Throughput and Coverage Improvement in Wireless Mesh Networks

by

Amit Kumar Saha

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Dr. David B. Johnson, Chair
Associate Professor of Computer Science

Dr. Edward W. Knightly
Professor of Electrical and Computer Engineering

Dr. T.S. Eugene Ng
Assistant Professor of Computer Science

Houston, Texas

February, 2007
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
Throughput and Coverage Improvement in Wireless Mesh Networks

Amit Kumar Saha

Abstract

In a wireless mesh network, nodes known as transit access points (TAPs) cooperatively forward traffic from users that may be multiple wireless hops apart. A limited number of TAPs also have a connection directly to the Internet, serving as gateway nodes that provide Internet connectivity to the entire mesh network. Wireless mesh networks are gaining in importance as an alternative to cable and DSL and are envisioned to provide fixed, nomadic, portable, and—eventually mobile—wireless broadband connectivity. In this thesis, I provide solutions to two important problems in wireless mesh networks and evaluate these solutions through simulation experiments. The two solutions, improving throughput and coverage, can simultaneously coexist and can complement each other in a single mesh network. Nevertheless, both the techniques can also exist independent of each other in a wireless mesh network and can individually prove advantageous.

First, I present two novel traffic-aware routing metrics that take into account existing user traffic flows in the network. Previous routing metrics have been
traffic-unaware, often causing routes with poor throughput to be selected when other better routes are available. These new traffic-aware metrics use information captured through measurements at the medium access control (MAC) layer, which is then exposed to the routing layer. I compare these traffic-aware metrics with existing traffic-unaware metrics under different network scenarios.

Second, I present the design and analysis of a new technique for increasing the coverage of a wireless mesh network through deployment of small, low-cost booster TAPs (bTAPs). These bTAPs are strategically deployed and controlled by the system operator to wirelessly forward traffic between users and TAP nodes. This deployment model is especially suitable for wide area wireless access networks that use centralized management of radio resources. I analyze the use of bTAPs across different frequency reuse patterns typical of those used in multicell wireless environments for efficient management of costly radio spectrum. The bTAP architecture provides dramatic improvements in outage performance and a sufficient capacity gain to compensate for the radio resources required for forwarding user traffic via bTAPs.
Acknowledgments

The completion of this thesis comes with a great sense of achievement and personal triumph. I am grateful to several people without whose generous help and support I would not have been able to complete my thesis. Though I am bound to forget some, I would like to make a humble attempt to thank a few people.

First and foremost, I thank my adviser Dr. David B. Johnson for his guidance throughout the process of this thesis. The innumerable discussions we had, on and off topic, was the single most instrumental reason that has brought this thesis to its present form.

I also thank the other members of my thesis committee: Dr. Edward W. Knightly and Dr. T.S. Eugene Ng for their reviews and suggestions, which greatly improved the quality of this thesis.

It was a pleasure and an honor to work with my colleagues, Theodoros Salounidis and Michele Garetto. The theoretical model used in Section 3.4.2 of this thesis, and the corresponding routing metric were a result of several months of fruitful collaboration with them. It was also a pleasure to work with “J” Kim and Shankar, researchers at AT&T Labs, NJ, USA. My research with them formed an integral part of this thesis.
I am grateful to the other members of the Rice University Monarch Project group. Santashil was a constant companion for around five years, right from the start of my graduate study. I had memorable experience of working with Khoa Anh To, Shu Du, and Yanjun Sun. I would also like to thank earlier members of the group, Jorjeta Jetcheva and Yih-Chun Hu, who helped me during the beginning of my graduate career at Rice University. I thank all of my groupmates and wish them the very best in life.

I would not have been able to complete this thesis had it not been for the constant love and caring of my wife, Deepanwita Bagchi. Throughout my Ph.D. process she has been a steadfast companion and friend when I was making good progress in my research, and, more importantly, when I felt I was going nowhere. She somehow was easily able to cope with my bouts of frustration and social hibernation. I am lucky to have her by my side and cannot thank her enough.

I would also like to thank my parents whose constant support and belief in me allowed me to explore higher studies. I thank my brother for his support all throughout my life. I owe a lot to a host of relatives whose support for me and my family has allowed me to dedicatedly pursue higher education.

Perhaps the most important aspect of my graduate study is the new friends that I made during my stay at Rice University. The time that I spent with the Indian Students at Rice (ISAR), Friends of Young Minds (FoYM), The Rice Badminton Club, The Rice Cricket Club, and The Rice Squash Club allowed me to easily tackle the
frustrations inherent in graduate study. I thank the well wishes of all my friends, old and new, and greatly acknowledge my inability to be effective without their support. I would like to especially thank Dhruv, ACP, Rajnish, Pradeep Rai, Vinod, Priyank, Sushant, and Pradeep Bhat for all the time that we have spent together.

Special thanks goes to the staff at the Office of International Students and Scholars for the continuous and personalized support with which they pamper international students. I also thank the wonderful staff of the Department of Computer Science at Rice.

My work was supported in part by NSF under grants CNS-0520280, CNS-0435425, CNS-0338856, CNS-0325971, and CNS-0209204; by NASA under grant NAG3-2534; and by a gift from Schlumberger. The views and conclusions contained here are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of NSF, NASA, Schlumberger, Rice University, or the U.S. Government or any of its agencies.
Contents

Abstract ii
Acknowledgments iv
List of Illustrations xi
List of Tables xvi

1 Introduction 1

1.1 Wireless Connectivity 1
1.2 Wireless Mesh Networks 4
1.3 Existing Status of Wireless Mesh Networks 7
1.4 Thesis Contributions 10
1.5 Thesis Organization 12

2 Overview of Complete System 13

2.1 Frequency Reuse 14
2.2 End-to-end Communication 17
   2.2.1 Communication with a TAP 17
   2.2.2 Communication with a Gateway TAP 18
2.3 Scheduling End-to-End Communication 19
2.4 Chapter Summary ................................................. 22

3 Throughput Improvement in Wireless Mesh Networks 24

3.1 Related Work .................................................. 27
  3.1.1 Routing for Best-Effort Traffic ......................... 27
  3.1.2 Admission Control for QoS Traffic .................... 34

3.2 Design Basics ................................................. 40

3.3 MAC Layer Share (MLS) .................................... 43
  3.3.1 Per Link Throughput .................................. 46
  3.3.2 End-to-End Throughput Along a Route ................ 47
  3.3.3 Queuing Theory Model of a Clique .................... 50

3.4 Metric Based on Analytical Model for IEEE 802.11 .... 53
  3.4.1 Review of Modeling Framework ....................... 54
  3.4.2 Extension to Multiple Receivers ...................... 59
  3.4.3 End-to-End Throughput Along a Route ................ 63

3.5 The Routing Protocol ....................................... 64

3.6 Evaluation ................................................... 66
  3.6.1 Simulation Setup ..................................... 66
  3.6.2 Flow Throttling ....................................... 68
  3.6.3 Evaluation Topologies ................................ 70
  3.6.4 A Single Additional Flow — Chaska Topology ......... 72
  3.6.5 Directional Antenna Model ............................ 78
3.6.6 Results with Directional Antennas — Chaska Topology . . . . 79
3.6.7 Overall Results for the Chaska Topology . . . . . . . . . . . . 80
3.6.8 A Single Additional Flow — Manhattan Topology . . . . . . . 85
3.6.9 Results with Directional Antennas — Manhattan Topology . . 96
3.6.10 Overall Results for the Manhattan Topology . . . . . . . . . . 99
3.6.11 Multiple Additional Flows — Manhattan Topology . . . . . . 100
3.6.12 Shortlived Flows — Manhattan Topology . . . . . . . . . . . 107
3.6.13 Multiple Simultaneous Flows — Manhattan Topology . . . . 109

3.7 A Comparison of the Two Topologies . . . . . . . . . . . . . . . . . 111
3.8 Chapter Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117

4 Coverage Improvement in Wireless Mesh Networks . . . . . . . . . 122

4.1 Coverage Issues . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 123
4.2 Variations of bTAP Deployment Model . . . . . . . . . . . . . . . . 126
4.3 Related Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
4.4 Design . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 133
  4.4.1 Design Assumptions and Goals . . . . . . . . . . . . . . . . . . . 133
  4.4.2 System Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 135
  4.4.3 Limitation of Single Sector Analysis . . . . . . . . . . . . . . . . 139
  4.4.4 Scheduling in the Presence of bTAPs . . . . . . . . . . . . . . . . 142
  4.4.5 Effective Data Rate . . . . . . . . . . . . . . . . . . . . . . . . . . . . 146
4.5 Evaluation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 147
4.5.1 Evaluation Parameters ........................................... 147
4.5.2 Erceg-Greenstein Channel Model for Large-Scale Fading ... 149
4.5.3 Directional Antenna Model ..................................... 151
4.5.4 Frequency Reuse Pattern ...................................... 152
4.5.5 Achievable Transmission Rates without bTAPs ............... 155
4.5.6 Determining the Scheduling Regions ......................... 156
4.5.7 Simultaneous and Dedicated Scheduling Regions ............... 158
4.5.8 Transmission Rate from the Perspective of Clients ............. 160
4.5.9 Effects of Pathloss Exponent and Lognormal Fading .......... 162
4.5.10 Coverage Improvement ...................................... 165
4.5.11 Throughput Improvement ................................... 167
4.5.12 A Different Frequency Reuse Pattern — (1,3,6) ............. 168
4.5.13 Stability Analysis ........................................... 171
4.6 Chapter Summary ............................................... 178

5 Conclusions and Future Work ................................. 180
5.1 Throughput Improvement .................................... 181
5.2 Coverage Improvement ...................................... 184
5.3 Future Work ................................................ 185

Bibliography ...................................................... 188
Illustrations

1.1 A typical wireless mesh network ........................................ 5

2.1 An idealized six sector cell, each sector using an orthogonal channel. 14
2.2 The (1,6,6) frequency reuse pattern. ................................. 15
2.3 End-to-end communication from a client node to a gateway TAP. 16
2.4 A possible scheduling map for the entire wireless mesh system. 19
2.5 Overlapping periods during scheduling. ............................. 21

3.1 Limitation of the minimum hop count metric. ...................... 25
3.2 Just taking modulation rate into account is not enough. .......... 26
3.3 Intra-flow contention graph. .............................................. 49
3.4 Queuing theory model of clique. ......................................... 50
3.5 Example evolution of channel state with time, perceived by a node. 55
3.6 Topology of the Chaska wireless mesh network. .................. 73
3.7 Results for experiments on the Chaska topology. The maximum
load on any gateway is 1.5 Mbps, and each node uses a single
omni-directional antenna. ........................................ 75

3.8 Results for experiments on the Chaska topology. The maximum
load on any gateway is 2.0 Mbps, and each node uses a single
omni-directional antenna. ........................................ 77

3.9 Directional antenna pattern. ........................................ 79

3.10 Results for experiments on the Chaska topology. The maximum
load on any gateway is 1.5 Mbps, and each node uses a single
directional antenna. ........................................ 81

3.11 Results for experiments on the Chaska topology. The maximum
load on any gateway is 2.0 Mbps, and each node uses a single
directional antenna. ........................................ 82

3.12 Average throughput achieved in experiments on the Chaska topology. 83

3.13 Topology of the Manhattan wireless mesh network. ............................ 87

3.14 Results for experiments on the Manhattan topology. The maximum
load on any gateway is 2.0 Mbps, and each node uses a single
omni-directional antenna. ........................................ 90

3.15 Results for experiments on the Manhattan topology. The maximum
load on any gateway is 3.0 Mbps, and each node uses a single
omni-directional antenna. ........................................ 91
3.16 Results for experiments on the Manhattan topology. The maximum load on any gateway is 4.0 Mbps, and each node uses a single omni-directional antenna. ................................. 92

3.17 Results for experiments on the Manhattan topology. The maximum load on any gateway is 2.0 Mbps, and each node uses a single directional antenna. ................................. 93

3.18 Results for experiments on the Manhattan topology. The maximum load on any gateway is 3.0 Mbps, and each node uses a single directional antenna. ................................. 94

3.19 Results for experiments on the Manhattan topology. The maximum load on any gateway is 4.0 Mbps, and each node uses a single directional antenna. ................................. 95

3.20 Average throughput achieved in experiments on the Manhattan topology. ........................................ 98

3.21 Results for experiments with multiple simultaneous flows in the Manhattan topology. Each node uses a single omni-directional antenna. ........................................ 102

3.22 Results for experiments with multiple simultaneous flows in the Manhattan topology. Each node uses a single directional antenna. ... 104

3.23 Average cumulative throughput achieved in experiments on the Manhattan topology. ........................................ 106
3.24 Performance with multiple shortlived flows. .................. 108
3.25 Performance with multiple simultaneous flows. ................. 110
3.26 Neighbors and interferers CDF for both topologies. ............ 112
3.27 Neighbors and interferers statistics for the two topologies. .... 113
3.28 Neighbors and interferers contour plots for the Chaska topology. 115
3.29 Neighbors and interferers contour plots for the Manhattan topology. 116
3.30 Gateways accessible at different hop counts. .................... 117

4.1 Low signal strength at periphery of cell. ......................... 123
4.2 Proposed addition of bTAP to augment coverage of TAP. ......... 124
4.3 An idealized six sector cell, each using an orthogonal channel. .... 135
4.4 Different bTAP placement possibilities. .......................... 136
4.5 A unit composed of two adjacent sectors. ......................... 140
4.6 A possible downlink scheduling map. ............................. 142
4.7 Directional antenna pattern ...................................... 151
4.8 Multicell scenario without bTAP for (1,6,6) frequency reuse pattern. 152
4.9 Multicell scenario with bTAPs for (1,6,6) frequency reuse pattern. 153
4.10 Scatter plot for conventional downlink case ..................... 155
4.11 Downlink scatter plot with bTAPs for (1,6,6) frequency reuse pattern without lognormal fading .......................... 158
4.12 Downlink scatter plot with bTAPs for (1,6,6) frequency reuse pattern with lognormal fading ........................................ 159
4.13 Uplink scatter plot with bTAPs for (1,6,6) frequency reuse pattern with lognormal fading ........................................ 160
4.14 CDF of data rate from client perspective ........................................ 161
4.15 Effect of pathloss exponent and lognormal fading on downlink QPSK 1/2 outage ........................................ 163
4.16 Effect of pathloss exponent and lognormal fading on downlink sector throughput ........................................ 164
4.17 QPSK 1/2 outage for (1,6,6). ........................................ 165
4.18 Sector throughput for (1,6,6). ........................................ 166
4.19 The (1,3,6) frequency reuse pattern ........................................ 168
4.20 Analyzed multicell scenario without bTAP for (1,3,6) frequency reuse pattern ........................................ 169
4.21 Analyzed multicell scenario with bTAPs for (1,3,6) frequency reuse pattern ........................................ 170
4.22 QPSK 1/2 outage for (1,3,6) ........................................ 171
4.23 Sector throughput for (1,3,6) ........................................ 172
4.24 The effect of transmission power on downlink performance ........................................ 173
4.25 The effect of transmission power on uplink performance ........................................ 175
4.26 The effect of varying bTAP height ........................................ 177
4.1 Modulation levels used in evaluation.       147
4.2 Parameters used in the simulation experiments 153
Chapter 1

Introduction

For over a decade, the definitive agent of change in the world has been electronic connectivity. The ability to electronically connect to remote people and their places has fundamentally changed the way we conceive of the world and live in it. Apart from the widespread societal changes brought about by the Internet, connectivity has, among other things, given rise to newer business models (e.g., outsourcing, online advertisement, and e-commerce), faster and more transparent methods of governance (e.g., e-governance), far reaching education and knowledge dissipation (e.g., remote education, and Wikipedia), better health services (e.g., telemedicine), better monitoring ability (e.g., sensor networks, and RFID), and energy and time efficiency (e.g., home offices, e-ticketing, and online banking). In particular, wireless connectivity has substantially accelerated this change in the later half of the past decade.

1.1 Wireless Connectivity

When I started my graduate school at Rice University in the Fall of 2000, cell-phones, the best example of wireless technology affecting our lives, had only just
arrived in my home country of India. In around five years, cellphones have become a constant companion of people belonging to almost all strata of Indian society. The number of wireless phone connections, which are increasing in India at the rate of around 3 million subscribers per month, has long surpassed, the number of wired phone connections; this ranking is true in almost all markets of the world. Ubiquitous in the developed world, WiFi networks have only increased the pool of wirelessly connected users.

However, such networks are still heavily dependent on wired backhaul network infrastructures, thus presenting a hurdle in their deployment in many places where the cost of laying down wires is prohibitively high. For example, in developed cities, the cost of laying down underground wires is extremely high, as is the inconvenience caused to traffic and the smooth functioning of day-to-day life. Even in rural areas, where it might not be too inconvenient to lay down wires, the cost of laying down wires to a remote place is extremely high.

On the other hand, mobile ad hoc networks are completely distributed and entirely independent of any infrastructure and are envisioned to be used primarily in construction sites, disaster relief, military environments, and network edges. However, among other things, a lack of service guarantees and compelling applications has prevented the wide deployment of the mobile ad hoc network architecture in its most basic form. Mobile ad hoc networks are the most general case of wireless networks and hence the most difficult to manage. Nevertheless, research
on such networks has affected several newer architectures such as wireless mesh networks, WiFi networks, and sensor networks; these newer architecture have all used some form of stationary infrastructure to tame down the problems faced in mobile ad hoc networks.

Wireless technology holds the potential for simplifying the traditional problems of wired networks, such as deployment and maintenance. For example, DSL subscribers must be within a limited physical distance of the central distribution office in order to get satisfactory service. Infrastructure-based wireless networks such as the common access point-based IEEE 802.11\(^1\) networks [40], seen in university, industry, and municipal [73, 11] environments, have proven to be very successful in providing end-user connectivity. However, current IEEE 802.11-based infrastructure networks generally have each access point connected to a wired backbone connected to the Internet. As explained earlier, this increases the cost of deploying and managing the network. *Wireless mesh networks* alleviate this situation by not requiring all access points\(^2\) to have a wired connection.

---

\(^1\)IEEE 802.11 technology is often referred to as WiFi, based on the WiFi Alliance (http://www.wi-fi.org), which is an industry alliance with an objective to enable and certify the interoperability of products based on IEEE 802.11.

\(^2\)Unless otherwise mentioned, the terms “access point (AP)”, “Transit Access Point” (TAP), and “base station” (BS) are used interchangeably.
1.2 Wireless Mesh Networks

Current infrastructure-based wireless architectures, such as the IEEE 802.11/WiFi architecture, suffer from the problem that each access point has an additional wired connection to the wired backhaul. Normally, the transmission cost for the wired backhaul is significant in high capacity radio networks. Moreover, deploying such networks faces the hurdle of providing wired connectivity to each and every access point. This issue is especially acute in the early stages of network deployment, where a large number of wired access points are needed for sufficient coverage to start commercial services, while the revenue in the early stages is too low to offset the operational costs. These costs are relatively insensitive to the cost reduction of electronic components of a network.

The term “wireless mesh network” is often used interchangeably with ad hoc or multihop wireless network, although all the terms have slightly different nuances. A wireless mesh network sometimes implies multiple routes between any two nodes. An ad hoc network suggests a temporary and self-configuring nature of network formation, often composed of mobile nodes. A multihop network points to multihop forwarding implied in both mesh and ad hoc networks. However, for the purpose of this thesis,

*a wireless mesh network consists of nodes known as transit access points (TAPs) cooperatively forwarding traffic from users that may be multiple wireless hops apart. A limited number of TAPs have a connection directly*
Figure 1.1: A typical wireless mesh network

to the Internet, serving as gateway nodes that provide Internet connectivity
to the entire mesh network.

Wireless mesh networks, a schematic example of which is shown in Figure 1.1, attempt to change the economics of wireless networking by aggregating traffic for Internet backhaul. As such, wireless mesh networking has garnered widespread interest in its applications to military, sensor, community, and public safety networks [6, 15]. Wireless mesh networks [41, 58, 8, 60, 70] are gaining in importance as an alternative to cable and DSL and are envisioned to provide fixed, nomadic, portable, and — eventually mobile — wireless broadband connectivity [80]. For example, the TAPs project [70] at Rice University, Houston, Texas, USA states the following as its vision:

“Our driving vision is to provide a high-performance, scalable and widely deployed wireless Internet that facilitates services ranging from radically new and unforeseen applications to true wireless broadband to residences and public spaces at rates of 10s of Mb/sec.”
A survey of wireless mesh networks have been provided by Akyildiz et al. [6] and by Bruno et al. [15].

Initially, there may not exist enough users in an ad hoc network to provide sufficient coverage and connectivity to outside resources such as in the Internet, thereby limiting the incentive to join the network to improve connectivity for other users. Thus such a deployment require a large number of gateway access points to provide Internet connectivity to a geographical area. In contrast, a wireless mesh network does not suffer from this initial seeding problem. This is one of the reasons why the majority of commercial access network services, which need a certain level of availability assurance, are better supported through managed infrastructure, instead of ad hoc mesh networks.

Moreover, wireless mesh networking offers a simpler security model since the access points are centrally operated. Additionally, central control of access points leads to predictable network topology and route dynamics. In contrast to ad hoc networks, access points in wireless mesh networks are powered, thus allowing for higher power transmissions. Finally, since the access points are deployed and controlled by the network operators, a wireless mesh system exhibits better manageability than do ad hoc networks.

Due to these practical advantages offered by a wireless mesh network, operators of most operational Wi-Fi mesh systems providing access services [63, 76, 57,
12] deploy and control the mesh access points and user Wi-Fi stations, i.e., users, do not participate in traffic forwarding.

In order to make wireless networks practically useful, there has been substantial amount of work in the field of wireless mesh networking, both in academia as well as in industry. The following section provides a brief overview.

1.3 Existing Status of Wireless Mesh Networks

Currently, the research and the commercial implementation of wireless mesh networks are predominantly Wi-Fi oriented, naturally using unlicensed spectrum. The proliferation of wireless mesh networks in the Internet access context has recently started in earnest [63]³, mainly encouraged by the promise of the economic advantages of unlicensed spectrum and mass-produced Wi-Fi technology. This is in contrast to conventional wide-area cellular technologies such as WCDMA [2], HSPA [4, 3], and CDMA2000 EV-DO [1] that employ radio interfaces designed for long ranges and high utilization necessary due to costly deployment and licensed spectrum. Theoretical and simulation results for wireless mesh networking in various scenarios have been shown previously [6, 15, 31, 34, 75, 19, 66]. These studies illustrate the potential of higher system capacity under certain assumptions as well as various tradeoffs with respect to factors such as latency, power consumption, capacity, per-user throughput, and multihop diversity.

³Metricom's Ricochet service [71] is one of the earliest uses of multihop wireless forwarding or mesh techniques for commercial access services.
In terms of industry standards activities, the IETF Mobile Ad Hoc Networking (MANET) Working Group has been focused on routing protocols at the IP layer for ad hoc networks. The IEEE 802.11s Task Group is working on introducing standard-based mesh networking into IEEE 802.11 Medium Access Control (MAC) layer. The IEEE 802.16\textsuperscript{4} [37], an IEEE working group on the broadband wireless fixed and mobile access standards, already defines a mesh option for the IEEE 802.16 MAC [16, 69]; however, the option does not enjoy widespread support from industry and is not compatible with the normal Point-to-Multipoint mode of IEEE 802.16 MAC. A new Task Group for Mobile Multihop Relay (MMR) [39] in IEEE 802.16 has recently been initiated to enable mesh networking that is compatible with Point-to-Multipoint mode of the IEEE 802.16 standard.

The next generation of networks are envisioned to use multiple orthogonal channels. Also, nodes constituting these networks are envisioned to possess multiple interfaces (radios) and hence can simultaneously communicate with multiple neighbors, over these multiple orthogonal channels. Yet, due to the limited availability of orthogonal channels, co-channel interference, and hence scheduling of transmissions, is still an important issue that has to be handled by the MAC layer. However, in this thesis, I present solutions to two remaining important problems

\textsuperscript{4}IEEE 802.16 technology is often referred to as WiMAX, based on the WiMAX Forum (http://www.wimaxforum.org), which is an industry forum with an objective to enable and certify the interoperability of products based on IEEE 802.16.
that exist in wireless mesh networks. I now give a broad description of these two problems.

Just like in current wired networks, wireless mesh networks are envisioned to have different types of traffic demands, which can be broadly classified into (1) traffic requiring Quality-of-Service (QoS) service and (2) traffic requiring best-effort service. Voice, video, etc. are examples of applications that generate traffic requiring QoS service such as throughput guarantee and delay bounds whereas web browsing, file transfer, messaging, email, etc. are examples of applications that require best-effort service from the network. While there has been noticeable work in the field of QoS traffic in wireless networks, very little has been done for best-effort traffic. In the first part of my thesis, presented in Chapter 3, I design and evaluate routing metrics which take into account existing traffic in the system in addition to wireless channel loss and modulation rates.

The economic case for any large-scale wireless access network deployments, especially those that aim to provide broadband experience, can be heavily influenced by costs related to TAP site acquisitions and operations, such as installation, back-haul costs, tower lease payments, etc [33, 28, 79]. In the second part of my thesis, presented in Chapter 4, I present how deploying low-cost booster TAPs (bTAPs) can increase the coverage of wireless mesh networks with the additional advantage of increasing the system capacity. These bTAPs are strategically deployed and
controlled by the system operator to wirelessly forward traffic between a user and a TAP node.

1.4 Thesis Contributions

In this thesis, I have provided solutions to the two previously mentioned important problems in wireless mesh networks. I have provided novel solutions to these problems and have evaluated my propositions through simulation experiments. My results show that application of my proposed solutions to wireless mesh networks is not only a promising way to address capacity and outage issues of broadband wireless mesh networks, but also to enable an economic deployment model for such networks.

First, I present a number of novel traffic-aware routing metrics that take into account existing user traffic flows in the network. Previous routing metrics have been traffic-unaware, often causing routes with poor throughput to be selected when other better routes are available. These new traffic-aware metrics use information captured through measurements at the medium access control (MAC) layer, which are then exposed to the routing layer. I compare these traffic-aware metrics with existing traffic-unaware metrics under different network scenarios—a topology from a real residential high-speed wireless mesh network commercialized by Chaska.net, and a synthetic grid-like topology, called the Manhattan topology. I find that the Chaska topology is severely limited by interference and hence the
traffic-aware as well as the traffic-unaware routing metrics performed similarly in that topology. Even the use of directional antennas can only marginally improve the average achieved throughput.

However, in a more structured Manhattan topology, the use of the traffic-aware metrics lead to the selection of routes that provide higher throughputs than routes selected with the use of traffic-unaware metrics. Additionally, with the use of directional antennas in the structured topology, all the routing metrics show a marked improvement in the average throughput achieved.

Second, I present the design and analysis of a new technique for increasing the coverage of the wireless mesh network through deployment of low-cost booster TAPs (bTAPs). These bTAPs are strategically deployed and controlled by the system operator to wirelessly forward traffic between users and TAP nodes. Users that have a weak direct signal from the TAP can communicate with the TAP using strategically placed bTAPs. This deployment model is especially suitable for wide area wireless access networks that use centralized management of radio resources, such as WiMAX/802.16, HSPA, and CDMA2000 EV-DO. I analyse the use of bTAPs across different frequency reuse patterns typical of those used in multicell wireless environments for efficient management of costly radio spectrum. The bTAP architecture provides dramatic improvements in outage performance and a sufficient capacity gain to compensate for the radio resources required for forwarding user traffic via bTAPs.
These two solutions for throughput and coverage improvement in wireless mesh networks can simultaneously coexist and can complement each other in a single network. I present an example scheduling scheme that can coordinate the transmissions that must take place at different stages of end-to-end connections between users and the Internet. Nevertheless, both the techniques can exist independent of each other in a wireless mesh network and can individually prove advantageous.

1.5 Thesis Organization

Before the presentation of the two solutions proposed in this thesis, Chapter 2 gives an overview of how the entire system, i.e., the entire wireless mesh network, can simultaneously employ both the presented solutions. Chapter 3 of this thesis first introduces the throughput improvement problem and then presents a discussion on the related work in the area. Following that, I present the design and evaluation of two new routing metrics designed to select high throughput paths. Chapter 4 introduces the coverage improvement problem in wireless mesh networks and discusses related work in the area. The chapter then elaborates on the design and evaluation of my solution to the problem. Finally, I conclude the thesis in Chapter 5 by summarizing the findings in the thesis and describing avenues of possible extensions to this work.
Chapter 2

Overview of Complete System

Before presenting the design and evaluation of the two techniques for improving the throughput and coverage in wireless mesh networks, in this chapter I present an overview of a complete wireless mesh network that can take advantage of the two techniques. In particular, I show how both of the techniques can simultaneously coexist and can complement each other in a wireless mesh network. Nevertheless, both the techniques can independently and individually prove advantageous to a wireless mesh network. This overview of the complete system is presented first in order to supply context for the two techniques presented next in Chapter 3 and 4.

The rest of the chapter is organized as follows. Section 2.1 provides some background information on frequency reuse. Section 2.2 describes how an end-to-end communication can be performed between a client and the Internet. Section 2.3 presents an example scheme to schedule the end-to-end communication, and finally, Section 2.4 summarizes the chapter.
2.1 Frequency Reuse

Frequency reuse is much more prevalent in cellular networks and cellular-like networks such as WiMAX networks than in WiFi networks. In the presence of a single transmitter, only one transmission can be performed at a time on any given frequency. However, a wireless mesh network is composed of multiple nodes, each having a single or even multiple transceivers. This allows for frequency reuse, i.e., the same radio frequency can be reused in a different area at the same time for a completely different transmission. The first generation cellular systems to use frequency reuse had a single transceiver and an associated omni-directional antenna with it, thus allowing for only a single frequency to be used in the entire cell.

However, current radio technologies as well as the development of directional antennas have allowed each cell to be divided into multiple sectors. For example, Figure 2.1 shows an idealized cell in which the TAP node at the center of the cell
Figure 2.2: The (1,6,6) frequency reuse pattern. Each cell is identical and uses the sectorization shown in Figure 2.1.

uses six radios and six directional antennas to implement a cell with six sectors, each sector using a unique orthogonal frequency.

The individual cells, shown in Figure 2.1, can be combined to form a frequency reuse pattern. Figure 2.2 shows a (1,6,6) reuse pattern, which has several cells, each with an identical sectorization. The three number notation for frequency reuse patterns, such as (1,6,6), with the third number representing the number of sectors in a single cell, is clearer compared to the conventional two number notations, such as (1,6). The first number in the conventional notation represents the number of cells in a reuse group, and the second number represents the minimum number of orthogonal channels needed per reuse group so that the interference between reuse groups is minimized. Utilizing advanced signal processing, modern systems
Figure 2.3: End-to-end communication from a client to a gateway TAP. As described later in Chapter 4, the client node A connects to a source TAP, either directly or via a bTAP. This TAP then selects a route, to one of two gateway TAPs (TAP1 and TAP2), using one of the traffic-aware routing metrics, as explained in Chapter 3.

can realize (1,3,6) or (1,3,3) patterns, which would be indistinguishable with the conventional two number notation. As an example, the notation (1,6,6) means that a reuse group consists of a single cell, represented by the “1”. This reuse group, containing only a single cell, uses six channels, represented by the “6”. Finally, the last “6” in the notation represents the number of sectors into which the cell is divided.

Having explained the concept of frequency reuse, I now describe the different steps that need to be take in order for a client to communicate to the Internet.
2.2 End-to-end Communication

In a wireless mesh network utilizing the two new techniques described in this thesis, suppose a client node wants to communicate to the Internet. First, the client has to reach a TAP node, either via a bTAP (booster TAP) or directly. For notational ease, let such a TAP node be called the source TAP. Second, this TAP node has to select a gateway TAP from among several gateway TAPs. Figure 2.3 shows a schematic in which a client node A is attempting to communicate to any of two gateway TAPs, TAP1 or TAP2.

2.2.1 Communication with a TAP

Before a client node can consider routing to any gateway TAP, the client first has to reach either the TAP managing the sector in which the client is physically located or the TAP managing the sector facing that sector. In the example shown in Figure 2.3, node A has chosen to communicate with the source TAP. However, depending upon the propagation characteristics of the environment in which the TAPs, the bTAPs, and the client lie, A might have chosen to communicate with the TAP node managing the sector that lies below the sector that A physically lies in. Whether node A will have to communicate directly to the source TAP, or will have to communicate via a bTAP is decided by the TAP node and is made according to the design to be presented later in Chapter 4. In particular, node A is scheduled to communicate directly with the TAP or via the bTAP depending on whichever op-
tion gives the higher effective rate. The decision of whether node A communicates via the bTAP need not be the same for both downlink and uplink transmissions. For example, node A can receive downlink transmissions via the bTAP, but can transmit uplink packets directly to its source TAP. Once the route from node A to the source TAP, either directly or via the bTAP, is decided, the source TAP needs to find a route to a gateway TAP.

2.2.2 Communication with a Gateway TAP

The source TAP uses one of the traffic-aware routing metrics presented later in Chapter 3 to select a gateway TAP, out of possibly multiple gateway TAPs present in the network. The traffic-aware metrics not only allow the source TAP to select a gateway TAP but also provide the route to the chosen gateway TAP. Such a route consists only of other TAP nodes; no other client node or bTAP node participates in a route from any TAP node to any other TAP. In the example shown in Figure 2.3, the TAP chosen by A, i.e., the source TAP, has two gateways from which to choose, the gateway to the left, TAP1, or the gateway to the right, TAP2. The source TAP would use, for example, the MLS metric presented in Section 3.3 in order to select between TAP1 and TAP2.

The entire route from a client to a gateway TAP consists of at most a single bTAP node. If a bTAP node is present in the end-to-end uplink route from the client to a gateway, then the bTAP is always the first hop from the client. Similarly,
Figure 2.4: A possible scheduling map for the entire wireless mesh system. The sector period and the inter-TAP period need not necessarily alternate. The periods are not shown to scale.

If the bTAP is present in the end-to-end downlink route from a gateway to a client, then the bTAP is always the penultimate node on the route.

Having described how an end-to-end connection can be setup, I now present a scheduling scheme that can be used to coordinate the transmissions in the various stages of the connection.

2.3 Scheduling End-to-End Communication

A number of scheduling schemes could successfully and efficiently coordinate various transmissions in a wireless mesh network such as presented in this thesis. Here I present a sample scheme. Figure 2.4 shows a possible scheduling map for end-to-end communication. The time elapsed at any TAP node is divided into two major parts

- **Sector period**: In this period, any TAP node communicates with the clients for which it is the source TAP, either directly or via a bTAP. The composition of the sector period will be described in further detail later in Section 4.4.4.
• *Inter-TAP period*: During this period, any TAP node can communicate to its neighboring TAP nodes. It is in this period that management packets, such as link-state updates, described in Section 3.5, as well as data packets, are forwarded among TAP nodes. The usage of this period is dependent on the Medium Access Control (MAC) protocol that is used for communication among TAP nodes. For example, an IEEE 802.11-style MAC protocol using Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) could be used in the inter-TAP period. The use of an IEEE 802.11-style MAC protocol is studied in detail in Chapter 3. However, since the TAP nodes are probably mounted on high outdoor towers, each TAP node can be fitted with a GPS receiver, which would allow time synchronization among TAP nodes. This synchronization would in turn allow a Time Division Multiple Access (TDMA)-style MAC protocol to be used for communication between TAP nodes. If so, then the inter-TAP period would be split according to the requirements of the specific TDMA MAC protocol being used.

The sector period and the inter-TAP period need not strictly alternate and can occur simultaneously as well. All sectors that are operating on different channels could operate simultaneously with the inter-TAP period. Figure 2.5 shows an example of a schedule for the (1,6,6) frequency reuse pattern; the periods shown in the figure are not to scale. In this example, for communicating to its neighboring TAP, the TAP uses the same logical channel as used by sector 6, and hence the sec-
Figure 2.5: Overlapping of sector periods with inter-TAP period for the (1,6,6) frequency reuse pattern. Each sector period operates on a separate logical channel, thus requiring six total logical channels for the (1,6,6) pattern. The inter-TAP period can be overlapped with five of the six sector periods except with the sector period using the same logical channel used by the TAP for TAP-to-TAP communication. The periods are not shown to scale.

tor period for sector 6 is the only period that has to be scheduled at a different time than when the inter-TAP period is scheduled. The remaining five sector periods can be scheduled simultaneously with the inter-TAP period. The actual lengths of the periods would depend on several other factors such as the number of active clients, the density of clients, the bandwidth requirements of active clients, and the TAP-to-TAP link bandwidths.
2.4 Chapter Summary

In this chapter, I have presented an overview of the complete system and have explained how the individual pieces, presented later in Chapter 3 and Chapter 4, fit into a wireless mesh system. For end-to-end communication between a client and a gateway TAP, the client has to first communicate with its source TAP, either directly or via a bTAP. The source TAP then selects one gateway TAP out of possibly many gateways present in the mesh network.

In order to coordinate the transmissions needed at different stages of an end-to-end connection, the scheduling of such communication is split into two periods, a sector period and an inter-TAP period. In the sector period, a client node communicates with its source TAP, and in the inter-TAP period, the source TAP communicates with a selected gateway TAP, via other TAP nodes. Sector periods that operate on a different logical channel than does the inter-TAP period of the source TAP (e.g., 5 out of the 6 sector periods in a (1,6,6) reuse pattern) can simultaneously operate with the inter-TAP period.

Using the scheme presented in this chapter, both the techniques presented in Chapters 3 and 4, those of throughput improvement using traffic-aware metrics and coverage improvement using bTAP deployments, can coexist and complement each other in a wireless mesh network.

Nevertheless, both the techniques can exist independent of each other in a wireless mesh network and can individually prove advantageous. In addition to wire-
less mesh networks, the traffic-aware routing metrics can be used in any static wireless network that has scope for multi-hop routing, such as static wireless ad hoc networks. Traffic-aware routing metrics can be used to select routes between any source and destination pair. Similarly, the technique of deploying bTAPs can be applied to any wireless coverage scheme that uses frequency reuse. Frequency reuse is very common in cellular networks and WiMAX networks, even though base stations in current architectures for these networks do not communicate with each other.
Chapter 3

Throughput Improvement in Wireless Mesh Networks

For the purpose of this work, a flow is defined to be a unique source-destination pair. For each flow in a wireless mesh network, often more than one possible route from the source to the destination of the flow exists in the network. For any given flow, the routing protocol deployed in the network typically selects one route among such multiple possible routes. The routing protocol makes this decision on the basis of a routing metric, which is a numerical value assigned to a route that lets the routing protocol compare (and select) between two routes\(^1\). A poorly chosen routing metric can lead to the selection of routes with lower throughput than other possible routes.

For example, a minimum hop count based routing metric does not take into account the modulation rates of the links constituting the route. In Figure 3.1, suppose source node S needs to find the best gateway to which to connect. The actual definition of the “best gateway” can be varied and many, yet for the purpose of this example and this thesis, the best gateway is the gateway to which a client would achieve the highest throughput. If S uses the minimum hop count routing metric for selecting routes, S would choose a two-hop route, S–A–B, each link of

\(^{1}\)Unless otherwise mentioned, the terms “route” and “path” are used interchangeably in this thesis
Figure 3.1: Node S needs to find a route to any gateway. If S uses a minimum hop count metric, then S will select a route to gateway B, which would give it lower throughput than if S were to choose the route to gateway E.

which can transmit at a maximum modulation rate of 1 Mbps, although the three-hop route S–C–D–E, each link of which can transmit at a maximum modulation rate of 11 Mbps, would generally result in higher throughput. As a result, S would get a maximum of \( \frac{1}{2} = 0.5 \) Mbps throughput to gateway B, since the two links of the route interfere with each other. On the other hand, had S chosen gateway E, S would get a throughput of \( \frac{11}{3} = 3.67 \) Mbps, since all the three links of the route to E interfere with each other.

However, just taking modulation rates of the links composing a route does not necessarily give the best route. For example, in the example scenario shown in Figure 3.2, if S were to just take modulation rates of the links into account, S would decide that it would get a higher throughput if it were to select the S–C–D–E route over the S–A–B route. The logic that S uses to come to this decision neglects the existing flow, \( f_1 \), that is transmitting at 4 Mbps. Hence, the throughput achieved over the route S–C–D–E would probably be much less than the throughput achieved on the route S–A–B.
These two examples demonstrate that a routing metric should not only account for modulation rates of links but should also account for existing traffic (and the interference caused by it). Most existing routing metrics are traffic-unaware, and hence routing protocols using such metrics often choose suboptimal routes. In this chapter, I present traffic-aware routing metrics and show that these outperform traffic-unaware routing metrics. I define \textit{traffic-aware} routing metrics as those routing metrics that explicitly account for existing traffic in the system, whereas \textit{traffic-unaware} routing metrics are those routing metrics that do not explicitly account for existing traffic in the system.

In the work presented in this chapter, I do not address the issue of routing to and from users but only consider routing between the TAP nodes. A user generally lies within the communication range of several TAP nodes. The user decides to communicate with that TAP with which the user has the strongest wireless link. As will be explained later in Chapter 4, there are other factors which dictate the communication between a user and a TAP.
The rest of this chapter is organized as follows. Section 3.1 discusses the existing work in the field of routing metrics for wireless networks. Following that, Section 3.3 and Section 3.4 present the design. I present evaluation of my design in Section 3.6 and finally, Section 3.8 summarizes the chapter.

3.1 Related Work

I first discuss current routing metrics that are used for finding routes for best-effort traffic. Following that, I discuss some of the existing work on how throughput guarantees are generally implemented for QoS traffic. Even though QoS-based approaches are not directly relevant to best-effort traffic, there are certain concepts present in QoS-based approaches that can be applied to best-effort traffic as well.

3.1.1 Routing for Best-Effort Traffic

Traditionally, ad hoc network routing protocols have considered minimum hop count as the routing metric for best-effort traffic. However, such a routing metric often leads to the selection of a sub-optimal route, since routes with greater or equal hop count but offering a higher throughput are not considered. Wireless links often show unstable behavior by dropping packets. Such drops might be due to several reasons such as, wireless interference and lossy channel. A metric that only takes hop count into account does not differentiate stable links from unstable ones. Consequently, such a hop count-based metric might select links that require
multiple transmissions for each successfully delivered packet, thus exhibiting a packet delivery ratio of less than 1.

The first routing metric that took into consideration the delivery ratios of the links composing a route is the expected transmission count (ETX) metric proposed by De Couto et al. [20]. The ETX metric of a link is defined as the expected number of data transmissions, including retransmissions, required to send a single packet over that link. As a result, such a metric only makes sense if the underlying MAC protocol does actually perform retransmissions (e.g., IEEE 802.11). Mathematically, the ETX metric of a link \( i \rightarrow j \) is

\[
etx_{i\rightarrow j} = \frac{1}{d_f \times d_r}
\]

(3.1)

where \( d_f \), the forward delivery ratio, is the probability that a data packet over this hop successfully arrives at the recipient, and \( d_r \), the reverse delivery ratio, is the probability that the link layer acknowledgement from the recipient is successfully received at the sender over the hop. The corresponding metric for a route of \( n \) nodes, ETX, is calculated as

\[
ETX = \sum_{i=1}^{n-1} \text{etx}_{(i\rightarrow (i+1))}.
\]

(3.2)

The authors suggested measuring the values for \( d_f \) and \( d_r \) using dedicated, periodic link probe packets.
In a separate work, Draves et al. [22] verified that the ETX metric indeed provided higher throughput than the hop count metric for static multi-hop wireless networks. The authors used a DSR-based routing protocol on a wireless testbed and found that ETX outperformed some other metrics as well, such as per-hop round trip time (RTT) and per-hop packet pair delay. The authors also observed that in a scenario with mobile nodes, the hop count metric outperforms all these metrics, including ETX, since all these metrics depend upon measurements, which, in the presence of mobility, do not react sufficiently fast.

The ETX metric was improved upon by Draves et al. [23] who proposed the weighted cumulative expected transmission time (WCETT) metric. The authors first extend the ETX metric to take raw channel transmission rate (e.g., 1 Mbps, 2 Mbps, 5.5 Mbps, or 11 Mbps for IEEE 802.11b) into consideration. The ETX metric was designed for a system with a single transmission rate. The expected transmission time (ETT) of a link \( i \rightarrow j \) is defined as

\[
ett_{i \rightarrow j} = \text{etx}_{i \rightarrow j} \times \frac{S}{B_{i \rightarrow j}}
\]  

(3.3)

where \( S \) is the size of the data packet, and \( B_{i \rightarrow j} \) is the raw data rate of the channel between \( i \) and \( j \). Consequently, the ETT metric for a route composed of \( n \) nodes is represented as

\[
\text{ETT} = \sum_{i=1}^{n-1} \text{ett}_{i \rightarrow (i+1)}
\]  

(3.4)
The ETX metric does not account for multiple orthogonal channels and considers that all links of a route operate on a single channel. In fact, the ETX metric assumes that the entire network operates on a single channel. Hence, in order to account for the interference among links (in a route) that use the same channel, the authors introduce another component

\[ X_j \ (1 \leq j \leq k) = \sum_{\text{Hop } i \rightarrow (i+1) \text{ is on channel } j} \text{ett}_{i \rightarrow (i+1)} \]  \hspace{1cm} (3.5)

where \( k \) is the number of orthogonal channels in the system, \( X_j \) is simply the sum of transmission times of all hops of a route that are on channel \( j \). The complete WCETT metric is a combination of the two components, represented by Equation 3.4 and Equation 3.5, and is defined as

\[ \text{WCETT} = (1 - \beta) \times \text{ETT} + \beta \times \max_{1 \leq j \leq k} X_j \]  \hspace{1cm} (3.6)

where \( \beta \) is a tunable parameter subject to \( 0 \leq \beta \leq 1 \). In the absence of multiple channels, \( \beta \) is set to 0 and the WCETT metric reverts to the ETT metric. The first term ("\((1 - \beta) \times \text{ETT}\)" in Equation 3.6) reflects the total transmission time taken over all hops of the route, whereas the second term ("\(\beta \times \max_{1 \leq j \leq k} X_j\)" in Equation 3.6) reflects the total time taken on the channel that is most bottlenecked.
Yang et al. [84] proposed the Interference-aware Resource Usage (IRU) metric for a link $i \rightarrow j$ as

$$iru_{i\rightarrow j} = ett_{i\rightarrow j} \times N_{i\rightarrow j}$$

(3.7)

where $N_{i\rightarrow j}$ is the set of neighbors with which the transmission on link $i \rightarrow j$ interferes. The corresponding metric for a route of $n$ nodes, IRU, is calculated as

$$IRU = \sum_{i=1}^{n-1} iru_{i\rightarrow(i+1)}$$

(3.8)

This metric, unlike the previously mentioned ones, attempts to take contention into account in an attempt to route around areas of high node density. However, the IRU metric equates node density to traffic contention, which is not always the case. For example, a link in a high density part of the network might get high throughput if none of its neighbors are transmitting, whereas a link in a relatively low density part of the network might get less throughput since its neighbors are transmitting.

The only dynamic variable in all of the above metrics is the packet loss probability. All other quantities depend only on network topology and thus are not affected by network load. Moreover, with a single transmission rate and constant packet-size, all metrics assign to the links a cost that is proportional to the ETX metric. Though the calculation of the ETX metric for a link involves sending periodic probes, the ETX authors [20] themselves explained that the presence of traffic
at either of the two end nodes of the link merely delays the transmission of probe packets, thus not affecting the effective loss rate of the link. Even the IRU metric, which includes the number of interfering nodes, is proportional to ETX when the node density is homogeneous.

Zhai et al. [87] identified that packets belonging to a flow face both inter-flow as well as intra-flow contention. Similar to De Couto et al. [20], the authors observed that in an IEEE 802.11-based MAC, the maximum throughput for a chain topology is only one quarter of the channel bandwidth. The authors proposed a few scheduling schemes to allow simultaneous transmissions at nodes that are physically distant enough to not cause mutual interference. Zhai et al.’s work, however, does not address the issue of routing.

Jain et al. [42] analyzed the theoretical maximum throughput that can be achieved given a specific traffic load on a specific placement of wireless nodes. The authors construct a conflict graph that models the interference among neighboring nodes, assuming the ability to make centralized scheduling decisions. They proposed a scheme to obtain the optimal throughput in the system. Using ns-2 [26] simulations with an IEEE 802.11 style MAC protocol (thus allowing uncoordinated transmissions and MAC contention), the authors tested the routes as selected by their omniscient algorithm and found tangible throughput improvement. They hence concluded that an interference-aware routing protocol can lead to throughput gains as compared to a routing protocol that is oblivious to interference.
Similar to the problem addressed by Jain et al., Kodialam and Nandagopal [47] addressed the problem of jointly routing flows and scheduling link transmissions, provided the rates required by the different flows are known. Kodialam and Nandagopal propose approximation algorithms for link scheduling, an aspect absent in the work by Jain et al. Later, Kodialam and Nandagopal [48], extended their work to include interference among nodes, i.e. they differentiated between transmission range and carrier sensing range. Their previous work [47] had avoided interference by assuming a sufficient number of orthogonal channels. This differentiation in their later work allowed them to infer whether a given end-to-end demand vector of flow requirements is feasible or not; if feasible they also present link channel assignment algorithms. In parallel work, Alicherry et al. [7] addressed the throughput maximization problem in a network that is restricted in the sense that the maximum distance at which two nodes can interfere with each other is considered to be some multiple of the transmission range. This is unlike Kodialam and Nandagopal's work [48], which uses a completely general network model thus allowing for interference between arbitrary nodes.

Raniwala and Chiueh [67, 68] proposed an architecture and related algorithms for an IEEE 802.11 based wireless mesh network. The architecture considers each gateway to be the root of a spanning tree, and each mesh node communicates with a single gateway. Based on such a system, the authors propose algorithms for load-aware channel assignment and routing based on metrics such as shortest path,
residual link capacity of the link connecting the gateway to the wired network, and residual capacity along an entire path. However, since the routing tree is already fixed, these metrics are used to select between multiple gateways and not to choose between multiple routes to the same gateway.

3.1.2 Admission Control for QoS Traffic

In routing for QoS traffic in multi-hop wireless networks, the spare bandwidth (or some other resource such as remaining battery life) along a route is the criteria used for selection of routes and for admission control [18, 86, 17, 77, 44, 74]. Any flow attempting to enter the network is expected to present its requirement, such as its minimum required bandwidth, and depending upon the available bandwidth, the flow is either allowed or disallowed into the network.

Lee et al. [49] designed a lightweight and adaptive in-band signaling (control information is transported along with data) framework to support reservation, restoration, and adaptation. Their framework is targeted towards adaptive applications that require a minimum bandwidth guarantee from the underlying network.

In the Contention-aware Admission Control Protocol [86], Yang and Kravets introduced the idea of c-neighborhood available bandwidth and local available bandwidth in an ad hoc network. Intuitively, the local available bandwidth at a node is the amount of unconsumed bandwidth as observed locally by the node. The
c-neighborhood available bandwidth is defined as the maximum bandwidth that a node can use for transmitting without decreasing the bandwidth being used by any existing flow in the carrier sensing range of the node. A new flow is admitted at a node only if there is enough c-neighborhood bandwidth and local available bandwidth. The authors use the fraction of idle channel time as an indication of local available bandwidth and propose two active approaches and one passive approach to calculate c-neighborhood available bandwidth. In the active approaches, each node broadcasts queries, over multiple hops or with higher transmission power, in order to reach all nodes within the carrier sense range of the querying node. Nodes receiving the query reply back to the querying node with information about their respective local available bandwidths. The querying node uses this information to estimate the c-neighborhood available bandwidth. In the passive approach, a node passively monitors the wireless channel for any activity above a particular threshold, called the neighbor-carrier-sensing threshold. This threshold is set to much lower than the carrier-sensing threshold. The node keeps track of the fraction of time, over some time period, during which it cannot sense any activity above the neighbor-carrier-sensing threshold and uses this fraction, along with the nominal bandwidth of the channel, to calculate the available c-neighborhood bandwidth.

In related work, Yang and Kravets [85] proposed the Multi-Priority Admission and Rate Control (MPARC) protocol for admission control of multi-priority QoS
traffic in ad hoc networks. The admission control scheme maintains the throughput of existing flows even with the later arrival of best effort flows and other QoS flows with lower priority. In addition to performing admission control for QoS flows, the MPARC protocol polices the rate of best effort flows, thus preventing best effort flows from degrading the performance of QoS flows.

Chakeres and Belding-Royer [17] propose the Perceptive Admission Control (PAC) protocol, which targets high network utilization without allowing congestion in mobile wireless networks. The PAC protocol monitors the wireless channel by considering the entire area that might be affected by a new flow. The protocol uses the time that the wireless channel is sensed as busy as an indicator of available bandwidth. This estimation of the available bandwidth is then used to make admission control decisions so that the necessary quality of service can be provided to admitted flows.

Tung et al. [77] suggested that there is little difference among routes that interfere with each other. The authors hence chose to divide the network into non-interfering clusters, each representing a separate interference domain. Routing operates at the cluster level, with shortest routes being chosen within each cluster. Using this approach, the authors provided admission control for flows with QoS requirements. However, real networks can seldom be partitioned into independent non-interfering clusters, and hence the authors suggested choosing from existing clustering schemes for mobile wireless networks.
Jia et al. [44] proposed an algorithm to solve the Ad Hoc Shortest Widest Path (ASWP) problem in order to admit QoS flows into the network. Given the source-destination pair of an impending flow, the widest feasible path is found. The *widest feasible path* is the path with the maximum spare bandwidth across all hops of the path such that each hop can at least support the requirement of the impending flow. Since there might be multiple such widest paths, the shortest among them is chosen as the route for the new flow; the shortest widest path chosen in this way is proven to be loop free. The ASWP is an NP-Complete problem, and hence the authors proposed a distributed algorithm that approximates the optimal solution.

Tang et al. [74] addressed the problem of joint channel assignment and routing in a stationary network consisting of $C \geq 1$ channels, where each node has $Q \leq C$ network interface cards (NICs). All nodes are assumed to have the same transmission range and same maximum distance at which any node can cause interference. In such a network, the authors provided a heuristic for channel assignment that induces a network topology that has the minimum number of interfering links among all $K$-connected\(^2\) topologies, where $K$ is a user defined parameter. Once the channel assignment is decided, Tang et al.'s algorithm sets up routes for QoS connections, each connection having some bandwidth requirement. The authors called this the Bandwidth-Aware Routing (BAR) problem and provided a polynomial time algorithm to optimally solve it, provided the flows are splittable; i.e., all

\(^2\)A graph is called $K$-connected if at least $K$ vertices (and their associated edges) have to be removed to make the graph disconnected
packets of a single flow need not be routed along a single route. However, when
the flows are not splittable, the authors provided a heuristic based on the maxi-
mum bottleneck capacity path.

Lindgren and Belding-Royer [52] suggested storing multiple paths to the des-
tination, thus allowing the use of alternate paths in the event of route breakages.
This is particularly effective in a mobile environment, where route breakages due
to mobility of nodes is common. The authors coupled this with an admission con-
trol scheme in order to provide the required QoS to real-time traffic.

In order to regulate resource consumption in the network, Maltz and John-
son [54] extended the Dynamic Source Routing (DSR) [45] protocol with two types
of soft-state, path-state and flow-state. Path-state allows the originator of data pack-
etts to control the use of network resources, such as remaining battery life of nodes
and available capacity in the network. On the other hand, flow-state allows the
intermediate nodes along a route to control the network resources.

Renesse et al. [21] proposed the Adaptive Admission Control (AAC) algorithm
for servicing QoS flows in a mobile ad hoc network. The AAC algorithm considers
mobility as well as multihop interference and can be adapted to almost any on-
demand routing protocol such as DSR or AODV.

Chen and Heinzelman [18] proposed having an admission control scheme along
with a feedback scheme in order to handle QoS traffic in mobile ad hoc networks.
The authors approximate residual bandwidth at each node of the network and use
it to perform admission control for new QoS flows or to provide feedback to ex-
isting flows. They suggest two techniques for estimating the residual bandwidth
at a node: a passive technique based on overhearing traffic around a node and a
proactive technique based on explicitly sending notification of bandwidth usage
among neighboring nodes.

All of the previous work on servicing QoS traffic in a multihop wireless net-
work, described above, calculates network utilization ($U$) at each node, and given
the maximum bandwidth of a link as $B_{max}$, calculates the spare bandwidth across
the link as

$$B_{spare} = (1 - U) \times B_{max}$$

If a flow, which as defined earlier, is a unique source and destination pair, attempts
to enter a network with a bandwidth requirement of $B_{required}$, the flow is allowed
if there exists a route consisting of nodes, each of which has $B_{spare} \geq B_{required}$;
otherwise, the flow is disallowed. Some common metrics, out of several possibili-
ties, used for estimating network utilization ($U$) are

- Channel busy time,
- Size of the MAC layer congestion window,
- MAC layer queue length, and
- Number of MAC layer collisions.
The work presented in this chapter does not perform any form of admission control. However, the ideas presented in my work can be applied to QoS traffic as well, albeit with the addition of admission control. Given an arbitrary source-destination pair, the traffic-aware routing metrics, presented in this thesis, can not only select routes between the source and the destination, but also can provide an estimate of the throughput along the selected route. This estimate can be used to control the admission of new flows by admitting a new flow only if the estimated throughput satisfies the requirement specified by the new flow.

3.2 Design Basics

The design of a routing metric that is to operate in a network with multiple simultaneous flows should take into account the following factors:

- *Link loss probability*. Transmitted packets might be lost due to various reasons other than interfering transmissions. Packets might be dropped due to existence of interference from sources outside the system, for example, microwave devices, due to buffer overflow, or due to several other reasons. Normally, any well planned system such as a wireless mesh network, and especially ones operating in licensed spectrum, would be engineered to avoid such lossy links (for example, by choosing a lower modulation rate), but intermittent losses might still occur. However, packet losses due to interference cannot be completely eliminated, especially in a CSMA-based MAC protocol.
TDMA-based MAC protocols might be able to assign unique time slots to interfering transmissions, thus avoiding packet collisions.

When packet losses cannot be avoided by scheduling alone, the packet loss probability across links needs to be quantified and taken into account. One way of quantifying the packet loss probability is to measure the packet loss rate across each link. For each link, such measurement can be done through dedicated probe packets as well as by keeping track of the number of transmissions that were required to successfully deliver a single data packet over the link. I use both the approaches in my work in this chapter.

I have implemented the broadcast packet probe technique suggested by De Couto et al. [20]. In this measurement technique, broadcast packet probes are periodically transmitted, and each node keeps track of how many probes it received from its neighbors in the past few seconds (I use a value of 20 seconds for this period). This information is then piggybacked on future probes sent by the node. The advantage of using broadcast probe packets is that a single broadcast packet can be used to measure all outgoing links from a node. However, since broadcast packets are transmitted at the highest data rate in the basic rate set, measurements made using broadcast packets might not accurately reflect the loss rate of data packets that may be transmitted at a different modulation rate. Many current wireless transceiver units are capable of transmitting at multiple data rates. For example, IEEE 802.11b-
compliant radios can transmit at any of four transmission rates: 1 Mbps, 2 Mbps, 5.5 Mbps, and 11 Mbps. The more recent IEEE 802.11a/g-compliant radios can choose a rate from an even higher number of transmission rates, up to 54 Mbps. Whenever data packets are transmitted, I augment the information collected using broadcast packets by counting the MAC layer re-transmissions for unicast data packets. This technique has the advantage of measuring the loss probability at the exact transmission rate at which the data transmissions take place.

- *Link modulation rate.* As mentioned earlier, IEEE 802.11a/b/g-compliant radios can choose to transmit at one of many possible transmission rates. The modulation rate at which any particular link operates can be obtained through offline or out-of-band measurements. Since my work is specifically geared to a static set of APs, such an assumption is not impractical. Better planning and placement of access points, as is expected in wireless mesh networks, might allow the use of offline measurements to statically determine modulation rates that should be used for reliable communication. Otherwise, protocols such as Receiver Based AutoRate (RBAR) [35] can be used to determine the appropriate modulation rate. However, the use of RBAR for this purpose might be more than needed, since for the work presented in this chapter, only the average link modulation rate is required, whereas RBAR
tries to determine the link modulation rate for each individual transmitted packet.

- *Interference caused by existing flows.* In a wireless mesh network, it is impractical to assume that there will be only a single flow in the network, especially since these networks are envisioned to provide simultaneous connectivity to numerous clients and are hence expected to have multiple flows active simultaneously. Since a wireless mesh network is intended for widespread usage, such an assumption would generally not be met in practice. Hence, the effect of other existing flows needs to be taken into account, something that is not done by existing routing metrics. Next, I introduce *MAC Layer Share* (MLS), which is used to take the interference of existing flows into account.

### 3.3 MAC Layer Share (MLS)

I define the *MAC layer share of a node* to be the fraction of time during which the node *is free to transmit and/or receive packets.* Unless otherwise mentioned, I always mean unicast packet by the term "packet". The presence of virtual or physical carrier at a node is an indication of the existence of interfering transmissions around that node. Any arbitrary period of time at a node $i$, can be decomposed into the following parts:

- $T_r$: The fraction of time during which node $i$ is receiving packets destined to node $i$ for successfully received packets.
• $T_i$: The fraction of time during which node $i$ is transmitting packets to neighbors of node $i$ for successfully transmitted packets.

• $T_d$: The fraction of time during which node $i$ is deferring for transmissions that interfere with node $i$.

• $T_c = $ The fraction of time during which node $i$ experienced packet collisions.

• $T_f = 1 - (T_r + T_t + T_d + T_c)$: The remaining free fraction of time at node $i$.

Hence, the MAC layer share (MLS) of node $i$ can be represented as

$$mls = T_f = 1 - (T_r + T_t + T_d + T_c) .$$  \hspace{1cm} (3.9)

All of the required components, i.e., $T_r$, $T_t$, $T_d$, and $T_c$ (or just $T_f$) are easy to keep track at the MAC layer, and even though most commonly available device drivers do not export interfaces to higher layers such as the routing layer for extracting these values, similar information is exposed by some wireless cards such as the DARPA GloMo Radio API [25]. Future device drivers may export the required interfaces, if such information proves beneficial to the higher layers.

The approach presented in this work does not incorporate the current contention window at the MAC layer. I have deliberately made this choice for two reasons. First, the contention window in a CSMA/CA-based medium access control (MAC) layer is highly dynamic, and as shown by Draves et al. [23], incorpo-
rating the contention window into a routing metric provides only marginal performance benefits and is hence not worth the extra complication. Second, contention windows are not used in some MAC protocols, such as TDMA-based MAC protocols (e.g., IEEE 802.16), used in emergent wireless mesh systems; incorporating the congestion window would make the definition of MLS inapplicable to systems with such MAC protocols.

Hu and Johnson [36] proposed keeping track of MAC layer utilization over some time window. They defined instantaneous MAC layer utilization at a node to be 1 if the node is either (1) backing off at the MAC layer, (2) deferring due to physical or virtual carrier sense or due to required interframe spacing, or (3) the node has at least one packet ready for transmission at the transmission queue of the wireless interface; instantaneous MAC layer utilization is 0 otherwise. The definition of MLS does not account for the transmission queue of the wireless interface because any significant effect of any packet in the queue is reflected in the measured values for $T_r$, $T_t$, or $T_c$. Though the definition of MAC layer utilization is somewhat similar to the definition of MLS, the use of MLS is entirely different. Whereas Hu and Johnson used some ad hoc thresholds (such as utilization exceeding some predetermined threshold value) to trigger certain routing layer and transport layer optimizations, I use the measured MLS value of a node to calculate an estimate of the available throughput across any outgoing link of the node. A very similar definition of utilization, as proposed by Hu and Johnson, has been used by Chen
and Heinzelman [18] to calculate the spare bandwidth at a node, which is then used to perform admission control for QoS flows.

3.3.1 Per Link Throughput

Having calculated the above mentioned three factors, link loss probability, link modulation rate, and MLS of the source node of the link, I calculated the estimated throughput across the link as

$$
e_l = mls \times m \times (1 - p_{loss})$$

(3.10)

where $mls$ is the MAC layer share of the source node of the link $l$, $m$ is the modulation rate of the link, and $p_{loss}$ is the link loss probability from the source of the link to the destination of the link. For example, if the link were to be perfect ($p_{loss} = 0$), and if there was no other traffic around the source node ($mls = 1$), then the estimated throughput across the link would be close to the modulation rate of the link, $m$. The achieved throughput across the link would always be less than $m$ because of periods in the MAC protocol that are not used for data transfer. For example, in IEEE 802.11, periods such as the Short Interframe Spacing (SIFS), the Distributed Interframe Spacing (DIFS), and the time taken to transmit the acknowledgement of a received data packet bring the effective data rate to lower than the nominal modulation rate. If however, $p_{loss} \neq 0$, but there is no traffic around the source node, then the estimated throughput would be $m \times p_{success} = m \times (1 - p_{loss})$. 
Similarly, if there is traffic around the source node then, \( mls < 1 \), and hence the estimated throughput is less than the modulation rate of the link.

### 3.3.2 End-to-End Throughput Along a Route

The end-to-end estimated throughput of a multi-hop flow over a path is the maximum throughput it can achieve subject to the condition that no queue along the path overflows. The estimated link throughput, presented in Equation 3.10, takes \textit{inter-flow} interference into account; this value is an estimate of the throughput across the link if the link were to operate in isolation from interference caused by other links of the same path. However, over a multihop route, different links constituting a route interfere with each other, i.e., routes exhibit \textit{intra-flow} interference. To handle this, I construct an \textit{intra-flow} link contention graph that captures interference relationships among the links of a route. Each vertex in the contention graph corresponds to a constituent link of the route. An edge exists between two vertices if they contend with each other. Each clique \( j \) of the contention graph, i.e., a complete subgraph of the contention graph, provides a constraint on the maximum achievable throughput \( T \) over the route in the form

\[
\sum_{l \in j} \frac{T_{el}}{T_{l}} \leq 1 ,
\]  

(3.11)
i.e., the sum of the normalized time shares required to send the amount of traffic $T$ over each link of the clique cannot exceed 1. From this it follows that

$$ T \leq T_j, \quad \forall j \text{ (i.e., for all cliques)} \tag{3.12} $$

where

$$ T_j = \left( \sum_{l \in j} \left( \frac{1}{\epsilon_l} \right) \right)^{-1}. \tag{3.13} $$

Finally, the available throughput across the entire route is computed as

$$ \text{Route Throughput} = \min_{\forall j} (T_j) \tag{3.14} $$

i.e., the maximum end-to-end throughput across the route is the minimum of the maximum throughputs achievable along each of the constituent cliques of the route. I now present an example for clarification.

Figure 3.3 shows an example of computation of the available path throughput in a simple case. Link 1, 2, and 3 here, with respective maximum link throughputs of 50 packets/s, 100 packets/s, and 25 packets/s, form a clique since each of these links interfere with the rest. Consequently, according to Equation 3.13, the maximum clique throughput achievable across a route composed of links 1, 2, and 3 is

$$ \left( \frac{1}{50} + \frac{1}{100} + \frac{1}{25} \right)^{-1} = \frac{100}{7} \text{ packets/s} > 14 \text{ packets/s}. $$
Figure 3.3: A 4-hop flow and its intra-flow contention graph. Dotted lines denote the interference range, and dotted circles denote cliques. The path available throughput is \[
\min \left\{ \left( \frac{1}{50} + \frac{1}{100} + \frac{1}{25} \right)^{-1}, \left( \frac{1}{100} + \frac{1}{25} + \frac{1}{20} \right)^{-1} \right\} = 10 \text{ packets/s}.
\]

Similarly, each of links 2, 3, and 4 interfere with the rest and hence form a clique. Again, according to Equation 3.13, the maximum clique throughput achievable across a route composed of links 2, 3, and 4 is
\[
\left( \frac{1}{100} + \frac{1}{25} + \frac{1}{20} \right)^{-1} = 10 \text{ packets/s}.
\]

From Equation 3.14, it follows that the maximum throughput achievable along the route composed of links 1, 2, 3, and 4 is the minimum of 10 packets/s and 14 packets/s, i.e., 10 packets/s.

Among multiple routes from a source to a destination, the route with the maximum route throughput is selected. Among multiple routes with the same value of
the route throughput, the shortest route is chosen. I call this traffic-aware routing metric, the MLS metric since the metric is based on mls (Equation 3.9).

3.3.3 Queuing Theory Model of a Clique

The maximum throughput achievable across a clique (i.e., Equation 3.13) can be derived from a queuing theory approach as well. Each link \( l \) forming a clique is modeled as a queue with packet arrival rate of \( \lambda_l \) and a service rate of \( \mu_l \). Since all the links of a clique interfere with each other, packets arriving at each of these links need to be handled exclusively. As shown in Figure 3.4, such a clique of interfering links can be modeled as a system with a single server and several queues, each queue modeling a different link of the clique.
The expected service rate $\bar{\mu}$ for such a system with $n$ queues can be represented as

$$\frac{1}{\bar{\mu}} = \frac{\lambda_1}{\sum_{i=1}^{n} \lambda_i} \times \frac{1}{\mu_1} + \frac{\lambda_2}{\sum_{i=1}^{n} \lambda_i} \times \frac{1}{\mu_2} + \ldots + \frac{\lambda_n}{\sum_{i=1}^{n} \lambda_i} \times \frac{1}{\mu_n},$$  \hspace{1cm} (3.15)$$

where each term of the form

$$\frac{\lambda_q}{\sum_{i=1}^{n} \lambda_i} \times \frac{1}{\mu_i}$$

is the product of the probability that the event to be processed is from queue $q$ (with arrival rate of $\lambda_q$) and the service time (inverse of the service rate $\mu_q$) required to process that event. The total arrival rate in the system is

$$\Lambda = \sum_{i=1}^{n} \lambda_i.$$  \hspace{1cm} (3.16)$$
In order to avoid overload in the system, the service rate of the system (Equation 3.15) should be greater than the arrival rate (Equation 3.16). That is,

\[
\lambda \leq \bar{\mu}
\]

\[
\frac{\lambda}{\bar{\mu}} \leq 1
\]

(3.17)

\[
\frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} + \ldots + \frac{\lambda_n}{\mu_n} \leq 1 \quad \text{(from Equations 3.15 and 3.16)}
\]

\[
\sum_{i=1}^{n} \frac{\lambda_i}{\mu_i} \leq 1.
\]

The network topology of a clique is such that the output of one wireless link is the input to the next wireless link. Thus, in order to avoid overloading any single link forming the clique, all the queues must have the same arrival rate, i.e.,

\[
\lambda_1 = \lambda_2 = \ldots = \lambda_n = T.
\]

Replacing \(\lambda_i\) in Equation 3.17 with \(T\) (the maximum arrival rate), we get exactly the same equation as Equation 3.11. The estimated throughput of link \(l\), i.e., \(\epsilon_i\) in Equation 3.11, is the service rate \((\mu_i)\) at which packets arriving at link \(l\) can be processed.
In this section, I have presented the calculation of estimated link throughput $\epsilon_l$ from measurement values. Nevertheless, it is also possible to calculate the value for $\epsilon_l$ based on an analytical model. Analytical models are normally tied to the MAC protocol being used, and since IEEE 802.11 is still the most commonly used distributed MAC protocol, I developed an analytical model for IEEE 802.11-style protocols; this model is presented in the next section.

3.4 Metric Based on Analytical Model for IEEE 802.11

The primary purpose in choosing an analytical model is to predict the estimated throughput along any route, chosen by any routing metric. Most routing metrics simply select a route and rely on higher layer congestion control mechanisms, such as TCP, to control the rate at which packets are inserted at the source end of the route. However, prior research has shown that TCP does not perform well in multihop wireless networks, and several attempts have been made to improve TCP’s performance in wireless networks [9, 32, 55]. In order to avoid the vagaries of TCP in a wireless network, I chose an analytical model that has been shown [72] to predict the estimated throughput of routes with reasonable accuracy, even in the presence of existing traffic. After choosing the analytical model, I evaluate a metric based on the analytical model and compare it with the other metrics. Since the model is traffic-aware, the resultant metric is likewise traffic-aware.
Garetto et al. [29, 30] introduced a general decoupling technique to analyze the behavior of each node in an IEEE 802.11 network with arbitrary topology. One important limitation of this previous work is that it was limited to the case in which each source sends traffic to a single neighboring node. Since this case allows to analyze only particular traffic patterns, the analysis has to be extended to the general case in which a node transmits to multiple neighbors. The term "node" is used to refer to one network interface, i.e., one instance of the IEEE 802.11 MAC protocol, which is shared by all flows passing through the node. I first review, in Section 3.4.1, the modeling framework introduced by Garetto et al [29, 30]. The extension of the analysis to the case of multiple receivers is described in Section 3.4.2.

3.4.1 Review of Modeling Framework

In IEEE 802.11, the behavior of a node is determined by what it senses on the channel, i.e., by the occupation of the "air" around it in the frequency spectrum used. In a wireless mesh network, the state of the channel can be perceived differently by different nodes. As a consequence, existing techniques developed in the "single cell" case, usually based on the classic analysis by Bianchi [13], are not applicable, and the individual channel view from each node's perspective has to be considered.
Figure 3.5: Example of the evolution of the channel state with time, perceived by a node. Possible time instants of state changes are shown by the arrows.

The evolution of the channel state experienced by a node can be described as a renewal process with four different states, as illustrated in the example of Figure 3.5. The four states are

1. Idle channel,

2. Channel occupied by a successful transmission of the node,

3. Channel occupied by a collision of the node, and

4. Busy channel due to activity of neighboring nodes, detected by means of either physical or virtual carrier sensing (the NAV).

The arrows in the figure point to possible time instants of state changes. The time intervals during which the node remains in each of the four states above are denoted by $\sigma$, $T_s$, $T_c$, and $T_b$, respectively. The idle channel state can last only for a fixed duration $\sigma$, equal to one backoff slot, after which the state can either change to one of the remaining states or it can continue to be idle (for another $\sigma$ period). The duration of the other intervals can be variable with general distribution, depending on the access mechanism (basic access or RTS/CTS), the frame size, and the
sending rate of the transmitting node(s). The terms $T_s$, $T_c$, and $T_b$ include a deterministic idle slot at the end [30]. Let $\Pi_c, \Pi_s, \Pi_c, \Pi_b$ be, respectively, the occurrence probabilities of the four states described above. To compute these probabilities, the events that can occur after an idle slot has elapsed need to be specified.

When a node running the IEEE 802.11 MAC protocol wants to transmit and finds the channel busy, the node sets a backoff counter to a random number of slots picked from its current congestion window. After waiting for a predetermined amount of time, equal to either the Distributed Interframe Spacing (DIFS) or the Extended Interframe Spacing (EIFS), if the backoff counter currently has a non-zero value, the backoff counter is decremented whenever the node experiences an idle slot, i.e., a duration of $\sigma$ during which the carrier is sensed to be free. Let $\tau$ be the probability that the backoff counter of the node reaches zero after an idle slot; let $e$ be the probability that when the backoff counter reaches zero, the transmission queue is empty; let $p$ be the probability that a transmission of the node is not successful; finally, let $b$ be the probability that if the node does not transmit after an idle slot, the channel becomes busy because of the activity of other nodes. Then,
the occurrence probabilities of the four channel states can be expressed as follows

\[ \Pi_s = \tau (1 - p)(1 - e) , \]
\[ \Pi_c = \tau p(1 - e) , \]
\[ \Pi_\sigma = [(1 - \tau) + \tau e](1 - b) , \text{ and} \]
\[ \Pi_b = [(1 - \tau) + \tau e] b . \]

(3.18)

Using standard renewal-reward theory, the throughput of the node (expressed in packets/s) is given by

\[ T_P = \frac{\tau (1 - p)(1 - e)}{\Pi_s \bar{T}_s + \Pi_c \bar{T}_c + \Pi_\sigma \bar{T}_\sigma + \Pi_b \bar{T}_b} . \]

(3.19)

The probability \( \tau \) is a deterministic function of \( p \), which depends only on backoff parameters such as the window size, and the number of backoff stages. The complete expression of \( \tau \) for IEEE 802.11 that takes into account the maximum retransmission limit jointly with the maximum window size, is given by

\[ \tau = \frac{2q(1 - p^{m+1})}{q(1 - p^{m+1}) + \frac{W_0}{1 - p - p(2p)^{m'}(1 + p^{m-m'}q)}} \]

(3.20)

where \( q = 1 - 2p \), \( W_0 \) is the minimum window size, \( m \) is the maximum retry limit, and \( m' \) is the backoff stage at which the window size reaches its maximum value, \( m' \leq m \).
The average durations $\bar{T}_s$ and $\bar{T}_c$ of a successful transmission or of a collision, respectively, in which the node is involved are also known *a priori* [13], depending on the distributions of packet sizes and data rates. The only unknown variables in Equation 3.19 are

1. The occurrence probability $b$ of a busy period, and its average duration $\bar{T}_b$,

2. $p$, the conditional packet loss probability, and

3. $e$, the conditional probability of an empty buffer.

The value of $e$ depends on the local traffic load. The values of $b$, $\bar{T}_b$, and $p$, instead, depend on the interaction of the node with the rest of the network. In the original model [29], the authors described an iterative technique to compute these quantities that allowed the authors to solve for the entire network and thus predict analytically the stationary behavior of each node, including the throughout achieved by the node. In particular, the authors have proposed a methodology to evaluate the fraction of time $f_B$ during which the channel is sensed busy, as well as the average duration $T_b$ of a busy period. The value of $f_B$ is related to the model through the expression

$$f_B = \frac{\Pi_b \bar{T}_b}{\Pi_e \bar{T}_e + \Pi_c \bar{T}_c + \Pi_\sigma \sigma + \Pi_b \bar{T}_b} .$$

Moreover, the authors proposed a technique to compute the packet loss probability $p$ on the (single) link used by the node. In the work in this thesis, I am not
concerned with the *analytical* computation of \( f_B, \bar{T}_b \) and \( p \) [29]. Instead, I measure the values for these parameters at the MAC layer.

### 3.4.2 Extension to Multiple Receivers

In the case of a node having multiple receivers, it is necessary to take into account the fact that each outgoing link experiences, in general, a different packet loss probability. Let \( n_L \) be the number of links used by the node to forward its traffic. Each link is associated to a packet loss probability \( p_{r_i}, 1 \leq i \leq n_L \), which maps into a per-link transmission probability \( r_i \) by Equation 3.20.

To analyze a node’s behavior, the arrival rates \( \lambda_i \) of packets destined to the node’s outgoing links need to be known. The arrival rate of packets for an outgoing link \( i \) at a node is the rate at which packets arrive for transmission on link \( i \). These packets can either originate at the node itself, or the packets could be those being forwarded by the node. Since the quality, such as modulation rate and link loss probability, of the outgoing links can be different, the traffic pattern is important. Let

\[
\Lambda = \sum_{i=1}^{n_L} \lambda_i
\]

be the total arrival rate of packets to the node, which has \( n_L \) outgoing links. For each link \( i \), the following weight is introduced

\[
w_i = \lambda_i / \Lambda, \quad 1 \leq i \leq n_L
\]
which is equal to the probability that a generic packet to be served by the MAC protocol has to go over link \( i \). Next, the average number \( c_i \) of channel slots (i.e., one of the four channel intervals described in Section 3.4.1) required to serve a packet destined to link \( i \) is computed as

\[
c_i = \frac{W_0}{2q} \left[ 1 - (2p)^{m'+1} + q(m - m')(2p)^{m'} \right]. \tag{3.22}
\]

Then, the probability \( s_i \) that a generic channel slot is part of the service time of a packet belonging to link \( i \) is computed as

\[
s_i = \frac{(w_i c_i)}{\sum_{j=1}^{n_L} (w_j c_j)} . \tag{3.23}
\]

Using the above quantities, the average transmission probability of the node can be computed as

\[
\bar{r} = \sum_{i=1}^{n_L} s_i \tau_i , \tag{3.24}
\]

and the average packet loss probability can be computed as

\[
\bar{p} = \sum_{i=1}^{n_L} s_i \tau_i p_i / \bar{r} . \tag{3.25}
\]
Finally, Equation 3.19 is used to compute the aggregate throughput of the node, where $\tau$ and $p$ are substituted by $\bar{\tau}$ (Equation 3.24) and $\bar{p}$ (Equation 3.25) respectively.

To evaluate probabilities $e$ and $b$ the node is assumed to be saturated, i.e., $e = 0$, and the maximum aggregate service rate of the node is evaluated. Under this condition, the value of $b|_{e=0}$ can be derived directly from Equation 3.21 and Equation 3.18 as

$$b|_{e=0} = f_B \frac{\bar{T}_s \bar{\tau} (1 - \bar{p}) + \bar{T}_c \bar{\tau} \bar{p} + \sigma (1 - \bar{\tau})}{(1 - \bar{\tau})(f_B \sigma + f_B \bar{T}_b)} .$$

(3.26)

Substituting this value into Equation 3.19, the maximum node throughput is obtained as

$$\mu_T = T_P|_{e=0} .$$

(3.27)

However, the computation of the maximum service rate is affected by the fact that packets end their service not only when they are sent successfully, but also when they reach the maximum number of retransmission attempts. The maximum rate $\mu_D$ at which packets are discarded by the MAC protocol due to the maximum retry limit is given by

$$\mu_D = \sum_{i=1}^{n_L} \mu_{D_i} ,$$

(3.28)

where

$$\mu_{D_i} = \frac{s_i \tau_i p_i^{m+1} (1 - p_i)}{\Delta (1 - p_i^m)} .$$

(3.29)
In the above equation, $\Delta$ is the average duration of a channel slot in the case $e = 0$. Finally, the maximum service rate of the node is calculated as

$$\mu = \mu_T + \mu_D \quad .$$

(3.30)

The queue behavior is modeled using a simple M/M/1/B model, where $B$ denotes the buffer capacity expressed in packets. The traffic intensity at the queue is $\rho = \Lambda / \mu$. The probability that the queue is empty is $\pi_0 = (1 - \rho)/(1 - \rho^{B+1})$. The aggregate node throughput can be computed as $T_P = (1 - \pi_0)\mu_T$, whereas individual link throughputs $T_{P_i}$ are given by

$$T_{P_i} = \frac{T_P s_i \tau_i (1 - p_i)}{\sum_{j=1}^{n_L} s_j \tau_j (1 - p_j)} \quad 1 \leq i \leq n_L \quad .$$

(3.31)

Standard equations of the M/M/1/B queue model allow the computation of the buffer overflow probability, the average number of packets in the queue, and the average queuing delay $\bar{Q}$. To compute the average queuing delay experienced by a packet destined to a particular link $i$, $Q_i$, first, the service rate on the link is computed as

$$\mu_i = \mu_{T_i} + \mu_{D_i} \quad ,$$

(3.32)

where

$$\mu_{T_i} = s_i \tau_i (1 - p_i)/\Delta \quad ,$$

(3.33)
and $\mu_D$, is given by Equation 3.29. Then, $Q_i$ is obtained by adding the average time spent in the queue before service (this component is common to all packets) to the specific service time on the outgoing link, yielding

$$Q_i = \bar{Q} - \frac{1}{\mu} + \frac{1}{\mu_i}.$$  \hspace{1cm} (3.34)

### 3.4.3 End-to-End Throughput Along a Route

As mentioned earlier in Section 3.3.2, the available bandwidth of a multi-hop flow over a path is the maximum throughput it can achieve subject to the condition that no queue along the path overflows. This estimation technique takes into account inter-flow and intra-flow interference separately and consists of two steps described below:

**Inter-flow step:** For each link $l$ in the path, I first find the maximum additional input rate $\epsilon_i$ that can be added under the constraint that no queue along the path overflows. The maximum allowable value of $\epsilon_i$ is found by iteratively applying the model to small increments of $\lambda_l$ (the packet arrival rate at link $l$). While doing so, I keep the measured values of $f_B$ and $p$ fixed relative to a node. However, when traffic is added on a link, the network state gets perturbed. Hence, the main approximation is to assume that $f_B$ and $p$ do not vary significantly when traffic is added on a single link, without saturating the link.
**Intra-flow step:** The quantities $e_l$ computed in the inter-flow step provide an estimate of the throughput that each link could obtain if operated in isolation. The same methodology, as presented previously in Section 3.3.2, is used to calculate the actual end-to-end estimate throughput for a route.

As was the case for the MLS metric, the route with the maximum estimated end-to-end throughput, among multiple routes between a source-destination pair, is selected. I call this metric the AVAIL metric since it depends upon the model based *available* throughput calculation.

I have so far presented the design of traffic-aware routing metrics only. However, in order to use any of these routing metrics, a routing protocol is required. I now describe the routing protocol.

### 3.5 The Routing Protocol

In order to make use of a routing metric (e.g., MLS and AVAIL, presented in Section 3.3 and Section 3.4, respectively), a routing protocol is needed that incorporates that routing metric. I choose a routing protocol that is a combination of a link-state protocol and a source routing protocol. I did not design an on-demand routing protocol because purely on-demand protocols do not search for newer (and possibly better) routes if a usable route, albeit not a good one, already exists. Hence, the routing protocol might get stuck with a route with lower throughput and/or higher delivery latency, even though a better route might have come into
existence since the current one was discovered. Each TAP in the network periodically floods to the entire network (each node in the network receives), the status of its links to its neighbors. This update contains, as appropriate for each of the metrics, the MLS value of the node, the loss probability on each of its links, the modulation rate for each of its links, and the input packet rate for each outgoing link of the node. For example, only the link loss probabilities are flooded for the ETX metric, whereas both the link loss probabilities and the MLS value of the node are contained in the flood if the routing protocol uses the MLS metric. On receipt of a sufficient number of link state updates, each node is able to construct a graph representing the topology of the network. The edge weights of this graph depend on the values contained in the advertisements. Each node then uses any standard algorithm to find a route from itself to any intended destination.

The exact algorithm used depends on the routing metric being used. For example, for metrics such as ETX, ETT, or IRU (Section 3.1.1), Dijkstra's shortest path algorithm can be used. However, for the metrics such as MLS and AVAIL, Dijkstra's shortest path might not be applicable since the values of these metrics do not increase monotonically with increasing path length, a requirement for the standard Dijkstra's shortest path algorithm. Instead, a slight variation of Dijkstra's shortest path algorithm can be used. Routes thus found are then used as source routes for outgoing packets. Routing loops are avoided since each node has a complete view
of the network and uses source routing, thus preventing intermediate nodes from making routing decisions on behalf of the source node.

Having presented the traffic-aware metrics, such as the MLS and the AVAIL metrics, and the traffic-unaware metrics, such as the ETX and the IRU metrics, and a routing protocol that can utilize them, I now present the evaluation of these metrics.

3.6 Evaluation

This section presents the evaluation of the metrics through simulation experiments. Simulation allows fast experimentation with reproduceable results and hence are convenient for evaluation. Simulation also allows the creation of networks with arbitrary numbers of nodes, thus allowing the testing of large scenarios that would be difficult to do in a real network deployment.

3.6.1 Simulation Setup

In this evaluation, I used the ns-2 network simulator with wireless extensions from the Rice University Monarch Project [61]. These extensions to ns-2 model channel contention, packet collision, channel capture, and backoff, in a manner similar to the specifications for IEEE 802.11.

Packet losses are measured by periodic MAC layer broadcast probe packets, each node broadcasting a probe packet every second. A MAC layer broadcast sent
out by a node can only be received by nodes which are within a single wireless hop of the broadcasting node. Whenever possible, packet loss statistics for unicast data packets are used to augment the statistics measured by broadcast probe packets.

For all experiments, the simulation used the IEEE 802.11 MAC protocol without RTS/CTS. Most WLANs operate with RTS/CTS turned off, and Bicket et al. [14] recently showed that the use of RTS/CTS was not effective at avoiding collisions over multihop routes. Additionally, due to the overhead of using RTS/CTS, the authors found a decrease in the throughput achieved over single hop routes. I used a data rate of 11 Mb/s, a transmission range of 150 m, and a node can cause interference at a maximum distance of 212 m. The simulation did not incorporate any power control mechanisms.

In an effort to understand fundamental properties of end-to-end throughput in wireless mesh networks, the evaluation was limited to the baseline case of single channel, single data rate, and UDP traffic. This is a necessary first step before considering more heterogeneous network scenarios and factors, such as Autorate Fallback and the role of transport protocols such as TCP.

For the purpose of this performance evaluation, the simulation did include all details of the routing protocol. The simulation emulated the behavior of a link state protocol using a centralized routing database, which is instantaneously updated whenever nodes provide new measurements. At any point in time in the simulation, the simulation considers the value of any parameter to be the average
of the samples collected for that parameter during the last 20 seconds. This interval of 20 seconds was chosen after experiments with several values. The values estimated by the model (Section 3.4) matched closely when the measurements were averaged over a period of 20 seconds; a higher value did not provide any further benefit.

When a source node S needs to send a packet to a destination node D, the simulation selects a route from the global database and puts the entire path in the packet header. Intermediate nodes relay packets only based on the source route carried in the packet.

3.6.2 Flow Throttling

After route selection, the flow has to be throttled so that it does not inject packets at a higher rate than the route can sustain. Irrespective of the routing metric used, I use the model presented earlier in Section 3.4 to predict the available throughput along the chosen route in order to rate-limit the newly added flow to the predicted throughput value. This rate-limiting is done for the sake of a fair comparison with ETX and IRU, since neither of these metrics can be used to compute the rate that a selected route can sustain.

The estimated throughput is used to throttle the flow at the TAP node originating the flow. Since the TAPs, unlike the actual end users of the flows, are controlled by the system, I consider it to be the responsibility of the TAP originating
the data packets to throttle the flow. In order to enforce rate-limiting, a TAP node can choose from a variety of methods such as per flow queues and per flow time slots.

- Per flow queues: Each TAP can keep separate queues for each flow originating at the TAP. Depending upon the route taken by each flow, the TAP can process packets from the queue for the flow at a rate that does not exceed the estimated throughput of the route. For example, if the selected route for a flow is able to handle at most 10 packets per second, the originating TAP would process packets from the queue for that flow at a rate of at most 10 packets per second.

- Per flow time slots: Instead of keeping per flow queues, each TAP can maintain a TDMA-style schedule for the flows originating at the TAP. In the previous example of a selected route for a flow being able to handle at most 10 packets per second, the originating TAP node would maintain a schedule that assigns at most one slot per second to the flow.

Given a route, the MLS metric does compute an estimated throughput for the route, but the results presented later in Section 3.6.4 show that the metric consistently over estimates the throughput sustainable by the route. Hence, I also present results in which the model is used to predict the available bandwidth along the route chosen according to MLS in order to rate-limit the newly added flow; this metric is represented as MLS-Throttle (MLS-T) in the result graphs.
3.6.3 Evaluation Topologies

I experimented with two different topologies: a topology from a real deployment of a residential high-speed wireless mesh network at Chaska, Minnesota, USA, and a synthetic grid-like, Manhattan topology.

I could use only the logical topology extracted from the Chaska deployment, and could not use the GPS coordinates of the TAPs, since I lacked details of the communication range and the maximum carrier sense range of the TAPs. The Chaska topology is comprised of 194 TAPs, of which I selected 14 to be gateway TAPs. In the actual deployment, 40 out of the 194 nodes (20% of total nodes) are gateway nodes. However, a topology with such a high percentage of gateways is inappropriate for evaluating and comparing routing metrics, since in such a network, routes hardly would be longer than a couple wireless hops. Consequently, all routing metrics, including the simple minimum hop count metric, would perform equally well in such networks. For any routing metric, the challenge of finding a suitable gateway exists only if gateways are not abundant. I manually selected the location of the gateways, trying to distributed the gateways over the entire network. In my judgement, only 14 gateways (7% of total nodes) were required to arrive at a topology in which each TAP node has at least one gateway, but possibly more, within a reasonable number of wireless hops, and the topology would still be able to differentiate between different routing metrics.
However, from the results presented later in this chapter, I concluded that the Chaska topology is severely limited by interference. As a result, results from the experiments with the Chaska topology do not convincingly illustrate the ability of the traffic-aware routing metrics to find routes with higher throughput than those found by traffic-unaware routes. Hence, I also considered a richer, structured topology, providing more spatial reuse and more degrees of freedom in the selection of routes. In particular, I considered a Manhattan network of 196 nodes arranged in a $14 \times 14$ grid. For easier comparison, the size of the grid was chosen to closely match the number of nodes in the Chaska deployment (194). Each node in the Manhattan topology is within sensing and transmission range of its north, south, east, and west neighbors, and is within sensing range of its northeast, northwest, southeast, and southwest neighbors. I randomly selected 10 of these nodes to act as gateways. I selected the gateways randomly because in a real deployment, the flexibility of placing gateways might be limited since gateway nodes typically require a dedicated link to the wired backbone. Here again, as in the case of the Chaska topology, 10 gateways seemed sufficient to reasonably cover the entire network, but still make it challenging for routing metrics to select and route to gateway nodes.
3.6.4 A Single Additional Flow — Chaska Topology

In the first part of my evaluation, I considered the Chaska topology. The topology is shown in Figure 3.6. Each node in the simulation used an omni-directional antenna.

I randomly picked 100 nodes and for each started an upstream flow to its nearest (according to wireless hop count) gateway. An upstream flow is defined as a flow from a TAP to a gateway TAP. The traffic load on each gateway is uniformly distributed between 30% and 100% of a maximum gateway load, and this load is equally distributed among the flows ending at the same gateway. By doing so, I randomized the load on the 14 gateways present in the network. For example, the maximum gateway load is set to 2 Mbps, and a particular gateway is loaded with 50% of the load, i.e., the total load on the gateway from all flows ending at the gateway is $2 \times \frac{1}{2} = 1$ Mbps; if there are five flows ending at that gateway, then each flow gets $\frac{1}{5} \times 1 = 0.2$ Mbps, since each flow gets an equal share.

I ran two sets of experiments for two values of this maximum gateway load, namely, 1.5 Mbps and 2 Mbps. After allowing the 100 flows to run for 30 seconds, I randomly chose another node, other than the 100 existing sources, and started an upstream flow from this node, which has the freedom to choose its destination gateway. For the new flow, I computed the best route to a gateway according to the different metrics, AVAIL, MLS, MLS-T, IRU, and ETX. Once the route is computed, the flow uses the same route for the rest of the simulation.
Figure 3.6: Topology of the Chaska wireless mesh network. There are 195 nodes in the network, 14 of which are gateway nodes, marked as squares in the figure. Nodes within transmission range of each other are connected by a solid line. Nodes with a dashed line between them wirelessly interfere with each other.
Moreover, upon selection of a route, the source is throttled to send at the estimated available throughput across the route. The traffic-aware metrics, except MLS-T, give an estimate of the available throughput. As mentioned earlier in Section 3.6.2, the MLS-T metric is used only to select routes, but the actual estimated throughput for the selected route is computed using the model presented in Section 3.4. Thus, the MLS-T metric selects exactly the same routes as does the MLS metric. Similarly for the traffic-unaware metrics, after the route has been selected according to the ETX and IRU metrics, the model is used to compute the estimated available throughput along the route.

For each metric, I repeated the experiment 50 times, each experiment having the same initial 100 flows but a different source node from which to start the new flow to be added in the network. I measured the throughput achieved by the new flow in each of these 50 experiments.

I first ran these experiments with a maximum load on each gateway of 1.5 Mbps. Figure 3.7(a) shows the throughput achieved by the additional flows, sorted in ascending order. All of the metrics perform very similarly, although MLS metric shows a slight benefit over the other metrics. However, the MLS metric also incurs the largest delay. Figure 3.7(b) shows the sorted delays incurred by the additional flows in these experiments. This high delay occurs because the MLS metric overestimates the available throughput, causing the queues to buildup at intermediate nodes. The other metrics all lead to similar delays. Figure 3.7(c) shows the route
Figure 3.7: Results for experiments on the Chaska topology. The maximum load on any gateway is 1.5 Mbps, and each node uses a single omni-directional antenna.
lengths of the additional flows. Compared to the other metrics, the AVAIL metric routes a few flows along longer routes, since this metric is able to route flows around areas of high wireless contention.

Since there was not much difference between the different metrics, I first suspected that the load on the network might be so low as to leave enough spare capacity in the network for the routing metric used for selecting routes to be irrelevant. In order to test whether the network was indeed this lightly loaded, I increased the load on the network by increasing the maximum gateway load to 2 Mbps. The achieved throughput is shown in Figure 3.8(a). Even with the increase in the offered load, the findings are very similar to those with the maximum gateway load set to 1.5 Mbps — there is very little difference between the different metrics; the MLS metric does give a slightly higher throughput, but this is because the MLS metric overestimates the available throughput along a route. This overestimation is reflected in the delay incurred by the additional flows, shown in Figure 3.8(b). The delay incurred when using the MLS metric is higher than the delays incurred when using any of the other metrics, and the delays incurred when using the other metrics are very similar to each other. However, with increased load, the choice of a routing metric does lead to different route lengths, as shown in Figure 3.8(c). When using metrics that take existing traffic into account, i.e., AVAIL, MLS, and MLS-T, flows are in general routed over longer hops in an
Figure 3.8: Results for experiments on the Chaska topology. The maximum load on any gateway is 2.0 Mbps, and each node uses a single omni-directional antenna.
attempt to route around areas of the network with high wireless contention, than when using metrics that do not take existing traffic into account, i.e., ETX and IRU.

The results presented to this point suggest that the topology is limited by interference. In order to verify that, I chose to use a directional antenna in an attempt to check if the lower interference caused by a directional antenna (compared to that caused by an omni directional antenna using the same transmission power) could alleviate the situation. I now describe the directional antenna model.

3.6.5 Directional Antenna Model

In the directional antenna model I used, the mathematical model for determining the gain at an angle of departure of \( \theta \) radians from the primary antenna direction is given by

\[
\text{gain}(\theta) = 0.25 \cdot (1 + \cos(\theta))^2 \cdot \left( \frac{\sin(\alpha \sin(\theta))}{\alpha \sin(\theta)^2} \right)^2, \tag{3.35}
\]

where \( \alpha \) is a parameter that controls the beamwidth of the directional antenna pattern. For example, a value of \( \alpha = 1.55 \) was used in the simulations to generate a pattern with a 3 dB beamwidth of 30°.

Figure 3.9 shows the normalized directional antenna pattern used for the antennas at the TAPs in these simulations. This pattern, based on the model of a rectangular aperture antenna [10], is an approximation of typical directional antennas in use for cellular systems. The directional antenna at the TAP has a 3 dB
beamwidth of 30°, which, as explained in the previous example, was generated by setting the value of \( \alpha \) to 1.55.

3.6.6 Results with Directional Antennas — Chaska Topology

As with the omni directional antenna, I first tried with the maximum gateway load set to 1.5 Mbps. These results are shown in Figure 3.10. In terms of throughput (Figure 3.10(a)), all of the metrics perform very similarly, even with a directional antenna being used for packet transmissions. This suggests that using a directional antenna does not reduce the inherent interference in the network. Interestingly, even though there is not much difference in the throughput achieved by using the different metrics, the delay incurred by the flows does vary between the different
metrics, as shown in Figure 3.10(b). This is because, as shown in the Figure 3.10(c), when using the metrics that take existing traffic into account, i.e., AVAIL, MLS, and MLS-T, around half the flows use a longer route than when using ETX or IRU. The performance with the maximum gateway load increased to 2 Mbps is very similar and is shown in Figures 3.11. The reason for the similar throughput is that even though directional antennas do help in decreasing the interference in the network, directional antennas do not decrease it enough to increase the capacity in the network.

3.6.7 Overall Results for the Chaska Topology

Figure 3.12 shows the average (over the 50 additional flows in 50 different experiments) achieved in the experiments for the Chaska topology; Figures 3.12(a) shows the results for maximum gateway load of 1.5 Mbps, and Figure 3.12(b) shows the same for a maximum gateway load of 2 Mbps. Even though these results show that the metrics that take existing traffic into account provide better throughput, the gain from using such metrics is not substantial. With the maximum gateway load set to 1.5 Mbps, the metrics perform very similar to each other. However, when the offered load on the network is increased by increasing the maximum gateway load to 2 Mbps, the use of traffic-aware metrics does increase the average throughput achieved by the additional flows. As shown in Figure 3.12(b), when using the AVAIL metric with omni-directional antennas, the flows get around 8%
Figure 3.10: Results for experiments on the Chaska topology. The maximum load on any gateway is 1.5 Mbps, and each node uses a single directional antenna.
Figure 3.11: Results for experiments on the Chaska topology. The maximum load on any gateway is 2.0 Mbps, and each node uses a single directional antenna.
Figure 3.12: Average throughput achieved in experiments on the Chaska topology. At lower load, all the metrics perform very similar to each other. At a higher load, there are some performance benefits, albeit marginal. This improvement at higher load is magnified when a directional antenna is used.

higher average throughput then when using the ETX or IRU metrics; the flows get roughly the same throughput between the ETX and IRU metrics. When using the MLS-T metric, the flows get around 3% higher average throughput than when using the ETX or IRU metrics. With the use of directional antennas, however, the traffic-aware metrics lead to higher increases in average throughput for the additional flows, compared to the traffic-unaware metrics. Compared to the average throughput achieved when using the ETX or IRU metrics, the use of the AVAIL metric gives 12% higher throughput, whereas the use of the MLS-T metric gives 8% higher throughput.
However, since a directional antenna helps to decrease the interference in the network, the effect of the directional antenna is less pronounced at the lower load (with the maximum gateway load set to 1.5 Mbps) than at the higher load (with the maximum gateway load set to 2 Mbps). As shown in Figure 3.12(a), there is very little change in the average throughput when a directional antenna is used, since the network is unsaturated at this load, and hence the decrease in interference owing to the use of directional antennas does not provide any further gain. There is a slight decrease when directional antennas are used in conjunction with either the MLS or the MLS-T metric. This decrease is because the MLS metric overestimates the available throughput along a route and hence allow the flows to inject more packets than the route can handle. This leads to lower achieved throughput, since intermediate nodes in the route compete with each other. As Figure 3.12(b) shows, when the load in the network is increased, the improvement with directional antennas is more pronounced. On average, the throughput performance of all the metrics are increased by around 15%. However, the performance of the MLS metric actually decreases by around 6% when a directional antenna is used. Increasing the load in the network leads to higher interference, and directional antennas help to reduce interference, thus allowing for higher average achieved throughput.

The hypothesis that sufficiently loading the network would differentiate the routing metrics, did not turn out to be correct, at least in the Chaska topology, suggesting that maybe the Chaska topology is severely limited by interference. That
the topology is limited by interference is indeed corroborated by the fact that, in order to provide more throughput to the access points, the actual Chaska deployment uses 40 gateways (I used only 14). Moreover, with 14 gateways, there is little freedom in the selection of gateways, because the network is composed of almost disconnected islands connected by a small number of bridges, yielding only a few independent paths that reach distant gateways. However, if more gateways are used in order to increase the throughput to each access point, as is done in the real Chaska deployment, then routing becomes less relevant, as most access points would have a gateway within one or two wireless hops.

Hence, from my experiments, I concluded that the Chaska topology was severely limited by interference, so that all possible routes from a node to a gateway get low throughput. In fact, the Chaska topology is so severely limited by interference that even using a directional antenna does not provide much gain.

At this point, it is not clear whether the traffic-aware metrics that I have proposed are effective in finding routes that would give higher throughput than routes chosen by traffic-unaware metrics. Hence, I chose a more idealistic network and experimented the metrics on such a network.

3.6.8 A Single Additional Flow — Manhattan Topology

As with the experiments using the Chaska topology (Section 3.6.4), the load on each gateway is uniformly distributed between 30% and 100% of a maximum gate-
way load, and this load is equally distributed among the flows ending at the gateway. As before, there are 100 existing flows. Also, once a route has been selected, the route is not changed, and the source throttles the flow to the estimated available throughput along the route (Section 3.6.2). I ran three sets of experiments for three values of this maximum gateway load, namely, 2 Mbps, 3 Mbps, and 4 Mbps. The maximum gateway loads are different from the ones used in the experiments with the Chaska topology (1.5 Mbps and 2 Mbps) because the Chaska topology was severely limited by interference, and hence the topology could not support high traffic load; the Manhattan topology is more structured and is able to support higher network loads.

Figure 3.14 shows the results of the experiments with omni-directional antennas in the Manhattan topology. Even with low load on the network, the metrics that take existing traffic into account provide higher throughputs than with the ETX or IRU metrics (Figure 3.14(a)). However, as with the Chaska topology, the MLS metric overestimates the available throughput along a route, and hence the delay incurred when using the MLS metric is higher than when using the other metrics (Figure 3.14(b)). As shown in Figure 3.14(c), flows using AVAIL, MLS, or MLS-T metric are routed over longer routes than flows using ETX or IRU.

Upon increasing the offered load on the network by increasing the maximum gateway load to 3 Mbps, the differences between the different metrics become clearer. These results are shown in Figure 3.15. As shown in Figure 3.15(a), the
Figure 3.13: Topology of the Manhattan wireless mesh network. There are 196 nodes in the network, 10 of which are gateway nodes, marked as squares in the figure. Nodes within transmission range of each other are connected by a solid line. Nodes with a dashed line between them interfere with each other.
traffic-aware metrics outperform the traffic-unaware metrics. Flows routed with
the MLS metric get the highest throughput, but this is mainly because the MLS
metric overestimates the available throughput along a route (leading to higher
packet delivery latency, as shown in Figure 3.15(b)). In general, the AVAIL metric is
able to find better routes than the MLS-T metric finds, because the AVAIL metric is
more detailed than is the MLS metric (for example, the AVAIL metric takes queue
sizes into account). The MLS-based metrics (MLS and MLS-T) are able to get bet-
ter throughputs even though, as shown in Figure 3.15(c), the lengths of the routes
found by these metrics are not considerably different from the lengths found by
ETX or IRU. In other words, between two routes of equal length, the MLS-based
metrics are able to choose the route with the higher estimated throughput. The
AVAIL metric finds longer routes than those found by all the other metrics.

Figure 3.16 shows the results upon further increasing the offered load on the
network by increasing the maximum gateway load to 4 Mbps. As shown in Fig-
ure 3.16(a), the performance of the metrics become even more separated. Many
flows get starved when the ETX or IRU metrics are used, whereas fewer flows get
starved when the MLS-T metric is used, and still fewer flows get starved when the
AVAIL metric is used. When the ETX or IRU metrics are used, the delay incurred
by the flows is higher than the delay incurred when the AVAIL or the MLS-T met-
ric is used (Figure 3.16(b)). Since the ETX and IRU metrics are not able to avoid less
congested routes, more packet collisions occur in the network. Similar to the pre-
vious scenario in which the maximum gateway load was set to 3 Mbps, the lengths of the routes found by the MLS-based metrics are very similar to those found by the ETX or IRU metrics, but are considerably shorter than the routes found by the AVAIL metric, as shown in Figure 3.16(c).

The important result from this set of experiments is that traffic-aware routing metrics are able to find routes with considerably higher throughputs as compared to traffic-unaware routing metrics. Even a simple traffic-aware model such as MLS is able to find much better routes than routes found with ETX or IRU. Since the delay characteristics of the MLS metric are worse than of the other metrics, at first glance it might seem that the MLS metric cannot be used independent of the model. However, the model is used to find the maximum sustainable throughput of a route, not to select the route. Any other congestion control mechanism, such as TCP, would be able to exploit the higher sustainable throughput of routes chosen according to the MLS metric. Since, across all scenarios, the AVAIL metric finds longer routes than does the MLS-T metric, especially when using an omni-directional antenna, without always providing substantial improvement over the MLS-T metric, the use of the model-based AVAIL metric might not be justified, given the complexity of implementing the model and the wireless resource usage required.
Figure 3.14: Results for experiments on the Manhattan topology. The maximum load on any gateway is 2.0 Mbps, and each node uses a single omni-directional antenna.
Figure 3.15: Results for experiments on the Manhattan topology. The maximum load on any gateway is 3.0 Mbps, and each node uses a single omni-directional antenna.
Figure 3.16: Results for experiments on the Manhattan topology. The maximum load on any gateway is 4.0 Mbps, and each node uses a single omni-directional antenna.
Figure 3.17: Results for experiments on the Manhattan topology. The maximum load on any gateway is 2.0 Mbps, and each node uses a single directional antenna.
Figure 3.18: Results for experiments on the Manhattan topology. The maximum load on any gateway is 3.0 Mbps, and each node uses a single directional antenna.
Figure 3.19: Results for experiments on the Manhattan topology. The maximum load on any gateway is 4.0 Mbps, and each node uses a single directional antenna.
3.6.9 Results with Directional Antennas — Manhattan Topology

As with the Chaska topology, I repeated the experiments for the Manhattan scenario with directional antennas instead of omni-directional ones. The results are shown in Figure 3.17. As in the case of omni-directional antennas, when the maximum gateway load was set to 2 Mbps, all of the metrics performed similarly, as shown in Figure 3.17(a). This similarity in performance is because the routing metric is irrelevant when the network is lightly loaded. The delay incurred when using the AVAIL metric is higher than when using the MLS-T metric, and both delays are higher than incurred when using the ETX or IRU metrics (Figure 3.17(b)). This difference in delay is because the traffic-aware metrics choose longer routes, with the AVAIL metric choosing routes even longer than those chosen with the MLS-based metrics, as shown in Figure 3.17(c).

Figure 3.18 shows the results of the experiments with the maximum gateway load set to 3 Mbps. Once the load is increased, the throughput performance with the traffic-aware metrics is much better than with the traffic-unaware metrics (Figure 3.18(a)). Moreover, all of the traffic-aware metrics perform similarly, with flows incurring higher delays with the MLS metric, as shown in Figure 3.18(b). As with the maximum gateway load of 2 Mbps, the traffic-aware metric finds routes that are longer but have higher available bandwidth, as is shown in Figure 3.18(c). Among the traffic-aware metrics, the AVAIL metric chooses longer routes than the MLS metric, thus increasing the delay incurred by the packets.
Figure 3.19 shows the results of the experiments when the maximum gateway load is further increased to 4 Mbps. At this load, the AVAIL metric outperforms the MLS-T metric, and all of the traffic-aware metrics provide much higher throughput than do the traffic-unaware metrics (Figure 3.19(a)). Even the delay incurred by the use of the AVAIL and MLS-T metrics are lower than that incurred by using the ETX and IRU metrics (Figure 3.19(b)), because the traffic-unaware metrics select routes which are shorter than the routes selected by the traffic-aware metrics (Figure 3.19(c)), but face higher contention from existing flows.

Figure 3.20 shows the average (over the additional flows) achieved in the experiments for the Manhattan topology; Figures 3.20(a) shows the results for maximum gateway load of 2 Mbps, Figure 3.20(b) shows the same for a maximum gateway load of 3 Mbps, and Figure 3.20(c) shows the same for a maximum gateway load of 4 Mbps. The use of directional antennas substantially improves the average throughput of the additional flows. When the maximum load on a single gateway is limited to 2 Mbps, the use of directional antennas gives as much as 28% average throughput improvement, for the traffic-unaware metrics (IRU and ETX), and as much as 25% improvement for the traffic-aware metrics (AVAIL and MLS-T), over the average throughput achieved using omni-directional antennas (Figures 3.20(a)). When the maximum gateway load is increased to 3 Mbps, the improvement with directional antennas increases to around 45% for the traffic-unaware metrics and around 50% for the traffic-aware metrics (Figures 3.20(b)).
(a) With 2 Mbps maximum load on any single gateway.

(b) With 3 Mbps maximum load on any single gateway

(c) With 4 Mbps maximum load in any single gateway

Figure 3.20: Average throughput achieved in experiments on the Manhattan topology. Across all offered load, the use of traffic-aware metrics gives higher average throughput than the user of traffic-unaware metrics. The use of the ETX or IRU metrics gives very similar average throughput.
By using directional antennas, the percentage increase in throughput over using omni-directional antennas increases as the offered traffic load in the network is increased, from a maximum gateway load of 2 Mbps to a maximum gateway load of 4 Mbps. However, at 4 Mbps maximum gateway load, the network is overloaded, causing high wireless contention in the network. Hence, the values of the average throughput achieved decreases substantially. For example, with omni-directional antennas and the maximum gateway load set to 4 Mbps, the average flow throughput when using the AVAIL metric is 0.3 Mbps (Figure 3.20(c)), compared to 0.8 Mbps when the maximum gateway load is set to 2 Mbps (Figure 3.20(a)).

3.6.10 Overall Results for the Manhattan Topology

Across all offered load, the use of traffic-aware metrics gives higher average throughput than with the use of traffic-unaware metrics. With the maximum gateway load set to 2.0 Mbps, shown in Figure 3.20(a), the use of the AVAIL metric improves the average throughput by 9%, and the use of the MLS-T metric improves the average throughput by 8%, compared to the average throughput obtained using the ETX or IRU metrics, which perform very similar to each other. With the use of directional antennas, these improvements are 10% using the AVAIL metric and 11% using the MLS-T metric. As shown in Figure 3.20(b), when the network load is further increased by increasing the maximum gateway load to 3 Mbps, the use of
the AVAIL metric improves the average throughput by 55%, and by 54% when directional antennas are used. Under this higher network load, the use of the MLS-T metric gives corresponding improvements of 41% with omni-directional antennas and 50% with directional antennas. With even higher load, shown in Figure 3.20(c), the network gets overloaded and although the use of traffic-aware metrics still improves average throughput, the average throughput achieved is very low.

3.6.11 Multiple Additional Flows — Manhattan Topology

As in the experiments with a single additional flow (Section 3.6.4 and Section 3.6.8), I randomly picked 100 nodes and for each started an upstream flow to its nearest (according to wireless hop count) gateway. The load on each gateway is uniformly distributed between 30% and 100% of a maximum gateway load, and this load is equally distributed among the flows ending at the same gateway. I ran two sets of experiments for two values of this maximum gateway load, namely, 1.5 Mbps and 2 Mbps. After allowing these 100 flows to run for 30 seconds, I added 50 additional flows, one flow every 5 seconds. The source of each of these flows was randomly chosen but was different from the 100 existing flows. Each of these additional flows continues for 50 seconds, and thus there are at most 10 additional flows at any time. I measured the throughput achieved by each of these 50 additional flows. The source node for each of this additional flows determines the best route for each packet that it generates, subject to the following constraints:
• A route to a destination is computed only if the existing route to the destination was computed at least one second ago. This separation ensures that computation overhead is bounded to one route calculation per second per destination.

• Once a new route is computed, that route is used only if the metric value of the route is better than the metric value of the existing route, if any, by at least 10%. Unless there is some threshold for preferring a newer route over an older one, the system might become unstable as routes start to flap. The 10% threshold for selecting a new route over an existing one allows for route hysteresis, which prevents rapid flapping among multiple routes to the same destination. In the particular case of routing with traffic-aware metrics, route hysteresis makes the process of route selection robust against bursty or transient changes in the available throughputs of probable routes.

As in the experiments with a single additional flow, a newly added flow is throttled (Section 3.6.2) at the TAP node originating the flow. However, the newly added flow in this case is throttled to half the estimated throughput along the selected route, because if a single flow sends at exactly the remaining estimated capacity of the selected route, then future flows might get starved. This experiment is repeated for each of the routing metrics — AVAIL, MLS, MLS-T, IRU, and ETX.

Figure 3.21 shows the results of the experiments done with three different values for the maximum gateway load, 2 Mbps, 3 Mbps, and 4 Mbps. Figure 3.21(a)
Figure 3.21: Results for experiments with multiple simultaneous flows in the Manhattan topology. Each node uses a single omni-directional antenna.
shows that when the maximum gateway load is set to 2 Mbps, the MLS metric gives the highest throughput, but as before, this performance is mainly because it overestimates the available throughput along a route. This is verified by the build up of the queues in the intermediate nodes. Flows using the traffic-aware metrics consistently find routes with higher throughput than the routes found by the traffic-unaware metrics. Among the traffic-aware metrics, the AVAIL metric finds better routes than does the MLS-T metric. As explained earlier, the better performance of the AVAIL metric is because the model that the AVAIL metric is based upon is more detailed and takes several parameters, such as network interface queue length, into account. As the offered load is increased by increasing the maximum gateway load to 3 Mbps, as shown in Figure 3.21(b), the difference between the throughputs achieved on routes selected by the traffic-aware metrics and by traffic-unaware metric increases. As the load is increased even more to the maximum gateway load of 4 Mbps, the difference becomes even more pronounced, as shown in Figure 3.21(c).

I replaced the omni-directional antennas with directional ones and repeated the experiments. These results are shown in Figure 3.22. Since the chances of finding routes that do not interfere with each other or slightly interfere with each other increases with the use of directional antennas, as shown in Figure 3.22(a), the traffic-aware routing metrics find higher throughput routes, even with the maximum gateway load set to 2 Mbps. As the value of the maximum gateway load is
(a) With 2 Mbps maximum load on any single gateway

(b) With 3 Mbps maximum load on any single gateway

(c) With 4 Mbps maximum load on any single gateway

Figure 3.22: Results for experiments with multiple simultaneous flows in the Manhattan topology. Each node uses a single directional antenna.
increased to 3 Mbps (Figure 3.22(b)), the throughput achieved by the traffic-aware routing metrics stays consistently higher than is achieved along routes chosen by the traffic-unaware metrics. Finally, when the maximum is set to 4 Mbps (Figure 3.22(c), several flows are starved when the traffic-unaware metrics are used. The number of starved flows reduces when the MLS-T metric is used and further reduces when the AVAIL metric is used.

Figure 3.23 shows the average (over all 50 additional flows) achieved in the experiments for the Manhattan topology; For all network loads, the throughput achieved by using the traffic-aware metrics is higher than is achieved by using the traffic-unaware metrics. The trend followed is very similar to that in the results for the experiment with a single additional flow, presented earlier in Figure 3.20. However, the average throughput achieved are substantially lower. This decrease is because in these experiments with multiple flows (Figure 3.23), there are up to 10 simultaneous additional flows, compared to a single additional flow in the experiments for a single additional flow (Figure 3.20). More flows cause higher contention among the flows, and hence the average throughput obtained by the flows is lower. For all the metrics, though, the throughput achieved increases with the use of directional antennas.
Figure 3.23: Average cumulative throughput achieved in experiments on the Manhattan topology. The results are similar in trend to the results presented in Figure 3.20. However, the absolute numbers are much less since in these experiments there are up to 10 additional flows, compared to a single additional flow in the experiment corresponding to Figure 3.20.
3.6.12 Shortlived Flows — Manhattan Topology

In this section, I present the performance of the different routing metrics when shortlived flows arrive in the Manhattan scenario with existing flows. As for the earlier experiments with multiple flows for the Manhattan scenario presented in Section 3.6.11, there are 100 existing flows in the network. The load on each gateway is uniformly distributed between 30% and 100% of a maximum gateway load of 3 Mbps, and this load is equally distributed among the flows ending at the same gateway. After allowing the 100 flows to run for 30 seconds, I added 50 additional flows. Each of these additional flows stays alive for 10 seconds, which is less than the 20-second interval used for calculating the value of any parameter used by the routing metrics.

Figure 3.24 shows the average throughput achieved by the additional flows in experiments performed with both omni-directional and directional antennas and for three different values for the flow arrival rate: one flow every 1 second, 2 seconds, and 5 seconds. For both omni-directional antennas (Figure 3.24(a)) and directional antennas (Figure 3.24(b)), as the rate of flow arrival decreases, the difference between the traffic-aware metrics and the traffic-unaware metrics increases. For example, with a new flow arriving every one second, and with the use of directional antennas, the AVAIL metric provides only around 10% higher average throughput than does either the IRU or the ETX metric. This percentage increases to around 35% when the flow arrival rate is reduced to one flow every 2 seconds.
Figure 3.24: Performance with multiple shortlived flows in the Manhattan topology. The maximum load on a single gateway is 3 Mbps, and each shortlived flow is alive for a duration of 10 s.

This increase is because with fewer simultaneous flows in the system, there is more scope for the traffic-aware routing metrics to select routes with higher estimated throughput.

The performance trend is also consistent across all flow arrival rates and antenna types. Moreover, as the flow arrival rate decreases, the average throughput increases because with decreasing arrival rate, the number of additional flows simultaneously active in the system decreases. Consequently, there is less contention in the network and each additional flow gets a higher individual share of the available bandwidth. Since the resource consumed by each newly arrived flow...
takes some time to affect the average values computed over a 20 s interval (used by the routing metrics, as described in Section 3.6.1), if the next flow arrives too early, then the routing decision made for this new flow is suboptimal to start with. If the arrival and departure rate of flows is very high, then the average values used by the routing metrics are unable to capture the current state of the network, thus leading to suboptimal routing decisions for the entire lifetime of the flow.

Also, comparing Figure 3.24(b) to Figure 3.24(a), not only is the average throughput achieved when using directional antennas higher than that achieved when using omni-directional antennas, the performance improvement obtained by using the traffic-aware metrics is more pronounced when directional antennas are used; this is consistent with the results presented earlier in the thesis.

3.6.13 Multiple Simultaneous Flows — Manhattan Topology

In the final experiment on the Manhattan topology, I simultaneously introduced five flows after allowing the 100 existing flows to run for 30 seconds. The parameters used for the existing flows are exactly the same as the previous experiment for shortlived flows described in Section 3.6.12. Each of the five additional flows stays alive for 50 second at the end of which the experiment ends.

Figure 3.25 shows the average and the standard deviation of the throughputs achieved by the five simultaneously introduced flows when using the different routing metrics and different antenna types. The standard deviation in the through-
Figure 3.25: Performance with multiple simultaneous flows in the Manhattan topology. The maximum load on a single gateway is 3 Mbps.

put achieved by these five additional flows when the AVAIL or the MLS-T metric is used is less than that when the IRU or ETX metric is used.

With omni-directional antennas (Figure 3.25(a)), the average throughput achieved by the use of the AVAIL metric is around 30% higher than the average throughput achieved by the use of the ETX or IRU metric. The average throughput achieved by the MLS-T metric is around 13% higher than that achieved by the traffic-unaware metrics.

However, with the use of directional antennas (Figure 3.25(b)), the performance benefits increase. All the metrics perform better than when using omni-directional antennas, and the difference in performance between the traffic-aware metrics and the traffic-unaware metrics is more prominent when directional antennas are used.
The use of the AVAIL metric leads to around 31% higher average throughput than that obtained by using the traffic-unaware metrics, whereas the MLS-T metric achieves around 26% higher average throughput.

3.7 A Comparison of the Two Topologies

In order to better understand the quantitative differences between the Chaska and the Manhattan topologies, I calculated the distribution of neighbors and interferers for both the topologies. Two nodes are considered neighbors of each other if in the absence of any interference, the two nodes can successfully communicate with each other. Two nodes are considered to be interferers of each other if first, they are not neighbors of each other, and second, according to the rules of the MAC protocol being used, the two nodes cannot transmit simultaneously because of the possibility of interference between them.

Figure 3.26 shows the cumulative density plots of the neighbors and interferers for both the Chaska and the Manhattan topologies. For example, in the Chaska topology (Figure 3.26(a)), slightly less than 60% of nodes have three or fewer neighbors, whereas more than 70% of nodes have have three or fewer interferes. Since the Chaska topology is a connected topology, i.e., all nodes are reachable from all other nodes, there are no nodes that do not have any neighbors. In contrast, there are around 15% of nodes that do not have any interfering nodes. Such a situation arises when the topology consists of chains in which consecutive nodes are within
transmission range of each other but the neighbor-of-the-neighbor of a node is not within transmission range. Moreover, the cumulative plot for interferers is distinctly above the plot for the neighbors. This indicates that most nodes have a higher number of interferers than neighbors. In contrast, the Manhattan topology (Figure 3.26(b)) has only a few nodes that have a higher number of interferers than neighbors. This similarity between the number of neighbors and interferers for the Manhattan topology is mainly because of the regular structure of the topology. Another difference between the two topologies is that for the Chaska topology, the maximum number of neighbors or interferers is 7, whereas for the Manhattan topology, the maximum number of neighbors or interferers is just 4. This differ-
Figure 3.27: Neighbors and interferers statistics for the two topologies. The Chaska topology has a wider range of values and also has a higher standard deviation that the Manhattan topology.

Figure 3.27 shows various statistical parameters for the two topologies. Compared to the Manhattan topology, the Chaska topology has a lower mean and median for the number of neighbors and interferers but has a higher standard deviation. This difference means that for the Chaska topology, certain parts of the network have significantly higher interference than other parts of the network. To illustrate this, Figure 3.28 shows contour plots of the number of neighbors and interferers in the Chaska topology, and Figure 3.29 shows the same for the Manhattan topology. As an example, in the Chaska topology, nodes in the area around the (x,y) coordinates (1500,1100) has significantly more neighbors and interferers (up
to 7) compared to nodes in the area around the coordinates (1800,3000) (as low as 0). However, for the Manhattan topology, the lower standard deviation shows that the interference characteristics is similar throughout the network. In the Manhattan topology, only nodes that lie along the edges of the topology have a different (and smaller) number of neighbors and interferers compared to the interior nodes. Indeed, it is due to these nodes at the edges that the standard deviation for the Manhattan topology, shown in Figure 3.27, is not 0.

Finally, I analyzed how the average number of gateways changes with increasing number of wireless hops. The result is shown in Figure 3.30. Even though any particular gateway might be accessible within, say, $k$ hops, the length of the selected route might be greater than $k$ when a routing metric other than the minimum hop count metric is used to select a route to that gateway. Except for gateways that are accessible within 1 wireless hop (for which a routing metric would not matter much), the Manhattan topology has a significantly more gateways accessible than in the Chaska topology. The gap is especially wide at distances of 4 and 5 wireless hops. This means that in the Manhattan topology, every routing metric has the ability to pick from a larger pool of gateways than in the Chaska topology. This is in spite of the fact that both topologies have almost the same number of nodes (196 nodes in the Manhattan topology versus 197 nodes in the Chaska topology) and that the number of gateways in the Chaska topology (14) is greater than the number number of gateways in the Manhattan topology (10).
Figure 3.28: The contour plots the neighbors and interfering nodes for the Chaska topology.
Figure 3.29: The contour plots the neighbors and interfering nodes for the Manhattan topology.
Figure 3.30: The average number of gateways accessible at varying wireless hop counts for the Chaska and the Manhattan topologies.

3.8 Chapter Summary

In this chapter, I have presented the design and analysis of three traffic-aware routing metrics: the MLS metric, the AVAIL metric, and the MLS-T metric. The MLS metric is based on the MAC layer share of nodes, which is the fraction of time that is free at the node. Along with link modulation rates and link loss probabilities, the MAC layer share is used to estimate the throughput across any link. The AVAIL metric is based on a theoretical model for IEEE 802.11-based MAC protocols. Given certain measured parameters, such as link modulation rate, link loss probabilities, packet arrival rate, and interface queue length, the model estimates the throughput across any given link. The estimated throughput of all links, computed either by measurement, or by the model, in a route are combined to estimate
the end-to-end throughput across the entire route. Since the results showed that the MLS metric consistently overestimated the throughput along a route, I used the MLS metric to select routes, but having selected a route, I used the AVAIL metric to estimate the throughput across the entire route; this modified metric is called the MLS-T metric. Since the traffic-unaware metrics, i.e., the ETX and IRU metrics, do not estimate throughputs across selected routes, for fair comparison, these metrics also use the model to estimate the throughput across selected routes. After a route has been selected, the source throttles the flow according to the expected achievable throughput.

I have evaluated different traffic-aware and traffic-unaware metrics, and under different network scenarios — a topology from a real residential high-speed wireless mesh network, called the Chaska topology, and a synthetic grid-like, Manhattan topology.

I found that the Chaska topology is severely limited by interference. The traffic-aware metrics showed an improvement of up to 8% when omni-directional antennas were used, and an improvement of up to 12% when directional antennas were used. Across all metrics, the use of directional antennas improved the average achieved throughput by only 15%, compared to the average throughput achieved when using omni-directional antennas.

However, in the more structured Manhattan topology, the use of the traffic-aware metrics, i.e., MLS and AVAIL, led to the selection of routes that provided
higher throughputs than did routes selected with the use of traffic-unaware metrics, such as ETX and IRU. Compared to the average throughput obtained by using the traffic-unaware metrics with a single additional flow in the network, the use of traffic-aware metrics improved the average throughput by up to 40%, when omni-directional antennas were used, and up to 50%, when directional antennas were used. Even with multiple additional flows, the improvement in average throughput provided by the use of traffic-aware metrics are similar in trend to the improvements achieved for a single additional flow.

In the structured topology, the use of directional antennas provided higher average throughput than did the use of omni-directional antennas; this improvement was consistent across all of the evaluated routing metrics, both the traffic-aware and traffic-unaware ones. This improvement was 28%–45% for the traffic-unaware metrics and 25%–50% for the traffic-aware metrics. Also, with the use of traffic-aware metrics, the gain in throughput achieved by flows inserted into a network with a large number of existing flows increased with the increase in contention in the network.

I further evaluated the performance of the different metrics for shortlived flows arriving at various rates in the Manhattan scenario. I found that the traffic-aware metrics provided benefits similar to those obtained in experiments with a single additional flow and multiple additional flows. Also, the benefits increased as the rate of flow arrival decreased. For example, with one flow arriving every
2 seconds, and with the use of directional antennas, the AVAIL metric showed up to 35% higher average throughput than the throughput achieved by the traffic-unaware metrics.

Furthermore, even in the case of multiple flows simultaneously arriving at the network, the use of traffic-aware metrics provided higher average throughput (and lower standard deviation) as compared to those achieved by the use of traffic-unaware metrics. For example, the use of the AVAIL metric leads to around 31% higher average throughput than that obtained by using the traffic-unaware metrics, whereas the MLS-T metric achieves around 26% higher average throughput.

I also presented a quantitative analysis of the Chaska and the Manhattan topologies. The analysis corroborated the fact that the Chaska topology inherently has higher interference than the more structured Manhattan topology. Moreover, the average number of gateways accessible within some fixed wireless hops is higher in the Manhattan topology than in the Chaska topology. This is in spite of the fact that even though both the topologies have similar number of total nodes, the Manhattan topology has just 10 total gateways, compared to 14 total gateways in the Chaska topology.

Since a wireless mesh network is intended for providing widespread wireless connectivity to a large geographical area, such network deployments will have simultaneous traffic between a large number of source-destination pairs. Hence,
from the results presented in this chapter, I conclude that in such a network deployment scenario, traffic-aware metrics must be used.
Chapter 4

Coverage Improvement in Wireless Mesh Networks

Coverage is an important issue in wireless networks. To maximize coverage, commercial deployment of wireless networks, such as cellular networks and commercial WiFi networks, as well as non-commercial deployments, such as academic and corporate WiFi networks, have to carefully plan the placement of access points (WiFi access points in WiFi network, TAPs in a wireless mesh network, or cellular base stations in a cellular telephony network). Though the issue of throughput, dealt with in the previous chapter, is less understood by the layman, network coverage, or the lack of it, is a commonly understood concept, and has been cleverly exploited in advertisements of cellular service providers.

For the sake of clarity, I define "coverage of a network to be the percentage of a geographical area where clients can expect to get a certain transmission rate from the network." For example, if only half the campus of a university has IEEE 802.11b WiFi access, then the coverage of the university is 50%. Since the lowest transmission rate for an IEEE 802.11b network is 1 Mbps, 50% of the campus can be expected to be able to transmit at 1 Mbps or higher. The actual throughput that the clients get would depend upon several other factors such as proximity to access points, packet loss, and interference, as have been described in Chapter 3.
Figure 4.1: Client C is further away from client B and hence has weaker signal from the TAP. The assumption here is that the TAP uses the same transmission power for transmissions to both B and C and that the only factor affecting the received signal strength is the geographical distance of the receivers from the TAP node. In reality, other factors, such as the presence of trees or buildings, might also affect the received signal strength.

4.1 Coverage Issues

Wireless signal strength decay is inversely proportional to the third (or even higher) power of the distance between the transmitter and the receiver. The received Signal-to-Noise-and-Interference-Ratio (SINR) deteriorates sharply as the location of the receiver moves from near the base station towards the extremity of the coverage area of the base station. Hence, even with careful placement of access points, clients that are at the edge of the coverage area of a base station, receive relatively weak signal from the base station compared to the signal received at nodes that are placed closer to the base station. Consequently, the clients at the edge need higher transmit power in order to transmit successfully to a base station. For example in
Figure 4.2: Addition of bTAP to augment coverage of TAP. The circle depicts the coverage area of the TAP. Client C is at the edge of the coverage area and hence has a low SINR from the TAP node. However, C has a higher SINR from the bTAP and hence can use the bTAP to forward C’s transmissions to the TAP.

Figure 4.1, the SINR at client C is much lower than the SINR at client B, assuming a constant transmission power at the base station.

My design to augment the coverage of wireless mesh networks is by deploying “low cost booster TAPs (bTAPs)”, as shown in Figure 4.2. These bTAPs are deployed for the sole purpose of digitally relaying messages between the TAP node and the clients. Each bTAP is connected to a single TAP node and all communication between a bTAP and the TAP occurs over a dedicated wireless link between the bTAP and the TAP, called the backhaul link. Thus, the bTAPs act as boosters for communication between a TAP and a client. Figure 4.2 shows a rather simplistic view and in later sections, I will clarify how bTAPs can be deployed in a more realistic representation of the coverage area of a TAP. These bTAPs conceptually
participate in both downlink and uplink transmissions. A downlink transmission is
defined to be a transmission from either a TAP or a bTAP to a client or from a TAP
to a bTAP. An uplink transmission is defined to be a transmission from a client to
either a TAP or a bTAP or from a bTAP to a TAP. Whether the bTAP actually par-
ticipates in the communication to and from a client depends upon several factors,
which I will clarify later.

Clients in a wireless mesh network generally do not forward any packets. A
client either directly communicates with a TAP or communicates to a bTAP that
then forwards traffic to and from a TAP, potentially over several hops. Since the
bTAP is deployed by network operators, the bTAP antenna can be better placed
than the antenna of a normal user and can be even directional if necessary. Thus,
the bTAP ↔ TAP link can be optimized to be a high transmission rate link, even
when using the same radio technology as a client, since most modern technologies
offer adaptive modulation based on channel conditions. Moreover, bTAP ↔ client
links are often shorter than bTAP ↔ TAP links, thus requiring less transmission
power at clients and consequently causing less interference to other clients.

The rest of the chapter is organized as follows. I first describe certain variations
of my scheme in Section 4.2. Then, in Section 4.3, I present some related work in
the field of coverage improvement. Following that, Section 4.4 presents the de-
sign used for bTAP deployment. Section 4.5 presents the evaluation of the design.
Finally, I summarize in Section 4.6.
4.2 Variations of bTAP Deployment Model

There are several variations of the bTAP deployment model in a wireless mesh. For example, the TAP ↔ bTAP backhaul can use a different spectrum or frequency channels than the spectrum used for the client ↔ TAP and the client ↔ bTAP links. This arrangement is a wireless backhaul system, which trades off the cost of extra spectrum, radios, and associated network engineering against the cost of reduced wired backhauls and their installation. Moreover, once a particular band of spectrum is dedicated to backhaul links, that spectrum often cannot be used for other purposes due to frequency reuse patterns and other engineering factors, thus preventing flexible use of the generally licensed and costly spectrum.

Another choice is to have the backhaul link dynamically share the same radio resources that is used for clients. Dynamically sharing the same radio resource has several advantages such as using the same type of mass-produced radio technology for the clients, the TAP, and the bTAP, no additional spectrum cost, and the possibility of dynamic resource reuse within the system. Since the spectrum is shared, it is important to design for flexible and frugal use of the shared spectrum, which is often licensed at high costs. In the work presented in this chapter, I choose to analyze this latter option, i.e., to analyze, in terms of outage and capacity, the performance of a wireless mesh network, augmented with low cost booster TAPs, in which backhaul links do not use dedicated radios and dedicated spectrum.
The work presented in this work presents a novel and practical wireless mesh suitable for wide-area wireless access networks such as WiMAX/802.16, HSPA, and EV-DO that use centralized MAC in licensed and limited spectrum. I examine the capacity and the coverage of the bTAP deployment model under multicell frequency reuse patterns and realistic wireless channel models, accounting for multicell co-channel interference and radio resource consumption by forwarding links.

4.3 Related Work

Wireless mesh networks utilizing IEEE 802.11/WiFi and its distributed MAC protocol have been studied and even deployed in commercial networks with products by companies such as Tropos Networks [76], Belair Networks [12], Firetide [27] and Mesh Dynamics [57]. Since the cost of 802.11-based radio systems is low and a large amount of unlicensed spectrum is available, WiFi-based systems have become very popular in both academic establishments as well as in corporate offices.

These systems often employ separate radio channels for backhaul links. For example, Mesh Dynamics [57] employs two 802.11 radios, on different channels in the 5.8 GHz unlicensed band, for backhaul links, thus allowing simultaneous reception and transmission on the backhaul, and a single IEEE 802.11 radio in the 2.4 GHz unlicensed band for the last hop to client devices. Belair Networks [12] uses a similar multi-radio mesh concept and operates the dedicated point-to-point
links between access points on separate channels in the 5.8 GHz band. Firetide [27] uses a single channel for the entire backbone network consisting of only access points and configures the channel to be different from the channels used to communicate with clients. Tropos' [76] Chaska, Minnesota, USA deployment includes a wireless backhaul of 250 MetroMesh routers. Each router is equipped with a single IEEE 802.11b radio, which is used both for communicating on the wireless backhaul and for communicating to the clients. Some of these routers function as gateways and are connected to the wired backbone network either by a wire or by fixed, line-of-sight, 5 GHz IEEE 802.11a links.

Pabst et al. [64] present an overview of the use of relays in cellular systems. They note that most existing 2G/3G cellular systems were designed for direct and bidirectional communication between access points and clients directly connect to these access points. As a result, the use of relays, placed between clients and access points, would require changes in the deployed architecture. Consequently, Time-division Multiple Access (TDMA)-based systems are better suited to introducing relays, since TDMA-based systems, compared to contention-based systems, can better allocate the radio resources between the client-to-relay link and the relay-to-access point link. The authors also report many previous investigations on multihop relaying, the hybrid use of WiFi and cellular technologies, and the use of separate radio channels and Time Division Multiplexing (TDM). The authors also outline several important research issues, such as radio resource management, and
routing, that need to be solved in order to practically realize relaying in a cellular system.

Aggelou and Tafazolli [5] propose Global System for Mobile Communications (GSM) client, with weak signal from the base station, to relay traffic to other client in order to improve coverage of GSM base stations. They propose adding an ad hoc mode into the GSM protocol suite; thus, the suite is divided into two parts: one handling GSM radio access and the other handling wireless multihop access.

There have been various papers in the field of integrating cellular and ad hoc network architectures. Xu et al. [82, 83] propose a load balancing scheme called Mobile-Assisted Data Forwarding (MADF) in which an ad hoc overlay is added to the fixed cellular infrastructure. This ad hoc overlay runs on a channel that is logically separate from the channels used to communicate directly between the clients and access points. Clients in an overloaded cell, i.e., a cell with high traffic, use this ad hoc overlay to route their traffic to lightly loaded neighboring cells.

Similarly, Wu et al. [81] proposed the Integrated Cellular and Ad Hoc Relaying system (ICAR) to balance traffic loads between different cells. The ICAR system uses stationary ad hoc relaying stations to borrow channels from non-congested adjacent cells so that traffic from congested cells can be routed to these adjacent cells.

In another similar approach, Luo et al. [53] proposed the Unified Cellular and Ad Hoc Network architecture (UCAN), which augments a cellular system with
IEEE 802.11-based peer-to-peer links among mobile clients. Data meant for clients with poor signal quality is forwarded by the access point to proxy clients with better signal quality. These proxy clients use the IEEE 802.11-based ad hoc network to forward the traffic to the intended destination. Such peer-to-peer forwarding allows increasing the throughput of clients with poor signal quality without decreasing the aggregate cell throughput.

Jorjeta et al. [46] propose a multi-tier ad hoc network system called Ad Hoc City, complete with a novel routing protocol called Cellular DSR (C-DSR). The backbone network is itself a mobile ad hoc network and is comprised of wireless devices mounted on fleets of buses or delivery vehicles. Additionally, the system includes a few access points, which communicate with each other over wired connections. If two communicating clients are within some wireless hops of each other, then the clients communicate over the ad hoc network. However, if the clients are too far away from each other, then the sending client uses the ad hoc network backbone to communicate to the nearest access point that then forwards the packets to the access point that is nearest to the destination client.

Miller et al. [59] proposed a similar hybrid architecture in which mobile clients are at most $K$ hops away from either the destination or some access point; access points are connected directly to the wired network. The authors suggested a combination of proactive routing at the access points and on-demand routing at the mobile clients. Lin and Hsu [51] presented the theoretical analysis of the per-
formance of a somewhat similar architecture, called the Multi-hop Cellular Networks (MCN). They found that if the transmission range of base stations decrease by a factor of $k$, then the number of simultaneous transmissions in the cell increase by a factor of $k^2$, and the mean hop count increases by a factor of $k$.

Cho and Haas [19] studied the advantage of multihop relaying on the downstream throughput of a cellular network. Their study suggests that most of the throughput gains can be achieved by two-hop and three-hop multihop relaying. The authors also showed that the use of multihop relays led to better QoS, since due to the use of these relays the effect of location-dependent signal quality is mitigated to a large extent. Additionally, the authors proposed allowing for two types of concurrent transmissions, each concurrent transmission using a unique CDMA spreading code: first, between the multiple hops of a single downstream flow, and second, between multiple hops of different downstream flows.

Viswanathan and Mukherjee [78] analyzed the effect of placing wireless relays in 120° sector and found that barring signaling overhead, the cell throughput gain was between 50% and 70% with four relays in a single sector. In a later work [62], the same authors analyzed a system having two types of base stations. Some of these base stations have a wired connection to the backhaul network and are called Access Points (APs) and the remaining base stations, called Extension Points (EP), are otherwise similar to the APs but do not have a wired connection to the backhaul network; EPs simply extend the range of the APs through multihop communica-
tion. The authors considered a specific geometry in which a plane was tessellated with hexagons and the AP was placed at the center of one of the hexagons whereas the EPs were placed at the center of several concentric (at the AP) rings of hexagonal cells; each AP or EP communicates with clients located in their respective hexagons. Moreover, only a single channel was used in the system. Given such a system, they studied how the downlink throughput per EP varied with the distance of the EP from the AP. They showed that the throughput per cell decreases faster than linearly with the increasing number of cells served by a single AP. Additionally, they showed that if the AP can transmit at a significantly higher transmit power than the transmission power of EPs and from a greater height than the antenna heights of EPs, then there is no gain from using EPs outside of the first ring (i.e. there is advantage from only one additional wireless hop). However, if the APs and the EPs use the same power and have their antennas at the same height then there is some, though not substantial, throughput gain from using multiple rings of EPs.

In my work, the focus is on incorporating multihop forwarding into existing systems based on wide-area cellular air interfaces with centralized MAC protocols such as 802.16/ WiMAX, HSPA, EV-DO. These systems are for commercial access services using limited licensed spectrum. For such systems, I examine the impact of booster TAPs on the coverage and capacity under multicell frequency-reuse patterns and realistic wireless channel models.
Next, I present the design of my proposed deployment of bTAPs in a wireless mesh network.

4.4 Design

In this section, I present my design to augment wireless mesh networks with additional booster TAPs (bTAPs) with a goal to increasing the coverage and system capacity of TAP nodes. I start by stating the design assumption and goals.

4.4.1 Design Assumptions and Goals

In order to proceed with the analysis, I set up the following design goals and assumptions. First, I assumed that the system deploys a centralized medium access control (MAC) packet radio system. Since centralized MAC allows better control over resources, centralized MAC is considered more suitable for wide-area wireless systems. Use of contention based protocols, such as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), suffer from hidden terminal problem and consequently are inefficient for large cell radii. Although WiFi, i.e., IEEE 802.11-based systems most commonly use a distributed MAC called DCF (Distributed Coordination Function) or its variants with QoS, the standard does offer a less widely used centralized MAC called PCF (Point Coordination Function) and its variants with support for QoS traffic. Second, for ease of analysis, I assumed that clients are uniformly distributed in the analyzed area and that clients always have
backlogged data to send and receive, i.e. each client has infinite bandwidth requirement. Finally, as a design principle for near-term applicability, I require that the bTAP must have "low complexity and low cost of deployment and operation". To elaborate

- The bTAP antenna should be at a lower height than the TAP antenna for more economic deployment. A lower height requirement enables the use of alternative site installation options such as power or street lamp poles, rather than conventional tower sites commonly used for cellular TAPs. Also, it is possible that the largely unused lower parts of conventional towers may be available at lower costs.

- The bTAP should use a single radio to communicate with TAPs, other bTAPs, and client nodes.

- The bTAP should use an omni-directional antenna to communicate with client nodes.

With the exception of the first requirement, the other requirements mean that the bTAP would have complexity, and hence price, comparable to that of a client. There are some differences, since the bTAP may use a directional antenna for backhaul links and may need to perform some local scheduling decisions. However, the bTAP, using a simple switch, would be able to use the same radio for communicating through either the omni-directional antenna or the directional antenna. If
Figure 4.3: An idealized cell consisting of six sectors, each using an orthogonal channel. Consequently, the TAP node placed at the center of the cell has six radios and six directional antennas.

the directional antenna is adaptive to point in different directions, the antenna can be used to configure the directional antenna of the bTAP to point towards the best TAP, and also to change towards a different TAP, if needed.

4.4.2 System Model

As shown in Figure 4.3, an idealized cell consists of six sectors, each using a different, non-interfering frequency channel, represented by the different colors. The TAP is placed at the center of the cell and has six radios and six directional antennas. I chose to add two bTAPs at the base of the triangle representing a single sector, as shown in Figure 4.4(a); this amounts to one bTAP for each sector, as shown in Figure 4.5. The location and the number of bTAPs require further justification.

First, the bTAP is placed at the region farthest from the TAP because these areas generally have the highest path loss and higher probability of poor channel conditions, and consequently, clients located in these areas require higher trans-
Figure 4.4: Different bTAP placement possibilities. Placing two bTAP nodes, each using an omni-directional antenna, provides a solution for low-cost bTAPs. Placing a single bTAP with an omni-directional antenna would require the bTAP radio to have a transmission range comparable to the transmission range of the TAP radio. This would violate the design principle of a cheap bTAP node.

mit power to communicate to the TAP. Transmissions at higher power, especially from the periphery of a sector, cause higher co-channel interference in nearby sectors. Also, depending on traffic load, bTAPs, placed at the periphery of a cell, can more readily choose an alternate TAP to connect to.

Second, instead of placing just one bTAP at the middle of the base of the sector triangle (as shown in Figure 4.4(b)), I choose to place two bTAPs, as shown in
Figure 4.4(a). This decision is based on the design choice for low-complexity, low-height, bTAPs using omni-directional antennas, although it does require a larger number of bTAPs to be deployed. In order to cover the large portion of the edge of the sector with a single bTAP, the range of such a bTAP has to be comparable to the main TAP, which typically employs high towers, high-power amplifiers, directional antennas with complex patterns, or other complexities that the design aims to avoid. Also, a higher power transmitter at the periphery of the sector would cause higher co-channel interference at nearby sectors.

In addition to the two bTAPs, if required, more bTAPs can be placed inside a sector to form links back to TAP via additional hops. However, I choose to limit this work to a case of one more additional hop than conventional systems, since it presents a good starting point for the analysis of the tradeoff between system cost and performance gain. Besides, results presented by Cho and Haas [19] show that the first additional hop provides the most performance gains. Moreover, a single additional wireless hop is easier to manage and schedule for, especially in a system with centralized MAC. Existing frequency reuse patterns are unaffected with the presented arrangement of bTAPs in initial planning as well in incremental deployments of bTAPs.

Once the location of the bTAPs are known, clients lying in the sector choose to connect to the TAP either directly, or through the bTAP. For downlink data in the example shown in Figure 4.4(a), client node A might directly receive from the
TAP whereas client node B might have a weak signal from the TAP. Hence, any packet intended from the TAP to node B might be first forwarded over the wireless backhaul link to the bTAP, and the bTAP in turn would deliver these packets to node B; a similar argument holds for sending uplink data from either node A or node B to the TAP.

Given such a model, I analyze whether the simultaneous use of same radio resources by two different subscriber stations within a single sector, hence using the same logical channel, is possible or not. Such simultaneous use of the same logical channel is in contrast to existing use of frequency within a single sector, which does not allow for simultaneous transmission to or from two clients from the same sector. If indeed it is possible to simultaneously schedule two users, one to the TAP and the other to the bTAP, I aim to find the resulting expected system throughput gain, taking into account the radio resources used on the TAP→bTAP backhaul links.

For example, in Figure 4.4(a), client A may be scheduled to receive from the TAP at the same time as client B is scheduled to receive from the bTAP. This simultaneous scheduling is possible if the interference from the bTAP at client A, in addition to the co-channel interference from other cells, is low enough so that the Signal-to-Noise-plus-Interference-Ratio (SINR) at node A is sufficient for robust reception from the TAP. Similarly, the SINR at node B, even in the presence of interference from the TAP along with other co-channel interference from other cells,
must be high enough for robust reception from the bTAP. For the downlink data in the example shown in Figure 4.4(a), if the received signal strength at a receiver is assumed to depend only on the distance from the transmitter, then if node A lies on the solid arc near the TAP, then node B has to lie further from the TAP than the lower solid arc. Similarly, if node B lies on the dashed circle near the bTAP, then node A has to lie further from the bTAP than the dashed arc. If both the conditions are satisfied, then both nodes A and B can be scheduled to simultaneously receive from the TAP and the bTAP, respectively. A similar reasoning applies to simultaneous uplink transmission from clients to the TAP and bTAP.

In the next section, I elaborate why analyzing a single sector might give an incomplete picture since the pair of clients that are best suited for simultaneous reception might not geographically be located in a single sector.

4.4.3 Limitation of Single Sector Analysis

I realize that analyzing a single sector along with co-channel interference from other cells (including additional co-channel bTAPs in my approach), as done in most multicell performance analysis, provides only an incomplete picture. For the schematic shown in Figure 4.4(a), suppose a candidate client node is to be chosen for transmission from the bTAP simultaneously and on the same logical channel as the TAP→A transmission. Instead of choosing node B node C, a client geographically inside the facing sector, could be chosen to be scheduled to receive simultane-
Figure 4.5: A unit composed of two adjacent sectors. A unit composed of a single sector, shown in Figure 4.4(a), gives an incomplete picture.

ously with node A, both node A and node C receiving on the same channel. Upon a transmission from the bTAP, the SINR at node C is more likely to be higher than the SINR at node B, since the TAP, being farther away from node C than from node B, is likely to cause lower interference at node C. This scheduling choice could yield better overall performance.

As a result, instead of analyzing just the single sector, I analyze a unit composed of two facing sectors, as shown in Figure 4.5. As shown in the figure, the left bTAP communicates with the upper TAP, but many of the users (for example node A) that it communicates to, geographically lies in the lower sector. Though node A lies in the lower sector and would have used a separate frequency had it chosen
to communicate to the lower TAP, if node A chooses to communicate with the left bTAP, then node A could communicate using the frequency that is being used in the upper sector. Similarly, the right bTAP communicates with many users that geographically lie in the upper sector (for example node B), even though the bTAP itself communicates only with the lower TAP.

I assume that the upper and lower TAP use different logical channels to avoid interference, as is anyway standard in most current frequency reuse patterns used to distribute a limited set of channels among different sectors. However, if the radio technology can work with the same channel in both sectors, the presented approach would still be applicable. Without loss of generality, I assume that the Upper TAP and the Left bTAP (Figure 4.5) use the same channel, which is different from the channel used by the Lower TAP and the Right bTAP (Figure 4.5). This approach demands no changes to an existing frequency reuse pattern. Also, it is straightforward to add additional bTAPs along the first ones, on the boundary or inside the sectors, and still maintain the same frequency reuse pattern, albeit at the cost of additional scheduling complexity.

Having presented the bTAP deployment model, I now explain how packets can be scheduled in a such a system.
Figure 4.6: A possible downlink scheduling map. The periods shown at the top are not to scale. Also, the schedule does not show certain essential periods, such as periods used for clock synchronization, and slot allocation. In this figure, nodes A and B can receive simultaneously from the TAP and the bTAP respectively, whereas nodes C and D need to be scheduled in isolation.

4.4.4 Scheduling in the Presence of bTAPs

Scheduling decisions for each sector can be taken independent of the adjacent sector (e.g., the sector managed by the lower TAP in Figure 4.5) since the adjacent sector uses a separate channel.

Figure 4.6 shows an example downlink scheduling frame that is compatible with the frame structure currently used in most centralized MAC systems, including IEEE 802.16/WiMAX. The figure does not show certain essential periods, such as periods used for clock synchronization, client slot allocation, and modulation.
scheme allocation. In the figure, client nodes A and B may be simultaneously scheduled to receive from the TAP and the bTAP respectively whereas nodes C and D may need to be scheduled separately to receive from the TAP and the bTAP respectively. Other dimensions of orthogonal resource sharing within a single sector such as tones in OFDMA and codes in CDMA systems can be treated as additional logical channels that can be independently scheduled within a sector. The downlink frame contains the following periods:

- **Backhaul period**: In this period, the data is forwarded from the TAP to the bTAP using the backhaul link. This data is meant for nodes B and D.

- **Dedicated Schedule period**: This period consists of two consecutive periods. In the first part, the TAP transmits data to node C and in the second part, the bTAP forwards data to node D that the bTAP received in the backhaul period.

- **Simultaneous Schedule period**: In this period, the TAP transmits data to node A and at the same time, the bTAP forwards the data meant for node B that the bTAP received in the backhaul period.

I assume no scheduling coordination among cells or sectors, except the normal downlink and uplink synchronization necessary in Time Division Duplex (TDD) cellular systems.

The order of these three periods can be different, but the current order is chosen to simplify radio design and to improve compatibility. For example, the bTAP
must switch from receive mode to transmit after the backhaul period, thus placing the dedicated schedule period for the TAP (instead of the dedicated schedule period of the bTAP) right after the backhaul period relaxes the switching time requirement at the bTAP. It may also enable the accommodation of clients that are unaware of the multihop forwarding operations and expect continuous frames for reasons such as synchronization. Such clients can be accommodated during the dedicated schedule period for the TAP, or during the dedicated and simultaneous schedule periods for the bTAP. Also, the backhaul period may be staggered to a previous frame if a longer processing time is needed, possibly for local scheduling decisions or packet aggregations for improved efficiency. The packet transmitted by a bTAP during its dedicated schedule period need not be received from the TAP in the backhaul period immediately preceding the dedicated schedule period, but can be received in an earlier backhaul period. It is straightforward to devise a similar uplink frame structure, most likely with the backhaul period at the end, for the minimum latency associated with forwarding to the TAP.

So far, the frame structures apply equally to Frequency Division Duplex (FDD) and TDD systems with a single radio bTAP. Under FDD, or with multiple radios or advanced antenna technologies such as MIMO and beam forming, it is possible to envision sharing the backhaul periods with some user traffic periods at the TAP or the bTAP. Such simultaneous scheduling opportunities would be more common with advanced PHY techniques.
Although it is not difficult to extend this frame structure to a few more additional hops, a coordinated scheduling as described here may be difficult to extend to a large number of wireless hops for various reasons, such as conflicting radio conditions, the availability of channel information across multiple hops, and algorithmic complexity. As mentioned earlier, Cho and Haas [19] suggest that most of the performance gains are achieved with just one (case analyzed in this work) or two additional wireless hops. Thus, the complications of additional hops may not be justified.

Without going into the details, I also assume the availability of simple wireless link condition measurements to help select clients that can be simultaneously scheduled within a single sector. In most existing systems, such measurements may also be available from normal traffic exchanges. Additionally, dedicated measurement packets may be needed while other clients are instructed to remain silent. One observation is that if clients are highly mobile, the measurements become stale quicker. Any scheduling algorithm must consider these effects along with other information. The decision to connect to either the TAP or the bTAP in a sector should be relatively long-lived, on the time scale of large-scale fading of wireless channels [43]. It is difficult to coordinate scheduling and route dynamics while adapting to small-scale fading or fast fading, and small-scale fading is usually more effectively handled by signal processing techniques exploiting frequency, time, and antenna diversity.
Having presented a scheduling scheme, I now calculate the effective rate that a client gets when the client chooses to communicate with the TAP via the bTAP. The effective rate is essential in order to calculate both the coverage and the capacity of my scheme.

4.4.5 Effective Data Rate

If a client is scheduled to communicate to the TAP via the bTAP, the effective rate to the TAP from the perspective of the overall sector throughput is \( \frac{r_{bh} \cdot r_i}{r_{bh} + r_i} \), where \( r_i \) is the rate at which a client \( i \) connects to the bTAP and \( r_{bh} \) is the backhaul rate at which the bTAP connects to the TAP. The effective rate is derived as follows. Suppose the client wants to send (or receive) \( B \) bits of data to the TAP via the bTAP. The client has to send the data first to the bTAP at a data rate of \( r_i \). The bTAP then forwards the data to the TAP at a data rate of \( r_{bh} \). Hence, the total time spent to transport the \( B \) bits of data from the client to the TAP is \( \frac{B}{r_i} + \frac{B}{r_{bh}} \). Thus, the effective data rate, \( r_e \), from the client to the TAP is given by

\[
\begin{align*}
    r_e &= \frac{\text{Total data transmitted to TAP}}{\text{Total time taken}} \\
    &= \frac{B}{\frac{B}{r_i} + \frac{B}{r_{bh}}} \\
    &= \frac{1}{\frac{1}{r_i} + \frac{1}{r_{bh}}} \\
    &= \frac{r_{bh} \cdot r_i}{r_{bh} + r_i}
\end{align*}
\]
Table 4.1: Modulation levels considered. The third row, in bold faced font, represents the modulation scheme and corresponding data rate used for transmissions on the backhaul link, i.e., the directional link between the TAP and the bTAP.

<table>
<thead>
<tr>
<th>MODULATION</th>
<th>CODE RATE</th>
<th>REQUIRED SINR (dB)</th>
<th>DATA RATE (Mbps) $r_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>QPSK</td>
<td>1/2</td>
<td>6.6</td>
<td>6.0</td>
</tr>
<tr>
<td>16-QAM</td>
<td>1/2</td>
<td>10.5</td>
<td>12.0</td>
</tr>
<tr>
<td>64-QAM</td>
<td>2/3</td>
<td>15.3</td>
<td>24.0</td>
</tr>
<tr>
<td>64-QAM</td>
<td>3/4</td>
<td>20.8</td>
<td>27.0</td>
</tr>
</tbody>
</table>

For clients directly communicating with the TAP, the effective rate is the same as the transmission rate of the client (or the transmission rate of the TAP for downlink communication).

I now proceed to evaluate the coverage and capacity of the augmented wireless mesh network.

4.5 Evaluation

I now describe the evaluation, performed using MATLAB 7.0 [56] simulations. Unlike the ns-2 simulations used in Section 3.6, a packet level simulator is not required because the metrics of interest are coverage and available sector throughput, which are not directly affected by packet level dynamics.

4.5.1 Evaluation Parameters

Table 4.1 shows the modulation levels used in the analysis and the corresponding average SINR thresholds and transmission rates for a channel with 6 MHz
usable bandwidth. Due to technology-dependent overheads such as guard bands, guard times, preambles, pilots, and packet headers, the spectrum per radio channel is larger than the usable bandwidth. I used 6 MHz for numerically calculating system capacity for relative comparisons without being specific to any particular technology. Discrete levels of modulations are used, instead of the continuous function of Shannon capacity, to examine more practical aspects of a system. The use of continuous capacity curves results in somewhat optimistic performance due to its infinite flexibility in utilizing available SINR. However, it may lead to overly optimistic results, especially in studying multihop mesh networks. Real systems cannot take full advantage of high SINR of short distances, due to limits on the amplifier linearity and the power consumption required for high-level modulations. If the analysis of mesh networks fails to account for this effective ceiling of peak rates, the analysis yields large but unrealistic performance gains. The values, shown in Table 4.1, are sufficient for relative comparisons between systems with or without bTAP.

The SINR values shown in Table 4.1 reflect the local average SINR over small-scale fading. At each modulation rate, small-scale fading yields a certain average bit error rate, such as $10^{-5}$ (1 bit error in $10^5$ transmitted bits). It is simpler and more realistic if the decision on whether a particular client connects to a TAP or a bTAP is made on the time scale of the large-scale fading instead of the small-scale fading of radio channels. Large-scale fading, also called shadow fading, is a much
slower fluctuation of radio signals than is small-scale fading, such as Rayleigh or Rician fading [43]. The chosen SINR values are representative values for an IEEE 802.16-like system with convolutional turbo codes under Rayleigh fading [37, 50, 65]. Additionally, the transmission rate used for the TAP → bTAP backhaul link, shown in bold in Table 4.1, has been verified under the large-scale pathloss model for the geometry used for the bTAP deployment.

4.5.2 Erceg-Greenstein Channel Model for Large-Scale Fading

I used the Erceg-Greenstein model [24] as the large-scale pathloss model, and most of the results are shown for Terrain A of the Erceg-Greenstein model. This model is the basis for the path loss model adopted by the IEEE 802.16 Working Group [38]. Terrain A is representative of hilly areas with moderate to heavy tree density. I carried out simulations with the other terrain types mentioned in the model (terrain B and C) and the results obtained were largely similar in trend to the results obtained for Terrain A.

The original Erceg-Greenstein model was proposed for propagation in suburban environment and was derived from measurements done with an omni-directional user antenna at a height of 2 m operating on a 1.9 GHz channel. The original model has been extended to incorporate change in carrier frequency and user antenna height [38]. In the extended model, the path loss, $PL(d)$, at a distance of $d$ from the transmitter, is given by
\[ PL(d) = PL_0 + 10 \cdot \gamma \log_{10}(d/d_o) + \text{fading} + \]
\[ \text{Mod}_{\text{frequency}} + \text{Mod}_{\text{user height}} \]  \hspace{1cm} (4.2)  
where

\[ PL_0 = 20 \cdot \log_{10}(4\pi d_o/\lambda), \]
\[ d_o \geq 100 \text{ m}; \text{ called the reference distance}, \]

Pathloss exponent \( \gamma = a - b \cdot h_b + \frac{c}{h_b}, \)

\[ \text{Mod}_{\text{frequency}} = 6 \log_{10}(f/2000); f \text{ in MHz}, \]
\[ \text{Mod}_{\text{user height}} = \begin{cases} -10 \log_{10}(h_u/2), & \text{channels A and B} \\ -20 \log_{10}(h_u/2), & \text{channel C} \end{cases} \]

The term \( \text{Mod}_{\text{frequency}} \) is for carrier frequency correction, and the term \( \text{Mod}_{\text{user height}} \) is for receiver antenna height correction. Additionally, \( a, b, \) and \( c \) are all data-derived constants and vary with each of the three terrain categories, A, B, and C, respectively. The Erceg-Greenstein model has the path loss exponent ranging from 4.8, for Terrain A, to 4.12, for Terrain C. The term \( PL_0 + 10 \cdot \gamma \log_{10}(d/d_o) \) is the mean path loss at distance \( d, \) whereas the term \( \text{fading} \) is the random shadowing variation about the mean. This fading is approximated as a zero-mean Gaussian variable with a certain standard deviation. Since Equation 4.2 is in log scale, this fading is often called lognormal fading.
Figure 4.7: Directional antenna pattern for a directional antenna with 3 dB beamwidth of 30°, which is obtained by setting $\alpha = 1.55$ in Equation 3.35. The concentric circles are in units of dB.

In addition to a large-scale fading model, I also used a directional antenna model since the TAP and the TAP→bTAP backhaul link use directional antennas.

4.5.3 Directional Antenna Model

For easier reading, I have repeated the normalized directional antenna pattern in Figure 4.7. As explained earlier in Section 3.6.5, this pattern, based on the model of a rectangular aperture antenna [10], is an approximation of typical directional antennas in use for cellular systems. The directional antenna at the TAP has a 3 dB beamwidth of 30°, as commonly used for 60° sectors. A directional antenna with a 3 dB beamwidth of 60° is too wide for a 60° sector, since its substantial
gain outside of $60^\circ$ causes and receives high interference. The clients use omni-directional antennas and so does the bTAP when communicating with the clients.

4.5.4 Frequency Reuse Pattern

Figure 4.8 shows the (1,6,6) frequency reuse pattern used to calculate throughput and coverage without bTAPs. The simulation considers interference from cells that are at most two co-channel rings away from the sector being analyzed. As described earlier in Section 2.1, (1,6,6) means that a reuse group consists of a single cell, represented by the “1”. This reuse group, containing only a single cell, uses six channels, represented by the “6”. Finally, the last “6” in the notation represents the number of sectors into which the cell is divided.
Figure 4.9: Multicell scenario with bTAPs for (1,6,6) frequency reuse pattern. The unit is the diamond shape that lies between cell 0 and cell 5.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency reuse</td>
<td>(1,6,6) and (1,3,6)</td>
</tr>
<tr>
<td>Cell radius</td>
<td>1000 m</td>
</tr>
<tr>
<td>TAP antenna gain</td>
<td>20 dBi</td>
</tr>
<tr>
<td>bTAP antenna gain</td>
<td>0 dBi</td>
</tr>
<tr>
<td>TAP antenna height</td>
<td>30 m</td>
</tr>
<tr>
<td>bTAP antenna height</td>
<td>15 m</td>
</tr>
<tr>
<td>Client antenna height</td>
<td>2 m</td>
</tr>
<tr>
<td>Transmit power</td>
<td>30 dBm</td>
</tr>
<tr>
<td>Lognormal fading standard deviation</td>
<td>10.6 dB (Terrain A), 8.2 dB (Terrain C)</td>
</tr>
<tr>
<td>Power control</td>
<td>No</td>
</tr>
</tbody>
</table>
Figure 4.9 shows the frequency reuse pattern used to calculate both the downlink and the uplink performance in the presence of booster TAPs. The pattern includes more cells than the pattern shown in Figure 4.8, due to the two-sector analysis required when bTAPs are present. When calculating the downlink co-channel interference with bTAPs, the simulation calculates the worst case interference from both the TAPs and the bTAPs transmitting or receiving simultaneously in all co-channel sectors. Although the results here are from the worst case, the analysis indicates that the total downlink co-channel interference from outside of a sector is dominated by that from the TAPs in those sectors, i.e., the bTAPs in the other sectors cause negligible interference. This domination is because TAP nodes use directional antennas and because antennas of TAP nodes are higher than those of bTAPs. When calculating the uplink co-channel interference in the presence of bTAPs, the simulation calculates interference from two randomly chosen clients from each of the interfering units; for the uplink co-channel interference without bTAPs, the simulations calculates the interference from only one randomly chosen client from each of the interfering sectors. Additionally, the simulation includes thermal noise in all SINR calculations. The remaining parameters used for the simulation are shown in Table 4.2.
4.5.5 Achievable Transmission Rates without bTAPs

Figure 4.10 shows the scatter plot of achievable transmission rates over a sector without any bTAP. For illustrative purpose, Figure 4.10(a) shows the scatter plot when no lognormal fading was applied to the pathloss. The directionality of the antenna used is reflected in the lobular regions of the figure. On the other hand, Figure 4.10(b) shows the scatter plot when realistic lognormal fading (Section 4.5.2) was applied to the pathloss. Due to non-uniform variations in the fading characteristics of the channel, the use of a directional antenna is barely noticeable from this scatter plot.
Algorithm 1 Pairing algorithm to schedule clients for downlink

Require: Unit to be analyzed: UpperTAP, LeftbTAP, LowerTAP, RightbTAP.
Require: UP = Frequency used for communication with left bTAP or upper TAP.
Require: DOWN = Frequency used for communication with right bTAP or lower TAP.
Require: \( \mathcal{P} = \) List of \( n \) points in the sector. \( \mathcal{P} = p_1, p_2, \ldots, p_n \)

1: Select \( \mathcal{I} \), sectors interfering with analyzed sector.
2: for \( p = p_1 \) to \( p_n \) do
3: \( p \text{UPInterference} = \) Interference caused by TAP’s and bTAP’s belonging to \( \mathcal{I} \) and using UP frequency.
4: \( p \text{DOWNInterference} = \) Interference caused by TAP’s and bTAP’s belonging to \( \mathcal{I} \) and using DOWN frequency.
5: end for
6: for \( p_i = p_1 \) to \( p_n \) do
7: for \( p_j = p_1 \) to \( p_n \) do
8: Schedule \( (p_i, p_j) \) from among the following possibilities such that the total throughput for \( (p_i, p_j) \) is maximized:
9: \( (p_i \rightarrow \text{UpperTAP \ AND \ } p_j \rightarrow \text{LeftbTAP}) \) OR
10: \( (p_j \rightarrow \text{UpperTAP \ AND \ } p_i \rightarrow \text{LeftbTAP}) \) OR
11: \( (p_i \rightarrow \text{LowerTAP \ AND \ } p_j \rightarrow \text{RightbTAP}) \) OR
12: \( (p_j \rightarrow \text{LowerTAP \ AND \ } p_i \rightarrow \text{RightbTAP}) \) OR
13: end for
14: end for
15: while Points can be paired do
16: Choose pair, say \( (p_i, p_j) \), from among all pairs that has maximum total effective rate from the TAP.
17: Remove \( (p_i, p_j) \) and \( p_i \) and \( p_j \) from all other pairings.
18: end while
19: for All remaining points \( p \) in \( \mathcal{P} \) do
20: Schedule \( p \) to one of UpperTAP, LowerTAP, LeftbTAP, or RightbTAP such that rate from TAP is maximized.
21: end for

4.5.6 Determining the Scheduling Regions

The design of an efficient scheduling algorithm has to take into consideration several details such as user density, traffic load, control information, and the existence of traffic requiring QoS service. However, I considered a simplified model characterized by infinite offered load and a high number of clients. The simulation calculates the throughput and outage via an exhaustive search algorithm (Algo-
rithm 1) that determines the respective scheduling regions in a sector for the frame structure shown in Figure 4.6.

The search algorithm works on a two-sector unit as shown in Figure 4.5. An example of a unit being analyzed is shown in Figure 4.9; the unit is the diamond shape that lies between cell 0 and cell 5. All co-channel two-sector units that are within two rings of the main unit are considered as sources of interference (Figure 4.9). All possible pairs of points in the two-sector unit are examined to find the pair with the highest total effective rate, when scheduled simultaneously, considering co-channel interference from other sectors as well as the mutual interference due to simultaneous scheduling. This pairing process is continued until there is no pair left eligible for simultaneous scheduling. At this stage of the algorithm, due to propagation and interference, certain points may not be eligible for simultaneous scheduling. Thus, these points need to be scheduled during a dedicated schedule period for the bTAP or the TAP depending on the effective rate. Even after dedicated scheduling, some points may not be able to communicate to the TAP or the bTAP at the minimum modulation level of QPSK 1/2. In all of the cases, a client chooses a bTAP only if its effective rate $r_e$ (Equation 4.1) to the TAP, via the bTAP is, higher than if the client directly communicates with the TAP.
Figure 4.11: Downlink scatter plot with bTAPs for (1,6,6) frequency reuse pattern without lognormal fading

4.5.7 Simultaneous and Dedicated Scheduling Regions

Figures 4.11, 4.12, and 4.13 show the scatter plot after the deployment of bTAPs in the (1,6,6) reuse pattern. Figures 4.11 shows the scatter plot for downlink transmissions when no lognormal fading is included in the simulation; The figure is presented here only for better visual illustration. Figure 4.12 shows the scatter plot for downlink transmissions but with lognormal fading included in the simulation, and Figure 4.13 show the scatter plot for uplink transmissions with lognormal fading included. The directions of the triangle at any point indicate whether a client
Figure 4.12: Downlink scatter plot with bTAPs for (1,6,6) frequency reuse pattern with lognormal fading

located at that point is connected to the TAP in the upper (△) or the lower (▽) sector, or to the bTAP on the left (<) or right (>) side. The colors of the triangles indicate the transmission rates (as opposed to effective rate (Equation 4.1)) seen by clients. The areas for dedicated scheduling for bTAP is quite small, i.e., there are few < or > symbols in Figure 4.12(b). It is because these areas have similar link conditions both to the TAP and the bTAP, thus communicating directly with the TAP is often advantageous. Given this observation, it may be practical to eliminate this dedicated scheduling period for bTAPs (Figure 4.6).
Figure 4.13: Uplink scatter plot with bTAPs for (1,6,6) frequency reuse pattern with log-normal fading

4.5.8 Transmission Rate from the Perspective of Clients

Figure 4.14 shows the cumulative density function (CDF) of the effective rates with and without bTAPs and for both uplink and downlink transmissions. The CDFs for the effective data rates in the presence of booster TAPs have steps that do not match with the transmission rates shown in Table 4.1. This mismatch is because when a client communicates with the TAP via the bTAP, the effective data rate from the system’s perspective (Section 4.4.5), to and from a wired TAP is less than the transmission rate that the client is able to transmit to or receive from the bTAP.
Figure 4.14: CDF of data rate from client perspective

If the transmission of a packet to and from the client and the forwarding of the corresponding packet over the backhaul link are both performed within a single frame, the client does not experience any additional latency compared to communicating directly with the TAP. However, the overall system observes the corresponding effective rate due to the radio resource spent on the forwarding over the backhaul link between the TAP and the bTAP. From the system’s perspective, this effective rate is not affected even if the packet is backhauling in a later frame, since the transmissions are merely staggered and the earlier frames are used for other packets. The client, of course, experiences higher latency in such a case.

Figure 4.14 shows that a large number of clients that were not able to communicate in the conventional case using the lowest modulation level in Table 4.1, are
able to do so in the presence of bTAPs. This ability explains the substantial improvement of the coverage at QPSK 1/2 (6 Mbps) or higher modulation rates with bTAPs, as shown later in Figure 4.17. However, the number of nodes with effective data rates of 12 Mbps or higher (modulation scheme 16-QAM 1/2 or weaker ¹) does not change substantially, because in the absence of bTAPs, say for the downlink case, there are several nodes, typically lying close to the TAP, that can get a high SINR from the TAP. However, when bTAPs are added to the system, several of these nodes are paired with other clients communicating with the bTAP. The simultaneous transmission from the bTAP node causes interference at the nodes that are receiving from the TAP and hence decrease the SINR at these nodes. This simultaneous transmission decreases the maximum data rate for the downlink to these clients. Of course, the benefit is that the client communicating via the bTAPs were earlier not able to communicate with the TAP or were able to do so at a very low data rate.

4.5.9 Effects of Pathloss Exponent and Lognormal Fading

Figure 4.15 and Figure 4.16 respectively show the change in outage and in sector throughput with changing pathloss exponent and lognormal fading standard deviation. The values are taken from terrain types in the Erceg-Greenstein model where the pathloss exponent of 4.8 and the standard deviation of 10.6 dB corre-

¹As modulation schemes improves, i.e., becomes more robust, the corresponding modulation rate decreases. Higher, the modulation rate, weaker the modulation scheme.
Figure 4.15: Effect of pathloss exponent and lognormal fading on downlink QPSK 1/2 outage

spond to Terrain A, and the pathloss exponent of 4.12 and the standard deviation of 8.2 dB correspond to Terrain C. However, I used these two parameters separately to examine the performance trends over the variation of each of the parameters.

Due to increased interference, the outage increases as the pathloss exponent drops, as shown in Figure 4.15. When compared on the basis of relative changes for the same lognormal fading standard deviation, the system with bTAPs appears more sensitive to the changes of the pathloss exponent, since there are more sources of interference due to the simultaneous scheduling. Given the same pathloss exponent, the higher the lognormal fading standard deviation is, the higher the outage in all cases.
The sector throughput decreases as the pathloss exponent decreases, again due to increased interference (Figure 4.16). As in the earlier result shown in Figure 4.15, the system with bTAPs appears more sensitive to the changes of the pathloss exponent, since there are more sources of interference due to simultaneous scheduling. However, the system with bTAPs is more robust to the increased lognormal fading standard deviation, due to the additional macro diversity offered by bTAPs. Thus, the deployment of bTAPs is more effective with higher pathloss exponent and higher lognormal fading standard deviation.

The Ercog-Greenstein model pairs the pathloss and the lognormal standard deviation into the terrain types based on measurement data. From Terrain A to C, the pathloss exponent decreases (from 4.8 to 4.12), which generally degrades sys-
Figure 4.17: QPSK 1/2 outage for (1,6,6). Terrain A is representative of hilly areas with moderate-to-heavy tree density and Terrain C is representative of flat areas with light tree density.

System performance, and the lognormal fading standard deviation decreases (from 10.6 dB for Terrain A to 8.2 dB for Terrain C), which generally improves system performance. The two effects, those of the pathloss exponent and the standard deviation of lognormal fading, roughly cancel each other out, and the outage and the sector throughput are largely similar between Terrain A and C, as shown later in Figures 4.17, 4.22, 4.18, and 4.23.

4.5.10 Coverage Improvement

Figures 4.17 shows the outage percentage with and without bTAPs, for both downlink and uplink transmissions. The addition of bTAP and multihop forwarding substantially improves the outage performance. If QPSK 1/2 is the most robust
modulation, the (1,6,6) frequency reuse is too aggressive without the use of bTAPs, since the outage, at around 25%, would be unacceptably high. In other words, around 25% of the area in a sector would not be able to communicate using QPSK 1/2 modulation (6 Mbps) or weaker (> 6 Mbps). In order to achieve reasonable outage, such a system would either require more robust, but lower rate, modulations at the cost of system capacity, advanced signal processing techniques with multiple antennas at the cost of increased complexity and power consumption, or less aggressive frequency reuse patterns such as (3,9,3) at the cost of more spectrum and/or substantially lower system capacity. Implementing partial loading, an artificial ceiling on the carried traffic that reduces the interference, enables the use of more aggressive frequency reuse at the cost of lower utilization. However, with the
addition of bTAPs, the outage is kept at an acceptable level (less than 5%) despite the aggressive frequency reuse. In other words, the coverage increases from 75% to greater than 95%. There is no sacrifice of capacity due to multihop forwarding; on the other hand, there are substantial gains in sector throughput.

4.5.11 Throughput Improvement

Figure 4.18 shows the sector throughput with and without bTAPs, for both downlink and uplink transmissions. Deploying bTAPs not only increases the coverage in the sectors but also increases the system throughput. For Terrain A, the system throughput increases from around 15 Mbps to around 22 Mbps, an increase of 46%, whereas for Terrain C, the sector throughput increases from around 15 Mbps to around 19 Mbps, an increase of 26%. This increase is in spite of the fact that there are several clients whose effective throughput decreases, as shown earlier in Figure 4.14. The increase in throughput is directly related to the increase in coverage area. As more of the sector is covered due to the deployment of bTAPs, the average throughput of the sector goes up. Additionally, due to the deployment of bTAPs, a large number of clients get a higher throughput to and from the TAP, when the communication is made via a bTAP (as opposed to communicating directly with the TAP).
Figure 4.19: The (1,3,6) frequency reuse pattern. Some parts of the figure have been shaded to avoid confusion.

4.5.12 A Different Frequency Reuse Pattern — (1,3,6)

Having demonstrated the benefits of deploying bTAPS in the (1,6,6) frequency reuse pattern, I repeated my experiments on a different frequency reuse pattern, (1,3,6) (Figure 4.19). As explained earlier, the notation (1,3,6) means that a reuse group consists of a single cell, represented by the "1". This reuse group, containing only a single cell, uses three orthogonal channels, represented by the "3". Finally, the last "6" in the notation represents the number of sectors that the cell is divided into, implying that there are two $60^\circ$ sectors in opposite directions, both using the same channel. In order to minimize co-channel interference in the reuse group, the cells are rotated at an angle of $60^\circ$ with respect to each other.
Figure 4.20: Analyzed multicell scenario without bTAP for (1,3,6) frequency reuse pattern. All the co-channel sectors have been marked in a different color.

Figure 4.20 shows such a scenario used for analysis of the base scenario, which has no bTAPs deployed. However, to analyze a system with bTAPs, I used the scenario shown in Figure 4.21. The positions of the deployed bTAPs is not dependent on the frequency reuse pattern. Both these figures show scenarios which appear less regular in their frequency reuse pattern than the scenarios shown for (1,6,6) reuse pattern. This apparent irregularity is because Figure 4.20 and Figure 4.21 show the typical rotation of the orientation of sectors across reuse patterns to minimize the co-channel interference. No such rotation is necessary for the (1,6,6) reuse pattern.

In general, the results, obtained in the (1,3,6) frequency reuse pattern, follow the same trend as was shown for the (1,6,6) reuse pattern. Figures 4.22 shows
Figure 4.21: Analyzed multicell scenario with bTAPs for (1,3,6) frequency reuse pattern. All the co-channel sectors have been marked with different colors. There are two channels that are used in the co-channels since the unit being analyzed consists of two facing sectors.

the outage percentage with and without bTAPs, for both downlink and uplink transmissions. Without bTAPs, more than 40% of the area cannot communicate at QPSK 1/2 modulation or weaker, making the system impractical for commercial deployment. With bTAPs deployed in the system, the outage decreases to less than 20%, i.e., the coverage improves from less than 60% to more than 80%. However, a coverage of 80% might still not make the system commercially attractive and the operator might need to allow for even stronger modulation schemes so that more area can be covered, albeit at a lower data rate.

Figure 4.23 shows the sector throughput for the (1,3,6) reuse pattern, with and without bTAPs, for both downlink and uplink transmissions. The overall sector throughput improves from around 10 Mbps to around 15 Mbps, an increase of
around 50%. Since (1,3,6) is more aggressive than (1,6,6), the coverage and the sector capacity are lower than (1,6,6). For example, the downlink sector throughput with the unmodified (1,6,6) reuse system (16.5 Mbps), without any deployed bTAPs, is higher than the downlink sector throughput achieved in the (1,3,6) system, with bTAPs deployed (16 Mbps). However, the overall system capacity compares better since (1,3,6) uses half as much spectrum as (1,6,6). As in the (1,6,6) reuse pattern, changing the terrain characteristics, from Terrain A to Terrain C, does not substantially alter the performance of the system.

4.5.13 Stability Analysis

In this section, I present the results of experiments done to test the effect of changing some of the parameters used in the experiments. In particular, I evaluated the
Figure 4.23: Sector throughput for (1,3,6)

effect of changing the sector length, the transmission power, and the bTAP antenna height. For these experiments, I used only the (1,6,6) frequency reuse pattern.

Figure 4.24 shows, at different sector lengths, the effect that changing transmission power has on the downlink coverage and on the downlink sector throughput. As for all the previously presented results, the TAP antenna was at a height of 30 m and the bTAP antenna was at a height of 15 m.

Figure 4.24(a) the resulting downlink coverage without bTAPs deployed in the system. In this case, the outage remains greater than 22% (i.e., coverage is lower than 78%) even at high transmission power. The outage does not decrease further, because with the increase in transmission power, the interference caused by the TAP at neighboring co-channel sectors also increases, thus nullifying any gains in received signal strength at the clients. At higher transmission powers, the outage
Figure 4.24: The effect of changing transmission power on downlink outage and downlink sector throughput at different sector lengths. The TAP antenna height was 30 m, the bTAP antenna height was 15 m, and the (1,6,6) frequency reuse pattern was used.
does not depend on the sector length, since the outage is dominated by interference. However, at lower transmission powers, the outage increases with the increase in sector length. Thus, for example, a 6 dBm transmission power causes 30% outage for a sector length of 500 m, but a much higher 65% outage for a sector length of 1250 m.

Figure 4.24(b) shows the downlink sector throughput without bTAPs deployed in the system. In this case, the sector throughput is limited to around 16 Mbps, irrespective of the transmission power and the sector length. At lower transmission power though, the sector length does affect sector throughput. For example, at 6 dBm transmission power, the sector throughput is 15 Mbps for a sector length of 500 m, but only 6 Mbps for a sector length of 1250 m.

Figure 4.24(c) shows the downlink coverage with bTAPs deployed. In this case, the outage of the system decreases to around 5% (i.e. coverage of around 95%). As in the case without bTAPs deployed, the performance is dependent on the sector length only for lower transmission powers, but for higher transmission powers (21 dBm and higher), the outage is independent of the sector length.

Finally, Figure 4.24(d) shows the downlink sector throughput with bTAPs deployed. Due to the improvement in coverage described above as well as due to the capability of simultaneous communication, the sector throughput also increases to around 21 Mbps. Again for higher transmission powers, the sector throughput achieved is independent of the sector length, whereas at lower transmission
Figure 4.25: The effect of changing transmission power on uplink outage and uplink sector throughput at different sector lengths. The TAP antenna height was 30 m, the bTAP antenna height was 15 m, and the (1,6,6) frequency reuse pattern was used.
powers, smaller sectors have higher throughput. For example, at 6 dBm transmission power, the sector throughput is 20 Mbps at a sector length of 500 m, but only 11 Mbps for a sector length of 1250 m.

Figure 4.25 shows the results of a stability analysis for uplink communication, similar to the one done for downlink communication (Figure 4.24). The results for the uplink experiments are very similar in nature to the results for the downlink experiments. Without bTAPs, the system exhibits an outage of at least 30% (a coverage of at most 70%), whereas with the deployment of bTAPs, the outage is around 5%. Consequently, without bTAPs in the system, sector throughput is limited to a maximum of 14 Mbps, whereas with bTAPs, the sector throughput increases to around 20 Mbps. Similar to the downlink scenario, the performance is independent of the sector length at higher transmission power, whereas at lower transmission power, smaller sectors exhibit better performance.

Figure 4.26 shows, for a sector of length 1000 m, how the height of the bTAP affects performance at different transmission powers. The TAP antenna was set at a height of 30 m, and the bTAP antenna height was varied from the height of a client antenna (2 m) to the height of the TAP antenna (30 m). Figure 4.26(a) shows how, at different transmission powers, the variation in bTAP antenna height affects the downlink outage, and Figure 4.26(b) shows the effect on uplink outage. At high enough transmission powers (21 dBm and higher), the outage is more or less independent of the bTAP antenna height. For example, for the downlink case, as
Figure 4.26: The effect of varying bTAP height on uplink and downlink outage and sector throughput at different transmission powers. The TAP antenna height was 30 m, the sector length was fixed at 1000 m, and the (1,6,6) frequency reuse pattern was used.
long as the bTAP antenna is at 8 m or higher, and the transmission power is 21 dBm or higher, the outage stays at around 5%. At lower transmission power (9 dBm), the outage is at least 15%.

The overall sector throughput for both the downlink scenario (Figure 4.26(c)) and the uplink scenario (Figure 4.26(d)) is even less affected by the height of the bTAP antenna. Even with a bTAP antenna at 4 m or higher and a transmission power of 21 dBm or higher, the downlink throughput is around 21 Mbps.

These results imply that even with a very low bTAP antenna height (for example, 8 m) it is possible to get the benefits of deploying bTAPs in the system. Such low bTAP antenna mount points would require lower rental charges, and hence the deployment of such bTAPs would be cheap and in line with the design goals set in Section 4.4.1.

4.6 Chapter Summary

In this chapter, I presented a novel and practical wireless mesh network suitable for wide-area wireless access networks such as WiMAX/802.16, HSPA, and CDMA2000 EV-DO, which typically use centralized MAC in licensed and limited spectrum. In order to increase the coverage of each TAP node, I have presented the deployment of low cost booster TAPs. The work examines the capacity and the coverage of a modified wireless mesh network, i.e., augmented with bTAPs, compared to conventional cellular networks under multicell frequency reuse pat-
terns and realistic wireless channel models. The designed architecture enables the spatial reuse of radio resources within a single sector and improves coverage via multihop forwarding through booster TAPs; the booster TAPs are designed for low complexity and economic deployment.

The presented analysis accounts for multicell co-channel interference and radio resource consumption by multihop forwarding links. The results show that in the (1,6,6) and (1,3,6) frequency reuse patterns, the presented architecture provides dramatic improvements in outage performance and a sufficient capacity gain to compensate for the radio resources required for forwarding user traffic via bTAPs. For example, in the case of downlink transmissions in the (1,6,6) reuse pattern, the evaluation shows an outage reduction of around 80%, whereas the sector throughput increases as much as 33%.

I have also studied how changing several parameters affects the performance in a (1,6,6) system. I found that at transmission powers 21 dBm and higher, the performance of the system, in terms of both coverage and throughput, is independent of the sector size. Moreover, bTAP antennas need to be deployed at heights of only 8 m or higher in order to achieve the benefits in coverage and throughput.

These results show that the augmented wireless mesh approach is a promising way to address outage and capacity issues of broadband wireless mesh networks, and enables an economic deployment model for such networks.
Chapter 5

Conclusions and Future Work

In this thesis, I have presented solutions to two important problems in wireless mesh networks: throughput and coverage improvement. I have evaluated these solutions through simulation experiments. My results show that application of my solutions to wireless mesh networks is not only a promising way to address capacity and outage issues of broadband wireless mesh networks, but also enable an economic deployment model for such networks.

Even though the two solutions for throughput and coverage improvement can exist independent of each other in a wireless mesh network and can individually prove advantageous, in Chapter 2, I have described how they can simultaneously coexist and can complement each other in a single network. I have shown how end-to-end communication, from a client to and from the Internet, can be setup. Such end-to-end communication has two parts. First, the client node communicates, either directly or via a bTAP, with a TAP. Second, the TAP, with which a client communicates, completes the end-to-end communication path by using traffic-aware routing metrics to locate a gateway TAP (and the corresponding route to the gateway TAP) among possibly multiple gateway TAP nodes. I have presented a scheduling scheme that can coordinate the necessary transmissions that
must take place at different stages of end-to-end connections between clients and the Internet. The scheduling of the such communication is split into two periods, a sector period and an inter-TAP period. In the sector period, a client node communicates with its source TAP, and in the inter-TAP period, the source TAP communicates with a selected gateway TAP via other TAP nodes. Sector periods that operate on a different logical channel than the inter-TAP period of the source TAP can simultaneously operate with the inter-TAP period.

5.1 Throughput Improvement

In Chapter 3, I have presented a number of novel traffic-aware routing metrics (MLS, MLS-T, and AVAIL) that take into account existing client traffic flows in the network. Previous routing metrics have been traffic-unaware, often causing routes with poor throughput to be selected even when other better routes are available. Since wireless mesh networks have the potential to provide widespread network connectivity, the routing metrics designed for such networks must account for existing traffic, which is not done by current routing metrics. These new traffic-aware metrics use information captured through measurements at the MAC layer, which are then exposed to the routing layer. Given a node, the MLS metric measures the fraction of time that the node experiences to be free. The AVAIL metric is based on an analytical model of the IEEE 802.11 MAC protocol. At any node, the model uses measurements, such as arrival rate of packets for each outgoing link from
that node, to calculate estimated throughput along the outgoing link. Finally, the MLS-T metric combines both the MLS and the AVAIL metric.

I have compared the new traffic-aware metrics with traffic-unaware metrics, such as ETX and IRU. I have evaluated, under different network scenarios, a topology from a real residential high-speed wireless mesh network commercialized by Chaska.net, and a synthetic grid-like topology, called the Manhattan topology. I found that the Chaska topology is severely limited by interference. Still, the use of traffic-aware metrics provides higher throughput than that of traffic-unaware routing metrics. The traffic-aware metrics show an improvement of up to 8% when omni-directional antennas are used, and an improvement of up to 12% when directional antennas are used. Across all metrics, the use of directional antennas improves the average achieved throughput by only 15%, compared to the average throughput achieved when using omni-directional antennas.

However, in a more structured Manhattan topology, the use of the traffic-aware metrics, i.e., MLS and AVAIL, led to the selection of routes that provided higher throughputs than routes selected with the use of traffic-unaware metrics. With a single additional flow in the network, compared to the average throughput obtained by using the traffic-unaware metrics, the use of traffic-aware metrics improve the average throughput by up to 40%, when omni-directional antennas are used, and up to 50%, when directional antennas are used. Even with multiple additional flows, the improvement in average throughput provided by the use of
traffic-aware metrics are similar in trend to the improvements achieved for a single additional flow. In this structured topology, all of the routing metrics show a marked improvement in the average throughput achieved when directional antennas are used. This improvement is 28\%-45\% for the traffic-unaware metrics and 25\%-50\% for the traffic-aware metrics. Also, with the use of traffic-aware metrics, the gain in performance, measured by throughput achieved by flows inserted into a network with a large number of existing flows, increases with the increase in contention in the network.

I further evaluated the performance of the different metrics for shortlived flows arriving at various rates in the Manhattan scenario. I found that the traffic-aware metrics provide benefits similar to those obtained in experiments with a single additional flow and multiple additional flows. Also, the benefits increase as the rate of flow arrival decreases. For example, with one flow arriving every 2 seconds, and with the use of directional antennas, the AVAIL metric shows up to 35\% higher average throughput than the throughput achieved by the traffic-unaware metrics. Even in the case of multiple flows simultaneously arriving at the network, the use of traffic-aware metrics provides higher average throughput (and lower standard deviation) as compared to those achieved by the use of traffic-unaware metrics. For example, the use of the AVAIL metric leads to around 31\% higher average throughput than that obtained by using the traffic-unaware metrics, whereas the MLS-T metric achieves around 26\% higher average throughput.
A quantitative analysis of the Chaska and the Manhattan topologies corroborates the fact that the Chaska topology inherently has higher interference than the more structured Manhattan topology. Moreover, the average number of gateways accessible within some fixed number of wireless hops is higher in the Manhattan topology than in the Chaska topology. This difference is in spite of the fact that even though both the topologies have a similar number of total nodes, the Manhattan topology has just 10 total gateways, compared to 14 total gateways in the Chaska topology.

5.2 Coverage Improvement

In Chapter 4, I have presented the design and analysis of a new technique for increasing the coverage of a wireless mesh network through deployment of low-cost booster TAPs (bTAPs). These bTAPs are strategically deployed and controlled by the system operator to wirelessly forward traffic between client nodes and TAP nodes. Clients that have a weak direct signal from the TAP can communicate with the TAP using the bTAPs. This deployment model is especially suitable for wide area wireless access networks that use centralized management of radio resources, such as WiMAX/802.16, HSPA, and CDMA2000 EV-DO. I have analyzed the use of bTAPs across different frequency reuse patterns typical of those used in multicell wireless environments for efficient management of costly radio spectrum. My analysis has accounted for multicell co-channel interference and radio resource
consumption by multihop forwarding links. The results show that in (1,6,6) and
(1,3,6) frequency reuse patterns, the bTAP architecture provides dramatic improve-
ments in outage performance and a sufficient capacity gain to compensate for the
radio resources required for forwarding client traffic via bTAPs. For example, in
the downlink communication case in the (1,6,6) reuse pattern, the outage is reduces
by around 80%, whereas the sector throughput increases by as much as 33%.

I have studied how changing several parameters affects the coverage and
throughput improvement of the proposed bTAP deployment in a (1,6,6) system.
I found that at transmission powers 21 dBm and higher, the performance of the
system, in terms of both coverage and throughput, is independent of the sector
size. Moreover, bTAP antennas need to be deployed at heights of only 8 m or
higher in order to achieve the benefits in coverage and throughput.

5.3 Future Work

I have presented a complete picture in which the solutions presented in this thesis
can be applied to solve the problems of selecting higher throughput routes and of
coverage improvement in wireless mesh networks. Nevertheless, there are several
future directions in which this work might be augmented.

The effect of using a TDMA-based MAC protocol for communication between
TAP nodes could be an interesting study. Since the TAP nodes are probably go-
ing to be mounted on high outdoor towers, each TAP node can be fitted with a
GPS receiver, which would allow time synchronization among TAP nodes. This synchronization would in turn allow the use of a TDMA-based MAC protocol for inter-TAP communication. Among other things, the use of a TDMA-based MAC protocol would simplify the measurement of the MAC layer share of each node, since free slots can be easily accounted for.

A large number of commonly used applications use the Transmission Control Protocol (TCP), or some variation of it, as the underlying transport protocol. The effect of the different routing metrics, traffic-aware as well as traffic-unaware, on the performance of the TCP needs to be studied. Since the traffic-aware metrics, apart from selecting routes, also predict the estimated throughput across the selected route, this estimation could possibly be integrated with TCP's window management. This integration might allow TCP to ramp up to a sustainable data rate faster than is possible with TCP's current slow-start and congestion avoidance mechanisms. This speedup could be an important improvement to TCP, especially in high bandwidth wireless networks (such as wireless mesh networks), where TCP is known to perform badly.

I have presented only a proactive routing protocol in this thesis. However, proactive protocols might waste resources keeping network status updated even if there is no need of such updated information. Hence, in order to decrease the overhead associated with a proactive routing protocol, a hybrid routing protocol, which borrows techniques from both proactive and reactive routing protocols, can
be designed. A purely reactive routing protocol is unsuitable for traffic-aware routing metrics because such protocols do not search for newer (and possibly better) routes if a usable route, albeit not a good one, is already known. Hence, such a routing protocol might continue using a route with lower throughput and/or higher delivery latency even though a better route might have come into existence since the current route was discovered.

The final and most convincing test for any routing metric that is suggested to be promising by analytical examination and simulation-based evaluation is to test the metric on real wireless mesh network deployments. An interesting study would be to evaluate different types of metrics, not limited to the ones analyzed in this thesis, on actual wireless mesh deployments.
Bibliography


[82] X. Wu, S. H. Gary Chan, and B. Mukherjee. MADF: A Novel Approach to Add an Ad Hoc Overlay on a Fixed Cellular Infrastructure. In IEEE Wire-


