disjoint paths from each user to the base station. We proved the isotonicity of the metric through virtual decomposition. In addition, we developed a fair resource allocation optimization formulation using max-min fairness (MMF). Specifically, we formulated a lexicographic MMF optimization formulation to allocate bandwidth to the routing paths in the network, determined using RI^3M. We showed that RI^3M outperforms established interference based routing metrics in the literature by improving end-to-end delay and throughput. In addition, we also showed that our lexicographic MMF bandwidth allocation algorithm provides better resource utilization in a fair manner.

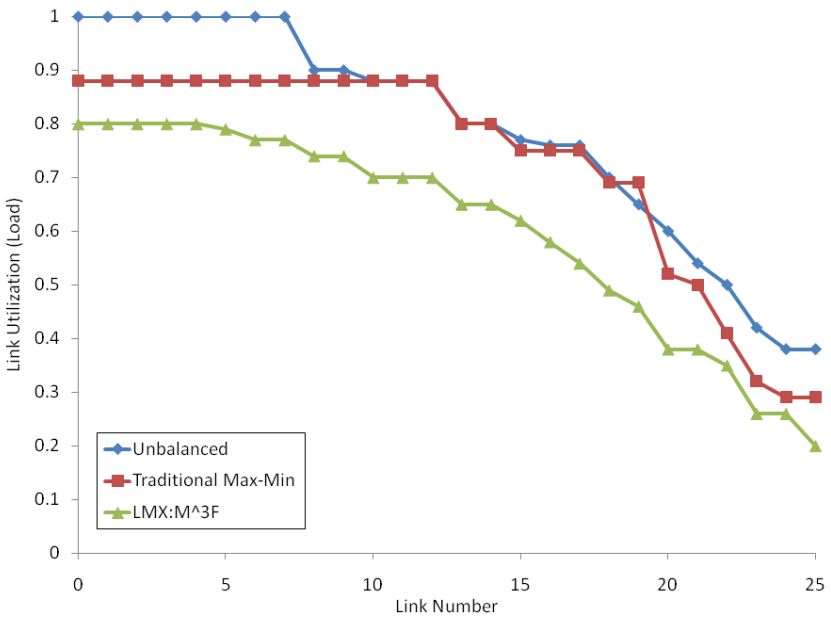


Figure 5.14: Link loads on various links for network with 10 nodes

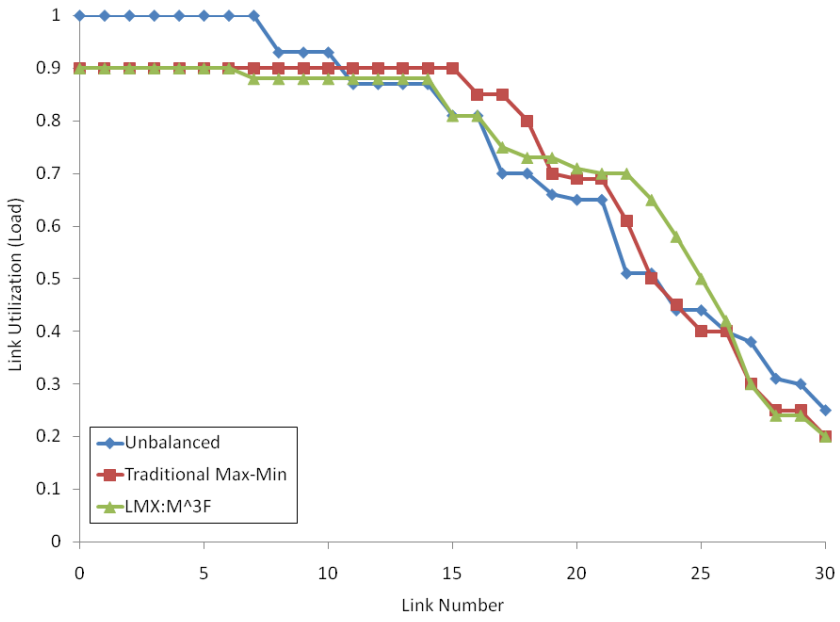


Figure 5.15: Link loads on various links for network with 15 nodes

Chapter 6

SINR Based Routing Using Distributed Mobility Prediction

In wireless networks, the movement of mobile terminals presents significant technical challenges to providing efficient wireless access to the Internet as the contact with the base station (i.e., wired infrastructure) changes with time. Thus, it is imperative that mechanisms are in place that can track and take into account this dynamic behavior, particularly when allocating resources to traffic [116].

Different groups of populations exhibit various types of mobility. For example, mobile users in a city or urban setting move very differently than those in rural, sparsely connected environments. However, within each individual population, the movements of mobile users are never totally random since they are constrained by local terrain and traffic conditions or they have a habitual route and purpose. There have been significant efforts on characterizing mobile user behavior and the resultant traffic patterns. Mobility prediction can be defined as learning and inferring from prior knowledge such as movement history, road information, etc.

To model random mobility in simulations and experiments, random mobility models such as Random Waypoint (RWP), Random Walk (Brownian motion), Random Direction

and Random Trip have been extensively used in mobile networks [117]. Along with these models, other parameters such as speed, direction, velocity etc. have been considered when used in simulations [118, 119, 120]. However, despite these various models, they can not be used directly on a single user's movement to predict its next position.

In traditional cellular networks, mobility management is performed by the base station. In such networks, mobility prediction is concerned with the user's path when it is within the coverage area of that base station. The base station manages and records the movement habits of each user within its cellular area. However, the traditional cellular architecture has a structural weakness in providing fair service because each user's QoS depends on its location and mobility within the cell. If a user is near the cell boundary, it experiences severe path loss and poor spectral efficiency compared to users near the base station. So more resources need to be allocated for cell boundary users to obtain the same throughput. To overcome these issues, fixed relays are deployed to reduce the hop transmission distance and improve spectral efficiency [74]. In RCNs, dedicated fixed relay nodes are placed to help forward traffic to and from the base station. The majority of user nodes in a RCN tend to connect to a relay node due to proximity; the relay nodes act as intermediaries between the mobile user and base station. From the point of view of the user, a relay acts like a pseudo-base station by collecting movement information directly from the users. Thus, in a RCN, mobility management control can be transferred to the relay nodes, thereby forming a distributed mobility management scheme.

Mobility management that involves movement prediction relies on the availability of prior information on the user's mobility behavior. Recently, prediction schemes using variations of the Markov model, particularly the Hidden Markov Model (HMM) have been proposed for resource management purposes in ad hoc networks [65, 121]. These schemes use control theoretic frameworks to dynamically allocate resources to users. Similarly, mobility prediction in cellular networks has also been researched [122, 123, 124]. However, except for [122], which deals with a call admission control scheme, the

other schemes are focused solely on user prediction without an emphasis on resource management. In addition, these works deal with the traditional cellular architecture in which mobility management can be performed in a centralized manner. Our network architecture facilitates a distributed solution to mobility management.

Interference is influenced by the node mobility and can lead to performance degradation. The mobility properties of the users (i.e., mobility patterns, speed, direction etc.) can cause new interference to be induced at neighboring nodes [47]. Specifically, if a node n moves from an area of low interference, A, to one of high interference, B, then any transmission from n will contribute to the interference of area B. Interference can be controlled/mitigated in the network layer i.e., with routing. Therefore, our focus is on the joint interaction of the physical layer and the network layer. The use of physical layer information in terms of interference from signal strength is used in our routing decisions.

In order to design an effective routing protocol that mitigates the interference experiences of the wireless links, the mobility of the users must be considered. Mobility assisted routing has been studied in the literature for several years, more recently focusing on ad hoc and delay tolerant networks [125, 126, 127]. Both [125] and [127] deal with modeling random user movements for the purpose of routing while [126] investigates group mobility patterns to implement routing. However, none of these works discuss the direct impact of interference on the routing protocols. More recently, in [47], mobility aware routing using interference constraints was developed. However, the interference is modeled using the protocol model which induces binary conflicts (either two links interfere or they do not despite neighboring simultaneous transmissions) which is not true in practice. Routing protocols using SINR to model interference has been studied in [81], [128] and [129]. Although SINR is used to model resulting interference, the routing is performed on static networks.

In this chapter, we first develop a distributed mobility prediction model using HMM to determine the locations of the user nodes at a time instant t. Second, we develop a

routing protocol which uses the location information of the mobile user to determine the interference level on links in its surrounding neighborhood. We use SINR as the routing metric to calculate the interference on a specific link (link cost). We minimize the total cost of routing as a cost function of SINR while guaranteeing that the load on each link does not exceed its capacity, thereby determining least interfering paths from each user to the base station. The routing protocol and the proposed solution are solved using a combinatorial optimization technique, known as the minimum-cost flow problem in the operations research literature. The work presented in this chapter is being prepared for submission [73].

6.1 Problem Preliminaries

In this work, we consider the network architecture discussed in Chapter 2, Section 2.1 in which each mobile user connects to a relay node or directly to the base station. Users move randomly across the network and can communicate with each other by exchanging data for routing purposes.

To understand the interaction between the various components of our framework, we provide a block diagram shown in Fig. 6.1. The block diagram describes the mobility prediction mechanism and its relationship to the SINR based routing algorithm. The prediction of the users' movement is driven by an HMM meaning that the HMM is used to represent the mobility pattern of the users. The current mobility information and the history of the users' past movements is used to make predictions. Each relay contains a mobility database that contains the mobility information of each user connected to it. Specifically, the database keeps track of which users are connected to the relay and which users have moved away to another relay, base station or cell. This mobility information along with the HMM is used to determine the SINR calculations and thereby the routes to each user.

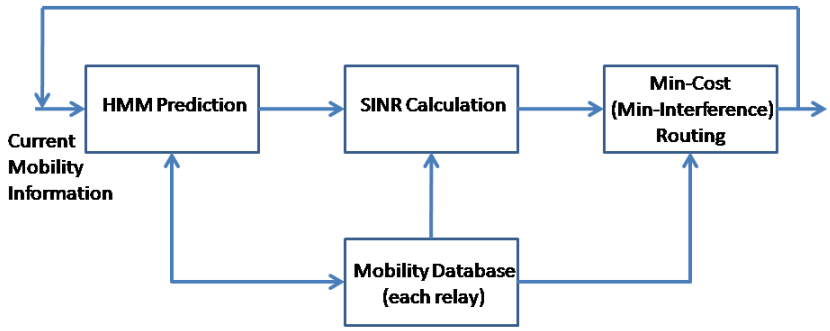


Figure 6.1: Block diagram that illustrates the interaction between the mobility prediction scheme and the interference aware routing protocol

6.2 Distributed Mobility Prediction Model

The prediction model discussed in this section aims to solve the following problem:

Consider a mobile user connected to relay node A. The user may move away from A to relay node B after some time. Using the history and transition paths, what is the likelihood that a user makes the transition from A to B?

This problem has been dealt with using a Markov chain model [121]. However, the drawbacks of using a simple Markov chain model can be illustrated as follows. Referring to Fig. 6.2, consider a RCN with 4 relay nodes, A,B,C and D. Initially assume that a user connected to A moves from A to connect to any of the other relays, B,C or D. The transition from A to any of the other relay nodes may depend on proximity, signal strength, etc. The Markov model given in Fig. 6.2 shows the changes in direction as a sequence of probabilities based on past states. The transition probability for the next state is based on the most recent state. However, an external observer may not be able to see all of these transitions. Some transitions may be hidden from the observer by the user or the system. For instance, if a user connects to any of the relay nodes, the

observer may only see the movement of the user from one relay to another but may not be able to determine which relay the user is connected to. Thus, the relay nodes are the hidden states and the locations are the observable states. Because there is no one-to-one mapping between these two states, the problem is to identify the relays corresponding to the location of the user.

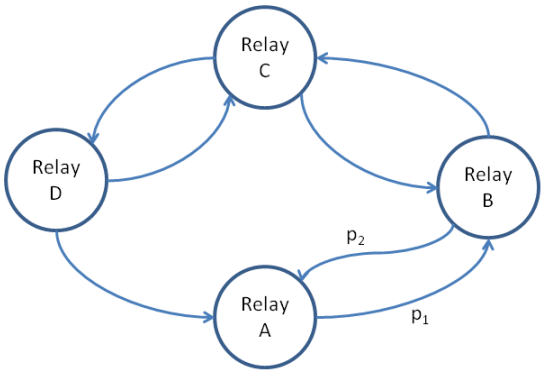


Figure 6.2: Example to show a simple Markov chain that depicts the transitions of a mobile user to various relay nodes

6.2.1 Hidden Markov Model (HMM)

An HMM has two kinds of stochastic variables: state variables (hidden variables) and the output variables (observable variables). An HMM can be defined as follows:

S : $\{s_1 s_2 ... s_N\}$ are the N hidden states of the system

O : $\{o_1 o_2 ... o_N\}$ are the values of the observed sequences

Π : $\{\pi\}$ is the initial state probabilities. π_i indicates the probability of starting in state i

$A = \{a_{ij}\}$ are the state transition probabilities where a_{ij} denotes the probability of moving from state i to j

$$a_{ij} = P(t_k = s_j | t_{k-1} = s_i)$$

$B = \{b_{ik}\}$ are the observation state probabilities where b_{ik} is the probability of emitting symbol k at state i

$$b_{ik} = P(o_k|t_k = s_j)$$

We shall also assume that the mobile user's time in a given state is a random variable taking values in the set $\{1, 2, ..., D\}$. $P = p_n(d)$ is the probability distribution function, where $n \in S$. The 4-tuple (A, B, P, π) provides a complete specification of the HMM for the system considered in this paper.

6.2.2 Prediction Model Using HMM

To track the state of a mobile user we apply two approaches: 1) forward-backward algorithm and 2) re-estimation algorithm for the HMM parameters discussed above. The main steps of the tracking algorithm can be summarized as follows:

1. Apply HMM re-estimation algorithm to obtain initial estimates of (A, B, P, π) of the HMM model.

2. Apply the HMM forward-backward estimation algorithm to predict at time t the next state of a user.

3. Obtain refined estimates of (A, B, P, π) by again applying the HMM re-estimation algorithm to the given observation sequences.

In mobile systems, up to date information regarding users' movements is difficult to obtain. Estimation of the mobility model parameters must in general be made based on incomplete data. Due to physical constraints, transmission of location data may not occur frequently enough to allow precise tracking of the user's state at all times. The task of estimation from insufficient data involves two important aspects: (a) estimation and prediction of the users' movement behavior and (b) re-estimation of the model parameters based on incomplete information. Before discussing the estimation and re-estimation

algorithms, we define the observation interval as the time interval during which observations occur. The observation interval is assumed to be segmented into T subintervals indexed by $1, 2, ..., T$. Observations may not be necessarily available in each of the T subintervals.

Forward-Backward Algorithm

A forward-backward algorithm is an algorithm for computing the probability of a particular observation sequence in the context of hidden Markov models. It is essentially an inference algorithm for HMM and consists of two steps. The first step of the algorithm computes a set of forward probabilities which provide the probability of observing the first k observations in the sequence and ending in each of the possible Markov model states (i.e., probability of ending up in any particular state given the first k observations). The second step of the algorithm computes a set of backward probabilities which provide the probability of observing the remaining observations given an initial state (i.e., probability of observing remaining observations given any starting point). These two sets of probabilities can then be combined to provide the probability of being in each state at a specific time during the observation sequence. The forward-backward algorithm can thus be used to find the most likely state for a hidden Markov model at any time. In [130], a forward-backward algorithm was devised to estimate a HMM from observations. The algorithm has a computational complexity proportional to D, where D is the maximum value of the time spent at a specific state for all states. For our model, we define the following forward and backward variables:

Forward variables:

$\alpha_t(n) = P[o_1^t, \text{ state } n \text{ sojourn ends at } t], t \geq 1$

$\alpha_t^*(n) = P[o_1^t, \text{ state } n \text{ sojourn begins at } t+1], t \geq 1$

Backward variables:

$\beta_t(n) = P[o_t^T |$ sojourn in state n begins at $t], t \leq T$

$\beta_t^*(n) = P[o_t^T |$ sojourn in state n ends at $t-1], t \leq T$

The forward variables are then computed inductively for $t = 1, 2, ..., T$. Similarly, the backward variables are computed inductively for $t = T, T-1, ..., 1$. After computing the forward and backward variables, a state estimate can be found. Define,

$$\gamma_t(n) = P[o_1^T; s_t = n]$$

as the probability that s is observed to be in state n at time t. Then the estimate of s_t is given by

$$\hat{s}_t = \arg\ \max_{1 \leq n \leq N} \frac{\gamma_t(n)}{P[o_1^T]}, t = T, T-1, ..., 1$$

Re-estimation Algorithm

A simple iterative procedure for re-estimating the HMM parameters is reported in [130]. By applying the well-known EM (Expectation/Maximization) algorithm [131], it can be shown that this iterative procedure is increasing in likelihood. The overall computational complexity of the re-estimation algorithm is essentially proportional to T. Thus, the parameters of the HMM model can be estimated effectively within our framework.

6.3 SINR Based Routing Using HMM Prediction

6.3.1 Challenge of Routing with Interference and Mobility

Using the HMM approach we are able to track the movement of the users to determine which relay it is connected to. Given this information, routing from the connected relay to the base station can take place through multiple hops. Note that knowing to which relay a user is connected is imperative to the calculation of interference. To route in the presence of mobility and interference using link based metrics is a fundamental challenge. Under generic shortest path routing, the path length (which depends on the link metric) is the only factor that decides the best route between any source and the BS. Various examples of link metrics in the literature, namely Euclidean distance, residual battery charge, and buffer occupancy, depend solely on the two nodes forming the link. They are independent of the existence of other paths from other users and the BS or their shortest path routes. This, in turn, has led to the notion of link metrics and link-based routing. However, interference depends on the existence of other sources/intermediate relays and their spatial separation. Hence, the routing decision of a given source-BS pair becomes coupled to the routing decision of other source-BS pairs.

To illustrate this, assume node a is transmitting to next hop b and node u is transmitting to next hop v as shown in Fig. 6.3(a). According to the non-linear decay of power with distance, governed by $P_r(z) = P_t * z^{-\alpha}$ where P_t is the transmitted power, z is the distance between transmitter and receiver and α is the pathloss exponent, the amount of interference at node v from transmitters other than u is given by $I_v^u = P_{ab} * z_{av}^{-\alpha}$. If node a was transmitting to a different node (i.e., node c), as shown in Fig. 6.3(b), then the amount of interference seen at node v would be different: $I_v^u = P_{ac} * z_{av}^{-\alpha}$. Thus, the interference induced on link (u, v) (needed to compute its link metric) depends on the routing decision of transmitter a which, in turn, depends on the routing decision of transmitter u. Couple this scenario with mobility in which node a is moving, then a more

refined time based routing metric is required to gauge both interference and the location of the node at that time.

To determine appropriate routing paths from the relay to the base station that are cognizant of interference, we use SINR as a routing metric. The SINR is an effective and practical metric to gauge link quality because it takes interference and noise as well as signal strength into account. Furthermore, with user nodes moving, poor links are unpredictable and thus SINR based routing decisions are useful to discover more robust paths.

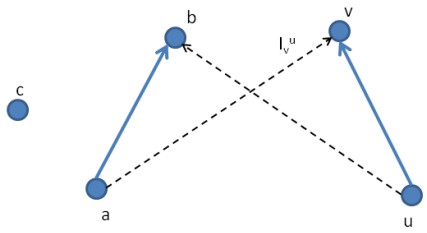
6.3.2 Problem Formulation

For our analysis, we model the RCN as a graph, $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes (relays, users and base station inclusive) and $\mathcal{E}$ is the set of links. Let V_N be the set of users and let V_M be the set of relays. Note that the network has only one base station, denoted BS. The successful reception of a packet depends on the received signal strength, the interference caused by the simultaneously transmitting nodes, and the ambient noise level η. The SINR of a link (i, j) is given as follows (same as Eq. 2.1).

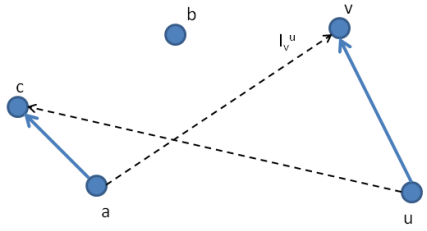
$$SINR_{ij} = \frac{P_j(i)}{\eta + \sum_{k \in V'} P_j(k)} \geq \beta \tag{6.1}$$

where $P_j(i)$ is the received power at node j due to node i, V' is the subset of nodes in the network that are transmitting simultaneously, and β is the SINR threshold. Our proposed routing protocol is implemented to route data using the least interfering path out of all path possibilities. If a link has a high SINR, it is an indication that it is experiencing low interference.

Each link (i, j) has an associated cost which is derived from the SINR value calculation. Each link also has an associated capacity denoted u_{ij}. The capacity is formulated using Shannon's formula, given in Eq.6.2.



(a) Node a is transmitting to node b and node u is transmitting to node v



(b) Node a transmits to node c while node u continues to transmit to node v

Figure 6.3: Illustration of the challenge of defining an interference aware routing metric in the presence of simultaneous transmissions and mobility

$$u_{ij} = log_2(1 + SINR_{ij}) \tag{6.2}$$

In addition, the flow of packets from node i to its neighbor j over wireless link (i, j) is represented by f_{ij}.

6.3.3 SINR Based Routing

The position of each user node at time t affects the cumulative SINR on each link (i.e., SINR will fluctuate with time depending on where the user node is and what noise or transmissions are present around it). The SINR is also affected by the path loss model and channel gain. The SINR at time t on link (i, j) is given by Eq.6.3.

$$SINR(t)_{ij} = \frac{G_{ij}P_j(i)(t)}{\eta + \sum_{k \in V'} X_k G_{kj} P_j(k)(t)} \geq \beta \tag{6.3}$$

where G_{ij} is the channel gain on link (i, j)[1], $P_j(i)(t)$ is the received power at node j due to node i at time t, and k is a simultaneously transmitting node. X_k is a binary variable which denotes whether node k is transmitting or not. The corresponding capacity, u_{ij}, is then modified to be

$$u_{ij} = log_2(1 + SINR_{ij}(t)) \tag{6.4}$$

In order to determine the least cost (least interfering) paths, we use the minimum cost flow optimization method. Essentially, the minimum cost flow problem is finding the cheapest possible way of sending a certain amount of flow through a network. In our case, the cost of a link is motivated by the amount of interference on that link due to neighboring transmissions and/or noise. As we are using SINR as the routing metric, the higher the SINR, the better the link quality. Therefore, we want to minimize the *inverse* of the SINR value.

[1]In the simulations, the channel gain of each link is calculated using a Rayleigh fading model and an appropriate path loss factor.

The objective of the SINR routing problem is to deliver all the data packets generated by the user nodes to the base station in the most cost-effective (least interfering) manner without exceeding the link capacities. Formally, the problem can be stated as follows.

$$\text{minimize} \quad \sum_{(i,j)\in\mathcal{E}} SINR(t)^{-1} f_{ij}(t) \tag{6.5}$$

subject to

$$\sum_{j:(i,j)\in\mathcal{E}} f_{ij}(t) - \sum_{j:(j,i)\in\mathcal{E}} f_{ji}(t) = d_i(t), \forall i \in V_N \tag{6.6}$$

$$\sum_{k:k\in V_M \cup BS} \Big(\sum_{j:(k,j)\in\mathcal{E}} f_{kj}(t) - \sum_{j:(j,k)\in\mathcal{E}} f_{jk}(t) \Big) = - \sum_{i:i\in V_N} d_i(t) \tag{6.7}$$

$$0 \leq f_{ij}(t) \leq u_{ij} \tag{6.8}$$

$$f_{ij}(t) \in Z^+ \tag{6.9}$$

In the above formulation, d_i represents the rate at which the data packets are generated at user node i per unit time. The first constraint (Eq. 6.6) ensures flow conservation at each node. The second constraint (Eq. 6.7) ensures that the base station and/or relay nodes receive all the packets generated by all the user nodes. The flow of packets on a link must not exceed its capacity and this is ensured by the third constraint (Eq. 6.8). The fourth constraint (Eq. 6.9) ensures that the packet flow values are integers.

The complexity of the above minimum cost flow problem is derived from [80] and shown to be $\bigcirc(\epsilon^{-2} m(m+n) log M)$ where m is the number of links in the network, n is number of nodes in the network (users plus relays) and M is an integer parameter that specifies the largest cost on the link (largest SINR value).

Solution

The above defined problem is similar to the minimum-cost flow problem, known in the operations research literature [132]. We will convert the above problem into the minimum-

cost circulation problem as follows.

1. Add a super source x, and a super base station node y, to the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$.

2. Add directed links (x, i), connecting the super source x to node i, for all $i \in V_M \cup V_N$. Set costs of these links to 0 and the capacities to d_i.

3. Add directed links (j, y) connecting the base station and relay nodes to the super base station y. Set costs of these links to 0 and the capacities to infinity.

4. Add a directed link (y, x) connecting the super base station y to the super source x. Set the cost of the link (y, x) to $-|\mathcal{V}|\beta$ and the capacity to infinity, where β is the minimum of any link cost (lower bound of SINR).

5. The modified graph is defined as $\mathcal{G}'(\mathcal{V} \cup \{x, y\}, \mathcal{E} \cup \mathcal{E}')$, where $\mathcal{E}' = \{(x, i) : i \in V_N\} \cup \{(j, y) : j \in V_M \cup BS\} \cup \{(y, x)\}$.

The minimum-cost problem given above is solved using the well-known minimum-cost flow algorithm given in [133]. An advantage of the minimum-cost flow algorithm is the integrality of flows. If all link capacities and expected data rates of nodes are integers, then the minimum-cost flow algorithm can find paths with integral flow values.

Analysis of the Solution

The minimum path cost formulation given in Eqs. 6.5-6.9 determines the least interfering paths by minimizing the inverse of the SINR values of the links in the network. In addition, it also routes flows such that the link capacities are not violated. Pushing more flow from node x to node y will decrease the overall cost of the flow due to fact that the link from node y back to node x has sufficiently large negative cost. It is clear that the maximum flow is bounded from above by $F = d_1 + d_2 + ... + d_{|V_N|}$ because F is the maximum possible flow going out of node x, the super source. There are two possibilities that have to be analyzed.

Case 1: $\sum_{i:i\in V_N} f_{xi} = \sum_{i:i\in V_N} d_i$

In this case, all the links of the form (x, i), $i \in V_N$ are saturated. The maximum-flow is restricted by the capacities of these links. Consider a link $(x, 1)$ having the capacity d_1. Since all the (x, i) links are saturated, the input flow at node 1 must be $d_1 + \sum_{j:(j,1)\in\mathcal{E}} f_{j1}$ and the output flow must be equal to the input flow (flow conservation). There must be paths from node 1 to the base station which carry the flow $d_1 + \sum_{j:(j,1)\in\mathcal{E}} f_{j1}$. The same argument holds for other nodes.

Case 2: $\sum_{i:i\in V_N} f_{xi} < \sum_{i:i\in V_N} d_i$

In this case the maximum flow is restricted by the capacities on the actual links ($(i, j) \in \mathcal{E}$) of the network. The minimum cost flow algorithm will identify the paths from the user node i to the base station which carries the flow d'_i where $0 \leq d'_i \leq d_i$, $\forall i \in V_N$. The flow on the links (x, i) would be d'_i, $\forall i \in V_N$.

6.4 Performance Evaluation

6.4.1 Simulation Model and Performance Metrics

The HMM prediction model and SINR based routing scheme have been simulated to verify their performance. The prediction engine based on HMM is first separately tested for accuracy in predicting the future mobility of users. For comparison, we use a generic Markov chain and a second-order Markov chain to gauge the prediction accuracy of the three methods. A second-order Markov chain can be defined as

$$P = P[Relay_{next}|Relay_{current}, Relay_{previous}]$$

When the users make first contact with a relay, there is no history of data from this user that can be utilized, so the initial parameters of the HMM are randomly generated using a uniform distribution. Specifically, the number of users, to which relay each

user is associated and the initial transition probabilities are generated. Once the users begin to move, its movement history is tracked and stored in the databases of each relay for prediction. To evaluate the SINR routing, we integrate the HMM and the routing protocol to show the packet loss, end-to-end delay and routing overhead. As benchmarks we compare with two interference aware routing metrics that use SINR as the routing metric and method of interference quantification, given in [81] and [128].

We use NS-2 to simulate our evaluations and use CPLEX to solve the optimization formulation for the minimum cost SINR based routing algorithm. It is assumed that 50 user nodes move over a cellular area of radius 3km. In our simulations, we use the Rayleigh fading model as the radio propagation model. The Rayleigh fading model allows us to capture radio propagation signals that are not in the line of sight (i.e., when there are many objects in the environment that scatter the radio signal before it arrives at the receiver). The received power, $P_j(i)$, is calculated according to the radio propagation model at the receiver. The noise, η, is AWGN. The propagation channel of the Rayleigh fading model is assumed to have a data rate of 2Mbps. The pathloss exponent (LOS/NLOS) is set to 2.35/3.76. We also assume the radio transmission range to be 250m. With a data transmission rate of 2 Mbps, each run has been executed for 1000 sec of simulation time. Constant bit rate (CBR) sources transmit UDP-based traffic at 4 packets per second and the data payload of each packet is 512 bytes long. The speed of each node is varied from 10 m/hr to 60 m/hr.

To evaluate the distributed HMM scheme, we look at the prediction accuracy of the mobility model. The prediction accuracy is one of the most important metrics for the verification of any mobility prediction model. Prediction accuracy is defined as the ratio of the number of times a user moves to different relays to the ability of the system to predict the location. For example if node n moves to relay A and then to relay B, and our prediction model predicts correctly that it moved to A but not B, then the prediction accuracy is 50%. To evaluate the SINR based routing scheme, we evaluate the following

performance metrics:

- Packet Delivery Ratio: ratio of the number of data packets successfully delivered to the destination over the number of data packets sent by the source.
- End-to-End Delay: the average delay for a packet to reach from the source to the BS.
- Routing Overhead: Routing overhead is defined as the number of routing messages transmitted per second.

As benchmarks we compare with two interference aware routing metrics that use SINR as the routing metric, given in [81] and [128].

6.4.2 Simulation Results of HMM

We first look at the performance of the distributed HMM for two random users in the network and compare against the Markov and 2-order Markov chains. Fig. 6.4 and Fig. 6.5 show the prediction accuracy in percentages for the two users in the network. From these figures we can conclude that the HMM has an advantage in prediction accuracy compared to the Markov and second-order Markov chains. The results also show that the HMM can better adapt to a user's change in movement. In other words, the HMM learns faster than the generic Markov based approaches.

6.4.3 Simulation Results of SINR Based Routing

The performance of the SINR routing algorithm is evaluated compared to two SINR based routing approaches given in [81, 128]. In [128], (denoted Kortebi in the simulation graphs after the primary author) the SINR based routing is performed using a variation of the shortest path algorithm. In [81], an algorithm, 2-HEAR, is developed in which a routing metric is used such that a node calculates the SINR to its neighboring links

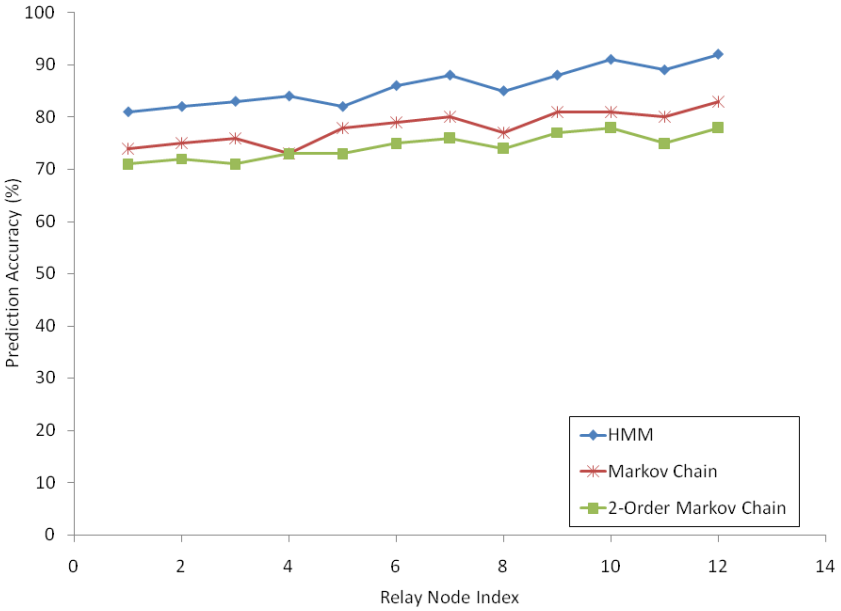


Figure 6.4: Comparison of prediction accuracy for our proposed HMM model, and a generic Markov chain and second-order Markov chain for User 1

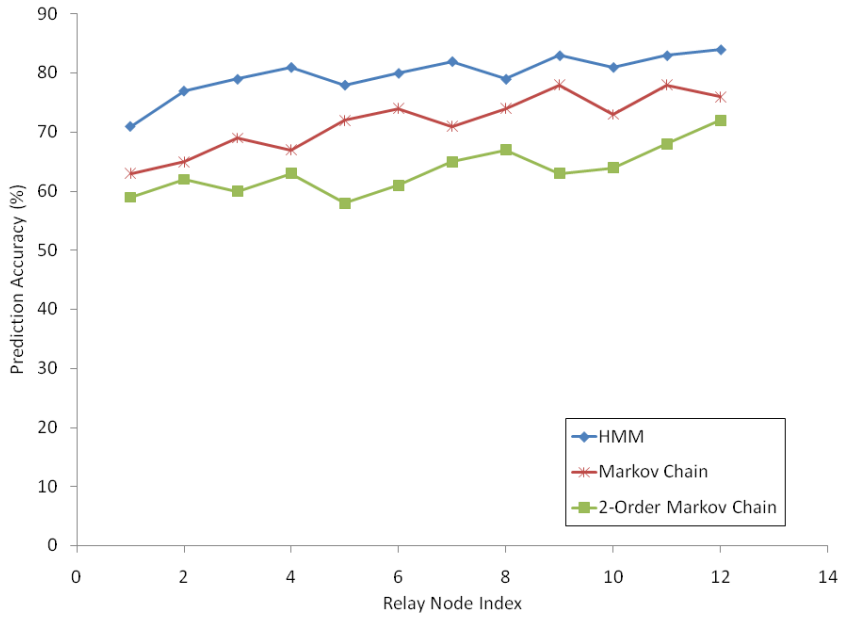


Figure 6.5: Comparison of prediction accuracy for our proposed HMM model, and a generic Markov chain and second-order Markov chain for User 2

based on a 2-Hop interference range only. For measuring the performance metrics, the noise level is varied within a meaningful range i.e., (-90dBm to -80dBm). When the noise is varied, the node speed is kept constant (50 m/hr). When the speed of nodes is varied, the noise level is kept constant, at -85dBm. Each simulation run has been executed for 900 seconds.

We first look at the packet delivery ratio for our SINR based routing scheme and its two relevant counterparts in the literature. In Fig. 6.6 and Fig. 6.7, the results of the packet delivery ratio for varying environmental noise and sampling rate are shown. The node speed is kept constant at 50m/hr. From the results it can be seen that our approach provides better packet delivery ratios when compared to the other approaches. We can justify the better performance of our results as follows: In both 2-HEAR and Kortebi, the paths are formed using incomplete interference information. In 2-HEAR the SINR calculated by each node only includes those nodes within a 2-hop range which means that even if interference beyond this range occurs, it is not captured in the routing metric. If the interference level is high beyond the 2-hop range, then any paths built may not be successful as interference may cause a packet drop and therefore a retransmission is required. A similar argument can be made with Kortebi's approach since only a shortest path algorithm is implemented using SINR. Thus, a long path with better cumulative signal strength may be available and is not captured by Kortebi's approach.

In addition, we also look at the effect of varying the sampling rate against the packet delivery ratio and show that with increasing T_w, the packet delivery ratio increases. The results are shown in Fig. 6.8.

We next look at the packet delivery ratio for varying node speed while keeping the noise constant at -86dBm. Fig. 6.9 shows the packet delivery ratio of our proposed routing scheme, Kortebi and 2-HEAR. Compared to the other two approaches, our scheme presents higher packet delivery ratio by reducing packet loss by up to 27%. The significant reduction of packet loss from our scheme can be attributed to more reliable routes and

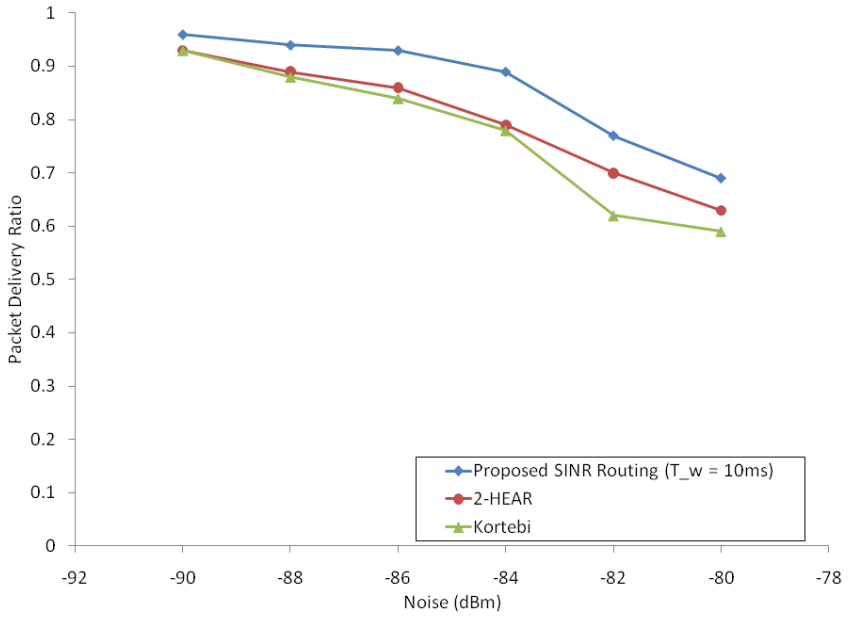


Figure 6.6: Packet delivery ratio for $T_w = 10ms$

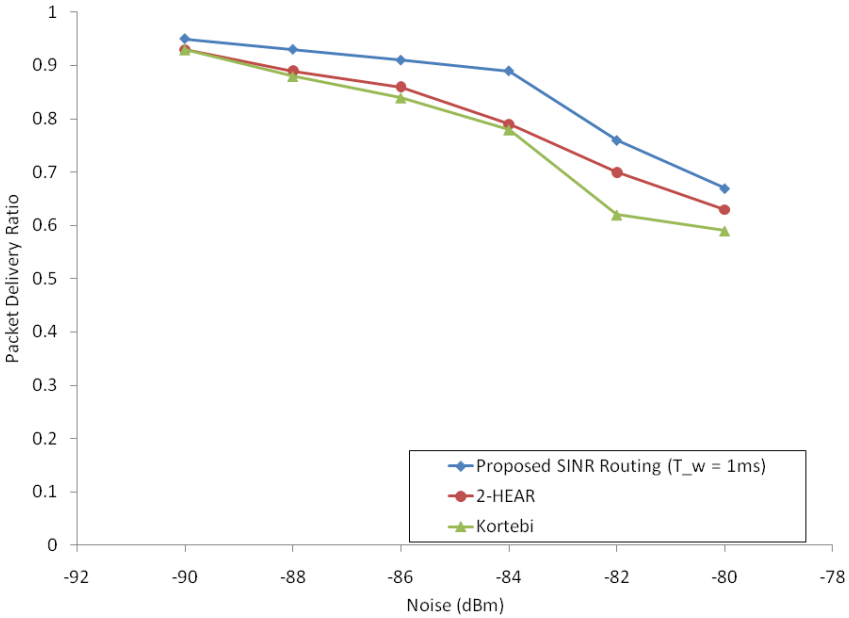


Figure 6.7: Packet delivery ratio for $T_w = 1ms$

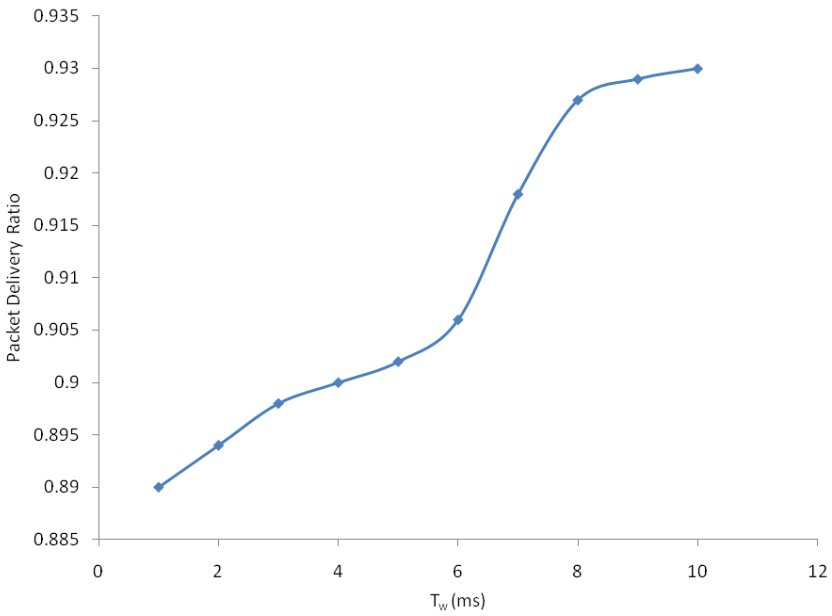


Figure 6.8: Effect of varying T_w values on packet delivery ratio

less overhead as explained above. It can also be observed that the packet delivery ratio for all three schemes decreases with increasing speed, primarily because of unavoidable errors in SINR measurement at high velocities.

We next evaluate the end-to-end delay and routing overheads of our approach for varying node speeds. The results are shown in Fig. 6.10 and Fig. 6.11, respectively. The average end-to-end delay is improved compared to Kortebi and 2-HEAR mainly due to more robust routes and less route discoveries, which minimize the potential possibility of link breakage. Note that the more reliable routes in our scheme significantly reduce the number of route discoveries and retransmissions. For all the three protocols, the average end-to-end packet delay is increased as the speed increases. Similarly, the routing overhead of our scheme is less than the other two approaches even at high speeds.

6.5 Chapter Summary

Mobility and interference jointly influence the performance of wireless networks. In this chapter, we developed a SINR based routing approach that determines least interfering

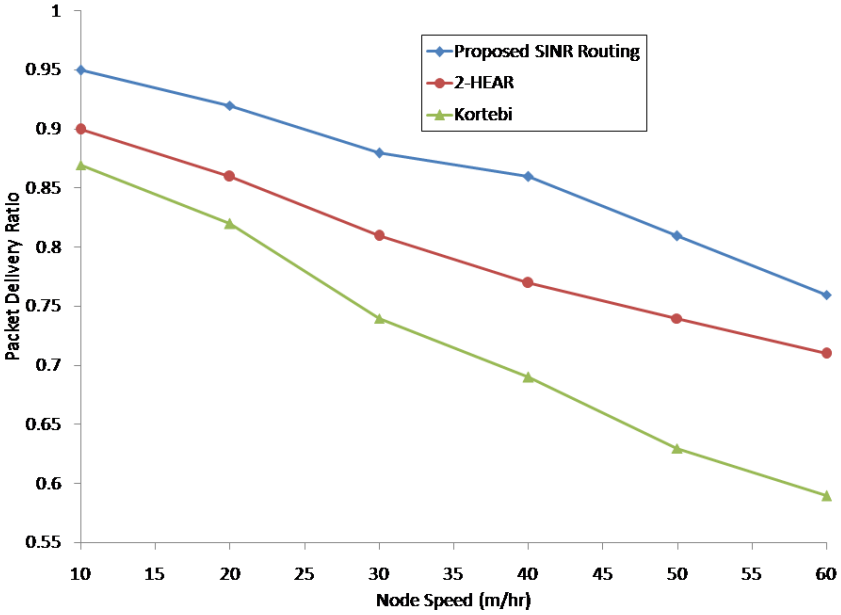


Figure 6.9: Packet delivery ratio for $T_w = 1ms$ and varying node speed

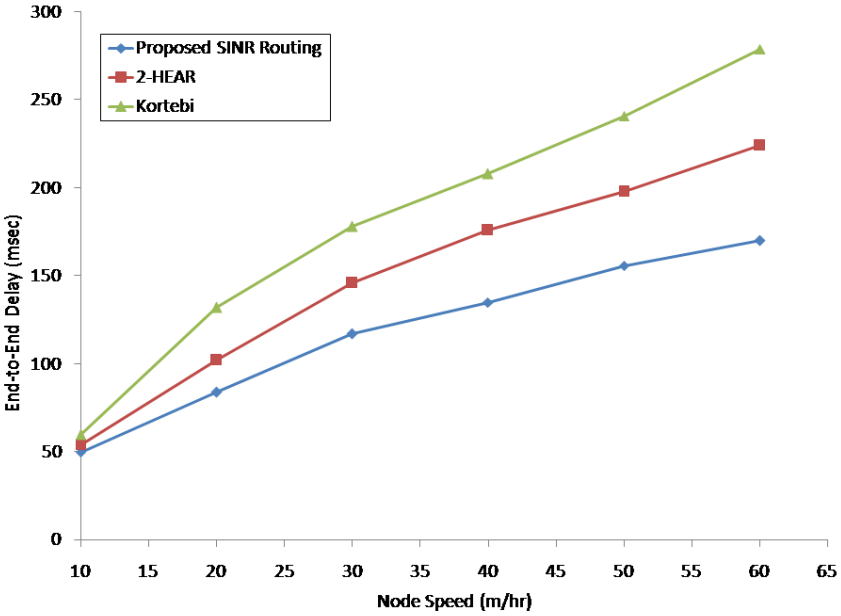


Figure 6.10: End-to-end delay for $T_w = 1ms$ and varying node speed

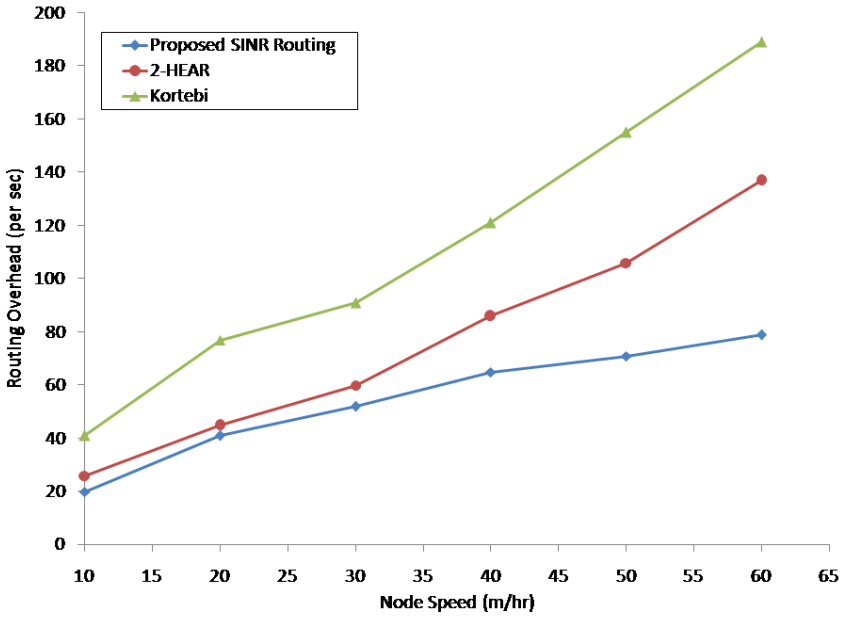


Figure 6.11: Routing overhead for $T_w = 1ms$ and varying node speed

paths for each user to the base station. We formulated and solved the routing problem using a minimum cost (in our case minimum interference) flow optimization technique such that the link capacities are not violated. The link cost is derived from the SINR and used to determine the paths. In order to take into consideration the mobility of the user nodes, we developed a simple mobility prediction model using the Hidden Markov Model (HMM) in which the predicted location of the user at time t is used to calculate the SINR value. We showed that our HMM prediction model provides good accuracy compared to conventional Markov based prediction models. We also showed that our SINR based routing approach guarantees minimum interference paths by increasing the packet delivery ratio and reducing latency and routing overhead compared to established SINR based routing approaches in the literature.

Chapter 7

Conclusions and Future Research

7.1 Conclusions

The success of achieving ubiquitous wireless connectivity in broadband access networks is contingent upon how resources are allocated to ensure that each user has service availability. With increasing number of users demanding multimedia services (i.e., video and voice data), the limited spectrum of wireless networks make resource allocation techniques indispensable. In addition to spectrum limitations, wireless networks are inhibited by other inherent characteristics. Specifically, wireless interference has been shown to be the most critical factors in hindering performance. Furthermore, the mobility of user nodes adds to the dynamic nature of wireless networks. Therefore, new and realistic paradigms for resource allocation considering the impact of interference and mobility are necessary to support high throughput and provide QoS guarantees. In this dissertation, we have proposed several novel and effective interference aware resource allocation approaches. Our results have showed that our proposed approaches are effective in providing performance improvements while mitigating the effects of interference. We have carried out performance comparisons to show the merits of our proposed approaches over their conventional counterparts established in the literature. The accomplishments of this dissertation can

be summarized as follows:

- In Chapter 3, we addressed the problem of interference aware rate adaptive subcarrier and power allocation using maximum multicommodity flow optimization. We proposed a novel method of solving the interference based MCF problem (MFCI). In addition, we developed a rate adaptive resource allocation algorithm that assigns subcarriers using spatial reuse and power to nodes considering the rate and time constraints of the users. We have showed that our novel approach to solve the MCFI routing algorithm appropriately discovers the least interfering paths while producing the maximum achievable throughput in comparison to other interference based routing protocols. In addition, we have showed that our subcarrier allocation technique performs better than that of assigning subcarriers with no spatial reuse. However, there is a tradeoff of too much reuse, which is detrimental to network performance. Furthermore, we have showed using a heuristic solution that our proposed rate adaptive joint subcarrier and power allocation algorithm garners better overall throughput than the two well-known joint resource allocation schemes. In addition, our proposed joint allocation algorithm considers the influence of interference on the system performance which is neglected by other schemes. We conclude that our approach, given the proper interference model and algorithmic measures, can mitigate the effects of wireless interference in dense wireless multihop networks thereby providing effective resource distribution. This work has been presented in [67] and [68].

- In Chapter 4 we have developed a decoupled approach to routing and scheduling optimization for relay enhanced wireless access networks which emphasizes physical interference constraints on capacity and spatial reuse of time slots to maximize throughput using multiple subcarriers. We used an interference based maximum concurrent flow method to route packets and we used spatial TDMA to schedule the packets. We modeled our scheduling framework so that the number of time slots

assigned to a link is proportional to the amount of traffic that traverses it. Since the problem is shown to be NP-hard, we proposed a modified column generation approach to provide a heuristic solution to the problem. We have showed that the decoupled routing and scheduling optimizations improve throughput and minimize time slots by mitigating interference and allowing time slots to be reused in a spatially effective manner while alleviating the need for joint interference constraints between routing and scheduling to determine schedulable flows. This work has been published in [69].

- In Chapter 5, we addressed the problem of fair resource allocation using multipath flow routing. We proposed a novel routing metric, RI^3M, by considering both inter-flow and intra-flow interference to enhance the selection of good quality paths. Using virtual network decomposition, we have showed that RI^3M is an isotonic routing metric that outperforms the most prominent and relevant routing metrics used in the literature in terms of end-to-end delay, packet loss and throughput. In addition, we have developed a max-min fair (MMF) bandwidth allocation algorithm for multipath flow routing in multihop wireless networks. To ensure QoS, our LMX:M^3F optimization formulation has been shown to provide better utilization of bandwidth resources in comparison to well respected MMF algorithms established in the literature particularly in terms of blocking ratio and link load. This work has been reported in [70], [71], and [72].

- To tackle the influence of mobility on interference aware resource allocation, in Chapter 6, we developed a distributed mobility prediction scheme using the Hidden Markov Model (HMM). Whereas in traditional cellular networks the base station controls mobility management, our scheme transfers this control to the individual relay nodes. The relays keep track of the users connected to it and their respective movements in a database. We then used the HMM mobility prediction engine

to calculate the resulting interference observed due to the mobility of the users. Using SINR and the HMM prediction scheme, we developed a SINR based routing algorithm that that determines the least interfering paths for each user to the base station. We solved this problem using the minimum cost optimization method. We showed that our HMM prediction model provides high prediction accuracy when compared to generic Markov based prediction approaches. In addition, we showed that our SINR based routing scheme outperforms existing SINR based routing protocols in the literature. This work is being prepared for submission [73].

7.2 Future Research

In this dissertation, we have studied key research issues related to interference aware resource allocation. Our results demonstrate the effectiveness of our proposed results and also reveal important future research directions to improve system performance. Despite its accomplishments, this dissertation is the first step to understand the impact of interference and mobility on resource management in broadband wireless access networks. Further research directions should address the following important issues:

- *Relay Node Placement*: Employing various relay nodes alleviates the problem of traffic congestion and single points of failure. In the presence of multiple gateways, traffic load can be balanced more effectively and efficiently, thereby facilitating traffic routing, packet scheduling, and QoS provisioning. With better traffic distribution, co-channel interference can be reduced to a greater extent. However, to achieve optimal interference reduction, the placement of the relay nodes has to be carefully determined. Interference aware algorithms for relay node placement need to be investigated to study the potential capacity gains that can be derived. Frequency reuse coupled with directional antennas can achieve interference mitigation.

- *Topology Control for Interference*: Related to the relay placement problem, mobility

plays a role in achieving interference limited performance. Topology control is a technique used mainly in wireless ad hoc and sensor networks in order to reduce the initial topology of the network to save energy and extend the lifetime of the network. The main goal is to reduce the number of active nodes and active links, preserving the saved resources for future maintenance. Much of the research in the literature deals with topology control for energy consumption. The natural question that arises is what are the best topologies from the radio interference point of view? Answering this question can be simplified if all the nodes use the same transmit power level, however, that is not a practical scenario. Thus, setting the transmitting range is critical for connectivity and reducing interference. The issue of determining interference-optimal topologies has not been addressed in the literature. This study would further enhance the deployment of broadband wireless access networks.

- *Handoff Management Exploiting SINR*: Handoff to base stations across a multi-cell network is an important aspect of mobility management. A handoff management architecture using the SINR of the present and neighboring base stations can improve service continuity. Maintaining this continuity is increasingly important for multimedia applications. Using a mobile user's speed, handoff signaling delay information can be maintained while enhancing the handoff performance. Specifically, integrating SINR into the handoff scheme, we can reduce false handoff initiations which create unnecessary traffic loads.

Bibliography

[1] R. Pabst et al., "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80–89, September 2004.

[2] D. Soldani and S. Dixit, "Wireless relays for broadband access," *IEEE Communications Magazine*, vol. 46, no. 3, pp. 58–66, March 2008.

[3] B.H. Walke, S. Mangold, and L. Berlemann, *IEEE 802 Wireless Systems: Protocols, Multi-Hop Mesh/Relaying, Performance and Spectrum Coexistence*, John Wiley and Sons Ltd, 2006.

[4] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1150–1158, November 2003.

[5] V. Genc, S. Murphy, Y. Yang, and J. Murphy, "IEEE 802.16j relay-based wireless access networks: An overview," *IEEE Wireless Communications*, vol. 15, no. 5, pp. 56–53, October 2008.

[6] S. W. Peters and R. W. Heath Jr., "The future of WiMAX: Multihop relaying with IEEE 802.16j," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 104–111, January 2009.

[7] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMAX and LTE-advanced mobile systems," *IEEE Communications Magazine*, vol. 47, no. 10, pp. 100–105, October 2009.

[8] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y. Kim, E. Kim, and Y. Cheong, "An overview of radio resource management in relay-enhanced OFDMA-based networks," *IEEE Communications Surveys and Tutorials*, vol. 1, no. 99, pp. 1–17, April 2010.

[9] T.S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, second edition, 2002.

[10] Y. Shi, Y.T. Hou, J. Liu, and S. Kompella, "How to correctly use the protocol interference model for multi-hop wireless networks," in *Proceedings of ACM MobiHoc*, 2009, pp. 239–248.

[11] A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2662–2671, May 2009.

[12] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multihop wireless network performance," in *Proceedings of ACM MobiCom*, 2003, pp. 66–80.

[13] M. Alicherry, R. Bhatia, and E. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 1960–1971, November 2006.

[14] K.N. Ramachandran, E.M. Belding, K.C. Almeroth, and M.M. Buddhikot, "Interference aware channel assignment in multiradio wireless mesh networks," in *Proceedings of IEEE INFOCOM*, 2006, pp. 1–12.

[15] J. Tang, G. Xue, C. Chandler, and W. Zhang, "Interference-aware routing in multihop wireless networks using directional antennas," in *Proceedings of IEEE INFOCOM*, 2005, pp. 751–760.

[16] G. Brar, D.M. Blough, and P. Santi, "Computationally efficient scheduling with the physical interference model for throughput improvement in wireless mesh networks," in *Proceedings of ACM MobiCom*, 2006, pp. 2–13.

[17] A.P. Subramanian, H. Gupta, and S.R. Das, "Minimum-interference channel assignment in multi-radio wireless mesh networks," in *Proceedings of IEEE SECON*, 2007, pp. 481–490.

[18] Y.H. Tam, R. Benkoczi, H.S. Hassanein, and S.G. Akl, "Channel assignment in multihop cellular networks: Minimum delay," *IEEE Transactions on Mobile Computing*, vol. 9, no. 7, pp. 1022–1034, July 2010.

[19] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, "Resource allocation for OFDMA based relay enhanced cellular networks," in *Proceedings of IEEE VTC-Spring*, 2007, pp. 3160–3164.

[20] H.T. Cheng and W. Zhuang, "An optimization framework for balancing throughput and fairness in wireless networks with QoS support," *IEEE Transactions on Wireless Communications*, vol. 7, no. 2, pp. 584–593, February 2008.

[21] L. Tassiulas and S. Sarkar, "Maxmin fair scheduling in wireless ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 163–173, January 2005.

[22] J. Maatta and T. Braysy, "A novel approach to fair routing in wireless mesh networks," *EURASIP Journal on Wireless Communications and Networking*, pp. 1–13, January 2009.

[23] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, August 2002.

[24] G. Li and H. Liu, "Resource allocation for OFDMA relay networks with fairness constraints," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2061–2069, November 2006.

[25] G. Kulkarni, S. Adlakha, and M. Srivastava, "Subcarrier allocation and bit loading algorithms for OFDMA-based wireless networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 6, pp. 652–662, November 2005.

[26] K. D. Lee and V. C. M. Leung, "Fair allocation of subcarrier and power in an OFDMA wireless mesh network," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2051–2060, November 2006.

[27] M. Sebastian, E. Weiss, and G. R. Hiertz, "Benefits and limitations of spatial reuse in wireless mesh networks," in *Proceedings of ACM MSWiM*, 2007, pp. 244–251.

[28] B. Alaweih, Z. Yongning, C. Assi, and H. Mouftah, "Improving spatial reuse in multihop wireless networks," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 3, pp. 71–91, 3rd Quarter 2009.

[29] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proceedings of IEEE VTC-Spring*, 2000, pp. 1085–1089.

[30] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171–178, February 2003.

[31] G. C. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless networks. Part I: Theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, February 2005.

[32] C. Mohanram and S. Bashyam, "Joint subcarrier and power allocation in channel-aware queue-aware scheduling for multiuser OFDM," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, pp. 3208–3213, September 2007.

[33] K. Inhyoung, I. S. Park, and Y. H. Lee, "Use of linear programming for dynamic subcarrier and bit allocation in multiuser OFDM," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1195–1207, July 2006.

[34] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2318–2328, September 2005.

[35] R. Kwak and J. M. Cioffi, "Resource-allocation for OFDMA multi-hop relaying downlink systems," in *Proceedings of IEEE Globecom*, 2007, pp. 26–30.

[36] M. Kaneko and P. Popovski, "Adaptive resource allocation in cellular OFDMA system with multiple relay stations," in *Proceedings of IEEE VTC-Spring*, 2007, pp. 3026–3030.

[37] O. Oyman, "OFDM2A: A centralized resource allocation policy for cellular multi-hop networks," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2006, pp. 656–660.

[38] E. Visotsky, J. Bae, R. Peterson, R. Berry, and M. L. Honig, "On the uplink capacity of an 802.16j system," in *Proceedings of IEEE WCNC*, 2008, pp. 2657–2662.

[39] T. Girici, C. Zhu, J. Agre, and A. Ephremides, "Optimal radio resource management in multihop relay networks," in *Proceedings of Symposium on Modelling and Optimization in Mobile Ad Hoc and Wireless Networks (WiOpt)*, 2008, pp. 443–451.

[40] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, "Resource scheduling for OFDMA/TDD based relay enhanced cellular networks," in *Proceedings of IEEE WCNC*, 2008, pp. 1544–1548.

[41] J. Cai, K.-H. Liu, X. Shen, J. W. Mark, and T. D. Todd, "Power allocation and scheduling for ultra-wideband wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1103–1112, March 2008.

[42] P. Wang, H. Jiang, W. Zhuang, and H.V. Poor, "Redefinition of max-min fairness in multi-hop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4786–4791, December 2008.

[43] X. Huang and Y. Fang, "Multiconstrained QoS multipath routing in wireless sensor networks," *Wireless Networks*, vol. 14, no. 4, pp. 465–478, August 2008.

[44] P. Thulasiraman, S. Ramasubramanian, and M. Krunz, "Disjoint multipath routing in dual homing networks using colored trees," in *Proceedings of IEEE GLOBECOM*, 2006, pp. 1–5.

[45] P. Thulasiraman, S. Ramasubramanian, and M. Krunz, "Disjoint multipath routing to two distinct drains in a multi-drain sensor network," in *Proceedings of IEEE INFOCOM*, 2007, pp. 643–651.

[46] A.P. Subramanian, M.M. Buddhikot, and S. Miller, "Interference aware routing in multi-radio wireless mesh networks," in *Proceedings of IEEE WiMesh*, 2006, pp. 55–63.

[47] R. Langar, N. Bouabdallah, and R. Boutaba, "Mobility-aware clustering algorithms with interference constraints," *Computer Networks*, vol. 53, no. 1, pp. 25–44, January 2009.

[48] K. Kar, M. Kodialam, and T.V. Lakshman, "Minimum interference routing of bandwidth guaranteed tunnels with MPLS traffic engineering applications," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 12, pp. 2566–2579, December 2000.

[49] B. Qi, S. Biaz, S. Wu, and Y. Ji, "An interference-aware routing metric in multi-radio multi-hop networks," in *Proceedings of the ACM 45th Annual Southeast Regional Conference (ACM-SE)*, 2007, pp. 549–500.

[50] J.-Y. Teo, Y. Ha, and C.-K. Tham, "Interference minimized multipath routing with congestion control in wireless sensor network with high rate streaming," *IEEE Transactions on Mobile Computing*, vol. 7, no. 9, pp. 1124–1137, September 2008.

[51] Y.-F. Wen and Y.-S. Lin, "Fair bandwidth allocation and end to end delay routing algorithms for wireless mesh networks," *IEICE Transactions on Communications*, vol. 90-B, no. 5, pp. 1042–1051, May 2007.

[52] Z. Xu, C. Huang, and Y. Cheng, "Interference-aware QoS routing in wireless mesh networks," in *Proceedings of 4th International Conference on Mobile Ad-Hoc and Sensor Networks*, 2008, pp. 95–98.

[53] L. Dekar and H. Kheddouci, "A cluster based mobility prediction scheme for ad hoc networks," *Ad Hoc Networks (Elsevier)*, vol. 6, no. 2, pp. 168–194, April 2008.

[54] W.-S. Soh and H.S. Kim, "QoS provisioning in cellular networks based on mobility prediction techniques," *IEEE Communications Magazine*, vol. 41, no. 1, pp. 86–92, January 2003.

[55] K. Wang and B. Li, "Group mobility and partition prediction in wireless ad-hoc networks," in *Proceedings of IEEE ICC*, 2002, pp. 1017–1021.

[56] W. Su, S.-J. Lee, and M. Gerla, "Mobility prediction and routing in ad hoc wireless networks," *International Journal of Network Management*, vol. 11, no. 1, pp. 3–30, January/February 2001.

[57] B. An and S. Papavassilou, "A mobility-based clustering approach to support mobility management and multicast routing in mobile ad hoc wireless networks," *International Journal of Network Management*, vol. 11, no. 6, pp. 387–395, November/December 2001.

[58] R. Chellappa Doss, A. Jennings, and N. Shenoy, "Mobility prediction for seamless mobility in wireless networks," in *Proceedings of the 4th International Network Conference*, 2004, pp. 435–443.

[59] G. Liu and G. Maguire, "A class of mobile motion prediction algorithms for wireless mobile computing and communication," *Mobile Networks and Applications*, vol. 1, no. 2, pp. 113–121, October 1996.

[60] S. Kwon, H. Park, and K. Lee, *A Novel Mobility Prediction Algorithm Based on User Movement History in Wireless Networks*, Lecture Notes in Computer Science. Springer Berlin/Heidelberg, 2005.

[61] S. Michaelis and C. Wietfeld, "Comparision of user mobility pattern prediction algorithms to increase handover trigger accuracy," in *Proceedings of IEEE VTC-Spring*, 2006, pp. 952–956.

[62] L. Song, U. Deshpande, U.C. Kozat, D. Katz, and R. Jain, "Predictability of wlan mobility and its effect on bandwidth provisioning," in *Proceedings of IEEE INFOCOM*, 2006, pp. 1–13.

[63] M.H. Sun and D.M Blough, "Mobility prediction using future knowledge," in *Proceedings of ACM MSWiM*, 2007, pp. 235–239.

[64] W. Ma, Y. Fang, and P. Lin, "Mobility management strategy based on user mobility patterns in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 322–330, January 2007.

[65] P.S. Prasad and P. Agrawal, "Movement prediction in wireless networks using mobility traces," in *Proceedings of IEEE CCNC*, 2010, pp. 1–5.

[66] H. Abu-Ghazaleh and A. S. Alfa, "Application of mobility prediction in wireless networks using markov renewal theory," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, pp. 788–802, February 2010.

[67] P. Thulasiraman and X. Shen, "Interference aware subcarrier assignment for throughput maximization in OFDMA wireless relay mesh networks," in *Proceedings of IEEE ICC*, 2009, pp. 1–6.

[68] P. Thulasiraman and X. Shen, "Interference aware resource allocation in hybrid hierarchical wireless networks," Computer Networks (Elsevier) 2010, to appear, doi:10.1016/j.comnet.2010.03.012.

[69] P. Thulasiraman and X. Shen, "Decoupled optimization of interference aware routing and scheduling for throughput maximization in wireless relay mesh networks," in *Proceedings of IEEE WiMesh (colocated with IEEE SECON)*, 2009, pp. 1–6.

[70] P. Thulasiraman, J. Chen, and X. Shen, "Max-min fair multipath routing with physical interference constraints for multihop wireless networks," in *Proceedings of IEEE ICC*, May 2010.

[71] P. Thulasiraman and X. Shen, "Disjoint multipath routing and QoS provisioning under physical interference constraints," in *Proceedings of IEEE WCNC*, April 2010.

[72] P. Thulasiraman, J. Chen, and X. Shen, "Multipath routing and max-min fair QoS provisioning under interference constraints in wireless multihop networks," IEEE Transactions on Parallel and Distributed Systems 2010, to appear.

[73] P. Thulasiraman, J. Chen, and X. Shen, "SINR based routing for relay enhanced networks using distributed mobility prediction," in preparation for journal submission.

[74] W. H. Park and S. Bahk, "Resource management policies for fixed relays in cellular networks," *Computer Communications (Elsevier)*, vol. 32, no. 34, pp. 703–711, March 2009.

[75] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," *Computer Networks (Elsevier)*, vol. 47, no. 4, pp. 445–487, March 2005.

[76] J. Shin, R. Kumarand, Y. Shing, and T. F. La Porta, "Multi-hop wireless relay networks of mesh clients," in *Proceedings of IEEE WCNC*, 2008, pp. 2717–2722.

[77] M. Mauve, J. Widmer, and H. Hartenstein, "A survey on position based routing in mobile ad hoc networks," *IEEE Networks*, vol. 15, no. 6, pp. 30–39, November 2001.

[78] F. Shahrokhi and D. W. Matula, "The maximum concurrent flow problem," *Journal of the ACM*, vol. 37, no. 2, pp. 318–334, April 1990.

[79] N. Garg and J. Konemann, "Faster and simpler algorithms for multicommodity flow and other fractional packing problems," *SIAM Journal on Computing*, vol. 37, no. 2, pp. 630–652, 2007.

[80] L.K. Fleischer, "Approximating fractional multicommodity flow independent of the number of commodities," *SIAM Journal of Discrete Mathematics*, vol. 13, no. 4, pp. 505–520, October 2000.

[81] R.M. Kortebi, Y. Gourhant, and N. Agoulmine, "On the use of SINR for interference-aware routing in wireless multi-hop networks," in *Proceedings of ACM MSWiM*, 2007, pp. 395–399.

[82] A. Capone and F. Martignon, "A multi-commodity flow model for optimal routing in wireless mesh networks," *Journal of Networks*, vol. 2, no. 3, pp. 1–5, June 2007.

[83] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional fairness," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726–2737, November 2005.

[84] M. Takai, J. Martin, and R. Bragodia, "Effects of wireless physical layer modeling in mobile ad hoc networks," in *Proceedings of ACM MobiHoc*, 2001, pp. 87–94.

[85] V. Friderikos and K. Papadaki, "Interference aware routing for minimum frame length schedules in wireless mesh networks," *EURASIP Journal on Wireless Communications and Networking*, pp. 1–13, August 2008.

[86] R. Nelson and L. Kleinrock, "Spatial TDMA: A collision-free multihop channel access protocol," *IEEE Transactions on Communications*, vol. 33, no. 9, pp. 934–944, September 1985.

[87] P. Bjorklund, P. Varbrand, and D. Yuan, "Resource optimization of spatial TDMA in ad hoc radio networks: A column generation approach," in *IEEE Proceedings of INFOCOM*, 2003, pp. 818–824.

[88] B. Hajek and S. Sasaki, "Link scheduling in polynomial time," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 910–917, September 1988.

[89] C. G. Prohazka, "Decoupling link scheduling constraints in multi-hop packet radio networks," *IEEE Transactions on Computers*, vol. 38, no. 3, pp. 455–458, March 1989.

[90] J. Gronkvist and A. Hansson, "Comparison between graph-based and interference-based stdma scheduling," in *ACM Proceedings of MobiHoc*, 2001, pp. 255–258.

[91] Y. Wang, W. Wang, X.-Y Li, and W.-Z Song, "Interference-aware joint routing and TDMA link scheduling for static wireless networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 12, pp. 1709–1725, 2008.

[92] M. Alicherry, R. Bhatia, and E. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 1960–1971, 2006.

[93] M.E.Lubbecke, "Selected topics in column generation," Tech. Rep., Braunschweig University of Technology, 2004, http://citeseer.ist.psu.edu/558048.html.

[94] D. Bertsekas and R. Gallagher, *Data Networks*, Prentice-Hall, 1992.

[95] J. Kleinberg, Y. Rabani, and E. Tardos, "Fairness in routing and load balancing," *Journal of Computer and System Sciences*, vol. 63, no. 1, pp. 568–578, August 2001.

[96] M. Allalouf and Y. Shavitt, "Centralized and distributed algorithms for routing and weighted max-min fair bandwidth allocation," *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, pp. 1015–1024, October 2008.

[97] D. Nace, L. Doan, O. Klopfenstein, and A. Bashllari, "Max-min fairness in multi-commodity flows," *Computers and Operations Research*, vol. 35, no. 2, pp. 557–573, February 2008.

[98] D. Nace, N. Doan, E. Gourdin, and B. Liau, "Computing optimal max-min fair resource allocation for elastic flows," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1272–1281, December 2006.

[99] D. Nace, "A linear programming based approach for computing optimal fair splittable routing," in *Proceedings of IEEE ISCC*, 2002, pp. 468–474.

[100] R. Pal, "A lexicographically optimal load balanced routing scheme for wireless mesh networks," in *Proceedings of IEEE ICC*, 2008, pp. 2393–2397.

[101] X. Huang, S. Feng, and H. Zhuang, "Cross-layer fair resource allocation for multi-radio multi-channel wireless mesh networks," in *Proceedings of 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCom)*, 2009, pp. 2639–2643.

[102] V. S. Mansouri, A. H. Mohsenian-Rad, and V. W. S. Wong, "Lexicographically optimal routing for wireless sensor networks with multiple sinks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 3, pp. 1490–1500, March 2009.

[103] J. Tang, R. Hincapie, G. Xue, W. Zhang, and R. Bustamante, "Fair bandwidth allocation in wireless mesh networks with cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1487–1496, March 2010.

[104] J. Chou and B. Lin, "Optimal multi-path routing and bandwidth allocation under utility max-min fairness," in *Proceedings of IEEE IWQoS*, 2009, pp. 1–9.

[105] Y.Yang, J. Wang, and R.Kravets, "Designing routing metrics for mesh networks," in *Proceedings of IEEE WiMesh*, 2005.

[106] D.S.J. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Proceedings of ACM MobiCom*, 2003, pp. 134–146.

[107] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proceedings of ACM MobiCom*, 2004, pp. 114–128.

[108] A. Bagchi, A. Chaudhary, and P. Kolman, "Short length menger's theorem and reliable optical routing," *Theoretical Computer Science*, vol. 339, no. 2, pp. 315–332, June 2005.

[109] J. Zhou and K. Mitchell, "A scalable delay based analytical framework for CSMA/CA wireless mesh networks," *Computer Networks*, vol. 54, no. 2, pp. 304–318, February 2010.

[110] G. Athanasiou, T. Korakis, O. Ercetin, and L. Tassiulas, "Cross-layer framework for association control in wireless mesh networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 1, pp. 65–80, January 2009.

[111] Y. Yang, J. Wang, and R. Kravets, "Interference-aware load balancing for multihop wireless networks," Tech. Rep. UIUCDCS-R-2005-2526, University of Illinois at Urbana-Champaign, 2005.

[112] R. Bhandari, *Survivable Networks: Algorithms for Diverse Routing*, Kluwer Academic Publishers, 1999.

[113] D. Nace and M. Pioro, "Max-min fairness and its applications to routing and load balancing in communication networks: A tutorial," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 5–17, January 2008.

[114] C.-C. Hsu, P. Liu, D. Wang, and J.-J. Wu, "Generalized edge coloring for channel assignment in wireless networks," in *Proceedings of the International Conference on Parallel Processing (ICPP)*, 2006, pp. 82–92.

[115] D. Willkomm, S. Machiraju, J. Blot, and A. Wolisz, "Primary user behavior in cellular networks and implications for dynamic spectrum access," *IEEE Communications Magazine*, vol. 47, no. 3, pp. 88–95, March 2009.

[116] P.S. Prasad and P. Agrawal, "A generic framework for mobility prediction and resource utilization in wireless networks," in *IEEE Proceedings of COMSNETS*, 2010, pp. 1–10.

[117] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, May 2002.

[118] W. Navidi and T. Camp, "Stationary distributions for the random waypoint mobility model," *IEEE Transactions on Mobile Computing*, vol. 3, no. 1, pp. 99–108, January/February 2004.

[119] C. Bettstetter and C. Wagner, "The spatial node distribution of the random waypoint model," in *Proceedings of German Workshop on Mobile Ad Hoc Networks (WMAN)*, 2002, pp. 41–58.

[120] C. Bettstetter, "Mobility modeling in wireless networks: Categorization, smooth movement, and border effects," *ACM Mobile Computing and Communications Review*, vol. 5, no. 3, pp. 55–67, March 2001.

[121] P.S. Prasad and P. Agrawal, "Mobility prediction for wireless network resource management," in *Proceedings of IEEE SSST*, 2009, pp. 98–102.

[122] W. Cui and X. Shen, "User movement tendency prediction and call admission control for cellular networks," in *Proceedings of IEEE ICC*, 2000, pp. 670–674.

[123] W.-S. Soh and H.S. Kim, "Dynamic bandwidth reservation in cellular networks using road topology based mobility prediction," in *Proceedings of IEEE INFOCOM*, 2004, pp. 2766–2777.

[124] H. Si, Y. Wang, J. Yuan, and X. Shan, "Mobility prediction in cellular network using hidden markov model," in *Proceedings of IEEE CCNC*, 2010, pp. 1–5.

[125] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 77–90, February 2008.

[126] L. Badia, N. Bui, M. Miozzo, M. Rossi, and M. Zorzi, "Mobility-aided routing in multi-hop heterogeneous networks with group mobility," in *Proceedings of IEEE GLOBECOM*, 2007, pp. 4915–4919.

[127] W. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *Proceedings of IEEE INFOCOM*, 2007, pp. 758–766.

[128] R.M. Kortebi, D.-E. Meddour, Y. Gourhant, and N. Agoulmine, "SINR-based routing in multi-hop wireless networks to improve VOIP applications support," in *Proceedings of IEEE CCNC*, 2007, pp. 491–496.

[129] S. Kwon and N.B. Schroff, "Energy-efficient SINR-based routing for multihop wireless networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 5, pp. 668–681, May 2009.

[130] S.-Z. Yu and H. Kobayashi, "Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1947–1951, May 2006.

[131] W. Turin, *Digital Transmission Systems*, McGraw Hill, 1998.

[132] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows*, Prentice Hall, 1993.

[133] J. Orlin, "A faster strongly polynomial minimum cost flow algorithm," *Operations Research*, vol. 41, no. 2, pp. 338–350, 1993.