EFFICIENT RESOURCE UTILIZATION IN
ADVANCED WIRELESS NETWORKS

by

Ouldooz Baghban Karimi

M.Sc., Iran University of Science and Technology, 2006
B.Sc., University of Tehran, 2003

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences

© Ouldooz Baghban Karimi 2013
SIMON FRASER UNIVERSITY
Summer 2013

All rights reserved.
However, in accordance with the Copyright Act of Canada, this work may be
reproduced without authorization under the conditions for “Fair Dealing.”
Therefore, limited reproduction of this work for the purposes of private study,
research, criticism, review and news reporting is likely to be in accordance
with the law, particularly if cited appropriately.
## APPROVAL

<table>
<thead>
<tr>
<th>Name:</th>
<th>Ouldooz Baghban Karimi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree:</td>
<td>Doctor of Philosophy</td>
</tr>
<tr>
<td>Title of Thesis:</td>
<td>Efficient Resource Utilization in Advanced Wireless Networks</td>
</tr>
<tr>
<td>Examining Committee:</td>
<td>Dr. Petra Berenbrink</td>
</tr>
<tr>
<td></td>
<td>Chair</td>
</tr>
<tr>
<td></td>
<td>Dr. Jiangchuan Liu, Senior Supervisor</td>
</tr>
<tr>
<td></td>
<td>Associate Professor</td>
</tr>
<tr>
<td></td>
<td>Dr. Qianping Gu, Supervisor</td>
</tr>
<tr>
<td></td>
<td>Professor</td>
</tr>
<tr>
<td></td>
<td>Dr. Mohamed Hefeeda, SFU Examiner</td>
</tr>
<tr>
<td></td>
<td>Associate Professor</td>
</tr>
<tr>
<td></td>
<td>Dr. Lin Cai, External Examiner</td>
</tr>
<tr>
<td></td>
<td>Associate Professor, Electrical &amp; Computer Engineering</td>
</tr>
<tr>
<td></td>
<td>University of Victoria</td>
</tr>
</tbody>
</table>

Date Approved: ____________________________
Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the “Institutional Repository” link of the SFU Library website (www.lib.sfu.ca) at http://summit/sfu.ca and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author’s written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011
Abstract

With advancements in wireless communication technologies, broadband wireless services will be prevalent in the near future. Meanwhile, increased capability of mobile devices is drastically increasing the mobile data usage. This increase is far in excess of mobile network capacities. Therefore, despite the higher availability of these networks, higher number of users they support, and their improved spectral efficiencies, effective utilization of wireless resources is required to keep up with the ever-increasing user demands for mobile content.

This thesis targets high-throughput data transmission in advanced cellular wireless networks and wireless local area networks. These wireless networks have been widely used for broadband wireless access and are constantly enhanced for future applications. We present efficient resource allocation solutions to meet the transmission requirements of high bandwidth applications, like video streaming, in these networks. Our solutions combat bandwidth limitations in different settings, including dense networks and high mobility. We compare our proposed methods with state-of-the-art solutions and prove their effectiveness.

We also discuss the availability of cost-effective wireless access solutions that are yet to be considered for expanding mobile data networks. If efficiently used, these networks can offload considerable traffic off the cellular wireless networks. We discuss dense wireless local area networks for their abundant resources and extensive deployment. We present an optimal solution to opportunistically use the dense deployment of local access points and collaboratively serve wireless users. We present our solution within a set of networks that share the same upstream provider. Our solutions show significantly improved throughput in dense implementations, a remarkable step towards pervasiveness in next generation of mobile communication systems.
To Rahimeh, Reza, Sevil, and Milad!
“As for me, all I know is that I know nothing.”
— Socrates
Acknowledgments

While I appear as the only author of this dissertation, many people contributed to it in different ways and supported me along the way. I would like to thank everyone for their help and support in the past four and half years.

First of all, thanks to my advisor, Prof. Jiangchuan Liu, for being a great mentor during this process. He influenced me by his research style, his focus, and his unwavering help and support. He kept me focused and motivated with his timely guidance, while letting me make my own mistakes, through which I found my own way in scientific research. I would also like to thank him for facilitating collaborations, and his confidence in my success.

Many thanks to Prof. Qianping Gu for helping me with his suggestions to find my research direction during my proposal stage, and his help and support along the way.

I would like to thank Prof. Jennifer Rexford. I enjoyed our collaboration and learned in every second of it. I hope this collaboration has also made my personality at least a bit like hers.

Many thanks to Prof. Zongpeng Li, who helped me take the first steps in my research, for always being a great help when I needed it along the way. Also, thanks to Dr. Chonggang Wang, and Prof. Dan Wang for our collaborations.

I also want to thank Prof. Mohamed Hefeeda, Prof. Lin Cai, and Prof. Petra Berenbrink for serving as my thesis examiners. I appreciate their valuable comments.

Finally, thanks to my family who have always supported me.

Thank you all, this was a great journey.
# Contents

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approval</td>
<td>ii</td>
</tr>
<tr>
<td>Partial Copyright License</td>
<td>iii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Quotation</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vii</td>
</tr>
<tr>
<td>Contents</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
</tbody>
</table>

## 1 Introduction

1.1 Cellular Wireless Networks .............................. 2
1.2 Pervasiveness and Offloading ............................ 3
1.3 Wireless Local Area Access ............................... 3
   1.3.1 Wireless Mesh Networks ............................... 4
   1.3.2 Collaborative Wireless Local Area Networks ......... 5
1.4 Thesis Contributions .................................... 6

## 2 Background and Related Work

2.1 Resource Management in Cellular Wireless Networks .... 8
5.3.4 The Channel Assignment Sub-problem ........................................ 56
5.3.5 The Routing Sub-problem ......................................................... 59
5.4 Performance Analysis ............................................................... 61
5.5 Summary .................................................................................... 64

6 Collaborative Wireless Local Access ........................................... 66
  6.1 Motivating Example ................................................................. 68
  6.2 Collaborative Wireless Access .................................................. 69
  6.3 Access Model and Notations ...................................................... 71
  6.4 Collaborative Access Point Association ..................................... 72
    6.4.1 Single Access Point Association ............................................ 73
    6.4.2 Multiple Access Point Association ......................................... 75
    6.4.3 Channel Assignment ............................................................ 77
  6.5 Performance Evaluation ............................................................ 78
    6.5.1 Data Collections ................................................................. 78
    6.5.2 System Configuration .......................................................... 80
    6.5.3 Evaluation Results and Analysis ............................................ 81
  6.6 Summary and Discussion .......................................................... 83

7 Conclusion and Outlook .............................................................. 86
  7.1 Contributions ............................................................................ 86
  7.2 The Big Picture and Future Outlook .......................................... 87

Bibliography .................................................................................... 89
List of Tables

3.1 Power-Efficient Multicast Formulation Variables ........................................ 15
4.1 Summary of the Variables ............................................................................. 38
4.2 Simulation Settings ....................................................................................... 41
5.1 Variables of Channel Assignment and Routing Sub-problems ...................... 53
6.1 Summary of the Variables ............................................................................. 71
List of Figures

3.1 Block Diagram of Solution Components and Their Relations .......................... 18
3.2 Sleep Potential for User(5MHz) ................................................................. 23
3.3 Sleep Potential for User(20MHz) ................................................................. 23
3.4 Average Delay (5MHz) .................................................................................. 24
3.5 Average Delay(20MHz) ................................................................................ 24
3.6 Average Delay (5MHz) .................................................................................. 25
3.7 Average Delay (20MHz) ................................................................................ 25
3.8 Transmitted Data (5MHz) ............................................................................. 25
3.9 Transmitted Data (20MHz) ........................................................................... 25
4.1 System Settings for LTE Network Access for High Speed Trains ................. 29
4.2 A Number of LTE Femtocells Can be Used to Cover Each Train Cabin .......... 30
4.3 Cell Array (CA) Reconfiguration Along the Railway Path ............................ 31
4.4 PHHO and PSSO Handover for High Speed Trains with Different Initiation Requests 33
4.5 Signal To Noise Ratio Throughout Cells ...................................................... 35
4.6 High Speed Train Movement Path ............................................................... 40
4.7 Urban Cell, Speed: 350kmph, Path:1. (a) Delay (ms) (b) Throughput (Mbps) (c) Handover Latency (ms) (d) SINR (dBm) (e) Handover Requests ......................... 41
4.8 Urban Cell, Speed: 350kmph, Path:3. (a) Delay (ms) (b) Throughput (Mbps) (c) Handover Latency (ms) (d) SINR (dBm) (e) Handover Requests ......................... 42
4.9 Urban Cell, Speed: 200kmph, Path:2. (a) Delay (ms) (b) Throughput (Mbps) (c) Handover Latency (ms) (d) SINR (dBm) (e) Handover Requests ......................... 42
4.10 Urban Cell, Speed: 350kmph, Path:2. (a) Delay (ms) (b) Throughput (Mbps) (c) Handover Latency (ms) (d) SINR (dBm) (e) Handover Requests ......................... 43
Chapter 1

Introduction

During the last couple of decades, advances in wireless network technologies have remarkably improved user experience. Sharing through only few wired interconnection of personal computers has been changed to ubiquitous wireless and mobile access through hand-held devices. Wireless networks continue to improve the user experience in numerous different ways by enabling high throughput data and multimedia applications.

Wireless networks differ fundamentally from wired networks due to unpredictable and shared nature of the wireless medium. Wireless radio spectrum is a scarce and expensive resource that must be shared among many different wireless systems using it for communications. This sharing is regulated by licensing blocks of the wireless spectrum for certain applications, e.g., cellular networks, and managing other parts openly for public usage. Licensed spectrum is controlled by regulatory bodies regionally and globally in attempt for cost-effective usage of it. The unlicensed and open spectrum, although abundantly and freely available for use, is also expensive due to short range and high interference if networks not deployed effectively. Therefore, a great number of researches are dedicated to spectrum efficiency and resource allocation to achieve higher throughput and lower costs through efficient utilization of wireless resources.

In this thesis, we discuss efficient resource utilization in advanced wireless networks. We investigate the advancements in the wireless technology with focus on resource utilization for high-bandwidth applications and provide effective solutions. We refer to wireless network technologies that are already in use and continuously improving towards accommodating future applications as advanced wireless networks. We investigate the advancements in the wireless network technologies with focus on three promising wireless solutions: LTE-based cellular wireless networks, wireless mesh networks (WMNs), and collaborative wireless local access networks.
CHAPTER 1. INTRODUCTION

1.1 Cellular Wireless Networks

Cellular wireless networks provide wide-area wireless access to mobile users. Long term evolution (LTE) of third generation partnership project (3GPP) is a promising radio access technology for mobile communication systems. High throughput and low latency in radio access network and connection setup are some of the reasons for using LTE, and LTE-advanced. Spectral and power efficiency, enhanced peak data rates, finalization of LTE-advanced in 2011, and cost-effectiveness of adoption for service providers are also among the reasons behind LTE being the technology of the choice [98] [86]. Another important improvement in LTE includes simplifying the architecture of the system, from the existing circuit and packet switching combined network, to an all-IP flat architecture system.

LTE is a mobile broadband system with enhanced multimedia services built on top as an upgrade towards fourth generation of wireless networks. It is targeting optimum packet switched services in data rate, throughput, spectrum efficiency, latency, bandwidth, multimedia services, cost, and mobility. Peak data rates target 100 Mbps and 3 Gbps in downlink and 50 Mbps and 1.5 Gbps in uplink in LTE, and LTE-advanced respectively. It includes a peak download rates of 326 Mbps for 4 × 4 antennas, and 172 Mbps for 2 × 2 antennas using 20 MHz of spectrum, and peak upload rates of 86 Mbps for every 20 MHz of spectrum using a single antenna. At the current stage of development it can support at least 200 active users with data usage in every 5 MHz cell spectrum. LTE provides increased spectrum flexibility by using multiple blocks of spectrum allocation [1].

Even though spectral efficient, advanced cellular networks need even more enhanced spectrum solutions to accommodate new applications with higher data rates. Enhanced multimedia broadcast and multicast services (eMBMS) is a video streaming and download solution provisioned for LTE-advanced [7]. Being under development with no current commercial deployments, eMBMS targets high throughput spectral efficiency through multicast. In this thesis, we investigate eMBMS and propose a practical resource management solution to further enhance its user experience and power consumption.

Another challenge in advanced cellular wireless networks is handling high-throughput connections in high-mobility. Although Doppler effect [141] is fairly controlled in physical layer of LTE networks, the connection transfer and handling is not as fast in handovers among the calls. This challenges managing high-throughput connections and the effect of handover latency on them. We propose a Cell Array architecture that solves this problem together with our predictive handover and resource scheduling solutions.
CHAPTER 1. INTRODUCTION

1.2 Pervasiveness and Offloading

Mobile data traffic has grown 70% to reach 885 petabytes in 2012 and will continue to grow 13-fold by the end of 2017. Mobile video traffic have been a large part of this increase, consisting over 50% of traffic by the end of 2012 [40] and increasing to be two thirds of traffic by the end of 2016 [39]. Even though the 3G and 4G cellular networks have increased data bandwidth by adopting new technologies and also can support a higher number of users in each cell, the limited spectrum cannot accommodate the number of current high bandwidth users. Also, improvement of bandwidth in the new technologies is not happening as fast as the increase in the user traffic demand.

Fortunately, other wireless network technologies can come to the rescue. Mobile devices with different wireless interfaces are increasingly being used for high-bandwidth connections. With faster growth of bandwidth usage for mobile users compared to cellular technology advances, switching among different interfaces is considered a practical solution for scarce spectrum problems. The hand-held mobile devices have the ability to switch to local wireless networks whenever available using vertical handover methods [97] [8]. This introduces new opportunities for migration of high-bandwidth loss-tolerant applications like video streaming from the cellular networks to other available wireless networks, known as mobile data offloading.

Mobile data offloading is one of the key concepts toward fifth generation of wireless communications. Providing ubiquitous computing through pervasive networks, where the user can simultaneously be connected to several wireless access technologies and seamlessly move between them, is one of the main goals in fifth generation of wireless communications. This much needed offloading technology is far from mature and needs further research. We believe the first step towards these networks is identifying and effectively utilizing the resources in target technologies, e.g., 802.11-based wireless networks, to increase the capacities they offer for offloading purposes.

1.3 Wireless Local Area Access

Wireless local area network access points provide local access, usually less than a 100 meters, to a handful of devices. Current wireless local area networks are usually 802.11-based products, commercially known as WiFi. They provide local wireless access and access to backbone network and internet through hot spots, local access points, or peer-to-peer wireless connection to a selected set of access point that have internet and backbone network connection.

Technology improvements include transition from 802.11b/g to 802.11n to increase the data
rates with multiple input multiple output antennas and the use of four spatial streams. This follows a recent update towards 802.11ac standard working in 5 $MHz$ band in order to further increase the data rates and coverage.

The popularity of local area wireless networks is due to their ease of installation, and cost effectiveness. However, these networks are usually poorly deployed. Installations typically focus on ensuring space coverage, rather than capacity or quality of service. Therefore, despite high data rates these standards offer, many users commonly experience significant performance degradation problems.

The simplest medium access control (MAC) for wireless local access is the single channel MAC, where all infrastructure and client mesh nodes should use the same frequency channel to communicate. Efficiency of such MAC layer algorithms and protocols can be improved by carrier sensing multiple access with collision avoidance (CSMA/CA), contention window size improvements, and modification of back-off procedures. However, these solutions show limited improvements since the high probability of contentions is a potential hinderance in leveraging the end-to-end throughput. Directional antennas, and power control schemes can also help improve the situation. Directional antennas direct the communication to the target neighbour and away from other receivers. Power control schemes decrease the transmission power levels to create spatial reuse opportunities [79] [77]. Both of these schemes are carried out in physical layer, need better interoperability among the layers, and may result in increased hidden terminal problem.

1.3.1 Wireless Mesh Networks

Wireless mesh networks are multi-hop peer-to-peer networks where wireless nodes are connected with mesh interconnections and cooperate with one another to route packets. Most of wireless mesh nodes are not connected to wired network infrastructure. They provide wireless connection over a fairly large area with low transmission power and cost using multi-hop wireless architecture [5]. They can be used for various applications such as extending the coverage of existing wireless networks or connecting remote users to the Internet where there is no wire-line backbone.

Wireless meshes can be implemented using different types of wireless interfaces, among which 802.11-based networks are the most popular technology. These networks are composed of two types of wireless nodes: mesh clients, and mesh routers. Mesh routers have minimal mobility, and perform dedicated routing and configuration. They are able to establish and maintain the mesh connectivity to form a self-organized mesh. Mesh routers can have multiple wireless interfaces to facilitate
CHAPTER 1. INTRODUCTION

multi-hop connection. Mesh clients usually have a single wireless interface with lightweight mesh communication functionalities. Mesh clients can maintain connectivity with minor mobility. They are wireless and mobile devices like laptops and cell phones [5].

Mesh routers can be classified into three sub-categories: mesh gateways, static relays, and access points. Mesh routers with gateway/bridge functionality can integrate wireless meshes to other wired and wireless networks and provide connection to the Internet. Static relays are the nodes which forward traffic to or from other mesh nodes. Access points are the nodes providing localized first-hop connectivity to mesh clients.

Wireless mesh networks provide a promising solution for pervasive and cost-effective broadband wireless connection [5]. This is due to their flexibility and cost-effectiveness in bringing a large number of users online [9]. Scalability, connectivity, quality of service, security, ease of use, compatibility, and interoperability are the main characteristics expected from these networks.

Nowadays, 802.11-based multi-channel MAC protocols are commonly used for wireless meshes to improve network capacity and performance. Mesh nodes can use different orthogonal or overlapping frequency channels to transmit data in the network with lower interference. This improves performance at the cost of complicating the routing layer functionalities.

Handling multiple channels is especially important in multicast routing [139]. Multicast applications need simultaneous transmissions to a number of nodes in a single network. This means higher number of nodes sharing the wireless medium at the same time. Therefore, in addition to gateway selection and multicast routing, the channel selection on each node is important in overall throughput of multicast in the mesh networks. There are existing solutions as well as continuing research to support multimedia multicast over wireless mesh networks. These solutions range from fundamental approaches like shortest path trees, minimum cost trees [93], and network coding [133] to mixed cross-layer solutions [138].

In this thesis, we focus on resource utilization efficiency of multicast routing in multi-channel wireless mesh networks. We investigate cooperative gateways and channel assignment effects and present a cross-layer design that jointly optimizes the multicast routing and channel selection.

1.3.2 Collaborative Wireless Local Area Networks

Wireless local are networks are popular for home and enterprise network access. In these networks, in contrast to wireless mesh networks, access points and hot spots are usually connected through a link to a backbone network infrastructure or an Internet service provider. In addition to increased
CHAPTER 1. INTRODUCTION

data rates with enhanced standards, deployment and management of these networks with focus on channel assignment, power management, and managing association decisions are important for better performance of these networks. Cooperative networking with relay deployment and network cooperation in access optimization [21] or mobility [134] have been proposed to improve WiFi access experience.

We discuss resource management in access point association. We investigate the available solutions [3] [50], applications [74], and their limited availability of current deployments for offloading purposes [22] despite the popularity and excessive use of the technology. We focus on dense urban areas and discuss the problem with current solutions in covering these areas. We discuss access point availability and association management and provide optimal solutions. We propose collaborative wireless local area access and investigate its benefits in dense urban areas.

1.4 Thesis Contributions

First we discuss LTE cellular wireless networks and examine power-efficient high-quality scalable video streaming. We consider scalable video streaming and download services offered by multimedia broadcast and multicast services (eMBMS). We propose an effective and practical solution to enhance user experience and power consumption in both user device and base station. To perform power efficient multimedia transmission, we face three key trade-offs: maximizing energy saving versus minimizing delay, maximizing sleep time versus minimizing lost packets, and maximizing quality of video versus minimizing unnecessary video transmissions. We provide a balanced solution that addresses the trade-off by including user preference. Our simulation results indicate 13% to 25% improvement in user device power conservation chances. The provided solution also decreases the transmitted data in the network while preserving the user perceived quality of the video.

We also consider high mobility users, and present a novel LTE-based solution to support high throughput and continuous multimedia services for high speed train passengers. Our solution is based on a CellArray design that smartly detects and organizes the cells along a railway, together with a femto cell service that aggregates traffic demands within individual train cabins. Given the movement direction and speed of a high-speed train, cell array effectively predicts the upcoming LTE cells in service, and enables a seamless handover that will not interrupt multimedia streams. To accommodate the extreme channel variations, our scheduling and resource allocation mechanism maximize the service rate based on periodical signal quality changes. Our proposed solution decreases the handover delay by up to 18% and improves user throughput by up to 35%. It also well
resists to network and traffic dynamics, thus enabling uninterrupted quality multimedia services for passengers in high speed trains.

For wireless mesh networks, we present a cross-layer design that jointly optimizes the multicast routing and channel selection. Our algorithm computes the optimal multicast flows from multiple cooperative gateways and selects appropriate channels for each mesh node to use at judiciously tuned power. We iteratively optimize the design through Lagrange relaxation and primal-dual decomposition. We introduce progressive channel assignment and power level adjustment in the MAC/physical layer, together with a smart link capacity allocation for cooperative gateways in the network layer. Our multicast algorithm improves performance by up to 100% over hop count routing and orthogonal channel assignment in multi-channel multi-hop wireless meshes.

Finally, we present our design for a collaborative network with centralized access point association. We propose using a share of each access point’s bandwidth to serve non-local users and extend the conventional proportional fair association through effective sharing and collaboration. Our model includes co-channel access points, sharing, and collaborative access in network utility maximization framework. We present optimal solutions for multiple access points to collaboratively serve wireless users within a set of networks that share the same upstream provider. We present measurement data from residential buildings to verify our design choices and show that our solution can improve the overall throughput by up to 40%.

The rest of this dissertation is organized as follows. Chapter 2 provides background information on target technologies and briefly presents a review of the related work. Chapter 3 presents our power efficient multimedia multicast algorithm for LTE cellular networks. In chapter 4, we present our cell array design, predictive handover mechanism, and scheduling solution that guarantees seamless connection in LTE cellular networks for high speed train passengers. In chapter 5, we present our multicast solution for high-throughput multicast with efficient resource utilization in multi-channel wireless mesh networks. Chapter 6 presents our proportional fair access point association in dense collaborative local area networks. Finally, chapter 7 concludes this dissertation, puts the bits and pieces of our work together into a big picture, and presents our outlook to the future.
Chapter 2

Background and Related Work

There has been significant research on efficient resource utilization in advanced wireless networks in recent years. The solutions range from physical layer wireless medium usage optimization and MAC layer standard upgrades to cross-layer system designs and resource allocation optimizations. In this chapter, we briefly review and categorize the related work on advanced wireless networks.

2.1 Resource Management in Cellular Wireless Networks

There have been a great number of researches on LTE cellular wireless networks. We focus on resource allocation solutions for multimedia multicast services, and service continuity for high-mobility users.

2.1.1 Multimedia Services

There are two different categories of solutions trying to alleviate power consumption in wireless multimedia multicast without degrading multimedia quality. First, application layer video multicast enhancement solutions. These solutions mainly discuss video coding, layering, and video transmission schemes for wireless multicast services. A well-known set of solutions in this category propose transmission of layered video to improve quality for users with better reception [108]. Base layer is sent in low rate, with high transmission power to be available to all users in the cell. Enhancement layers are coded in higher rates and transmitted with lower power. This will decrease power consumption and increase perceived quality for receivers with better signal quality.
Second set of solutions are power efficient multicast solutions. These solutions range from multicast beamforming in physical layer [65] [121], and scheduling in MAC layer [121] [132], to cooperative and opportunistic routing in network layer [81]. Multicast beamforming solutions work on efficient use of the spectrum by steering power in the directions of multicast subscribers [80] [115]. This will minimize leakage and interference in other directions [116] [65]. However, feasibility is the key concern in beamforming context due to power or mutual interference limitations.

Efficient scheduling mechanisms on LTE downlink can also alleviate the problem. There is a trade-off between multicast gain and multi-user diversity in multicast scheduling in time-slotted cellular wireless networks. This trade-off is considered as the main concern in many scheduling algorithms. There are different researches on enhancing multicast scheduling in LTE networks [49] [128]. Multicast and broadcast services over single frequency network (MBSFN) multicell cooperative solutions are also provided as a part of LTE standards. However, longer cyclic prefix and synchronization complexities dictated in single frequency multicast nullifies cooperative gains.

2.1.2 Service in High Mobility

Satellite communications were traditionally used for wireless access over vehicles moving across vast geographical areas [41]. The satellite service however would be disconnected in tunnels or at terminals. Solutions that adaptively switch to wireless local area networks [57] or distributed antenna systems (DAS) [27] [123] in none line of sight (NLOS) places to enhance connectivity have been proposed. Nevertheless, the satellite communications remain expensive, not to mention the severely limited bandwidth, the high propagation delays, and the high overhead with the enhancements for NLOS.

A recent proposal expands the use of heterogenous wireless links to provide continuous slow connections and intermittent fast connections for high-speed vehicles [4]. Given the vast geographical coverage of the high speed train railway, deploying such a new network infrastructure can be costly.

The availability of wireless cellular network in both urban and rural areas, which eliminates the need for implementation of a new infrastructure for high speed trains, makes it a better alternative. Digital wireless broadcasting with global system for mobile communications - railway (GSM-R) for high speed train passengers has been demonstrated by Gui et al. [56], which complies the Chinese digital TV broadcasting standard [118].

For high-quality general purpose wireless accesses, a series of measurements have shown that,
even with such new generation of standards as high speed packet access (HSPA) [125], 3G, and 3.5G, remarkable performance degradation still occur over high speed trains, particularly when the speed is over 250 km/h [63]. Including the user mobility information in link adaption for LTE to estimate the signal quality to choose the modulation and coding scheme (MCS) and multiple input multiple output (MIMO) options have been studied [103]. The latest wireless standard, LTE-advanced, could maintain quality links for speed up to 350-500 km/h. Yet, unlike a satellite that covers a huge area, an LTE cell is relatively small. Highly frequent handovers and fast variations in signal quality thus become severe problems that must be addressed [68].

There are proposals to enhance the handover latency for high speed trains. A quick review of handover schemes in LTE-advanced and WiMAX networks is presented by Kim et al. [68]. A three tier network topology based on heterogeneous wireless networks for seamless wireless connectivity has been proposed. They also propose a probabilistic fast handover procedure with reduced latency as compared to normal IPv6 handoff [73], but it still has higher average handoff latency compared to normal LTE seamless handover and WiMAX EBB handover mechanism.

On the other hand, moving extended cell (MEC) [101] is proposed as a solution to accommodate mobility in radio over fiber (RoF) networks. RoF is a new paradigm for high speed wireless access at 60GHz [101]. The line of sight requirement of 60GHz signal and air propagation losses limit the cell coverage in these networks, which complicates the access at high speeds. Our cell array solution is initially inspired by the MEC concept, though ours is based on LTE femto-cell [51].

2.2 Resource Management in Local Wireless Access

2.2.1 Wireless Mesh Resources

Wireless mesh networks [5] and resource management in these networks have been focus of many researches in recent years [10] [104]. Early researches focus on preserving quality of service in integration of these networks with wired networks [23], and benefits of multiple gateway [105] and multiple radio node [70] deployments in these networks. Minimizing wireless mesh installation costs [10], gateway placement [58] [30], router placement [131], and relay selection [117] have been discussed in literature.
CHAPTER 2. BACKGROUND AND RELATED WORK

Routing and Multicast

Different routing metrics have been proposed and used for multi-hop routing in wireless mesh networks. These routing metrics include hop count, link quality, expected transmission time, expected transmission count, per-hop round trip time, and weighted cumulative expected transmission time [45]. Based on these routing metrics, multi-radio routing [46] [45], multi-path routing, hierarchical routing [25], and geographic routing [47] have been proposed.

A notable challenge in wireless mesh networks is to provide support for multicast applications. Such applications usually serve a large number of users, and consume high network bandwidth. Providing high-throughput multicast applications in wireless meshes is challenging due to the interference and the change of channel quality. Nguyen and Xu [93] systematically compared the conventional minimum spanning trees and shortest path trees in wireless meshes. Novel approaches customized for wireless meshes have also been proposed [107] [139] [138]. Network coding [77] [133] has also been discussed as an alternative multicast solution.

Interference and Channel Selection

The optimal channel assignment in a multi-channel mesh topology is proven to be NP-Hard based on a problem mapping to the graph colouring problem [105]. Therefore, channel assignment has consistently been a focus [102] [69] [52] [89] [90]. Diverse fixed [2] [43] [85] [67], dynamic [19], and hybrid [67] [102] channel assignment solutions being proposed.

Adya et al. [2] proposed a greedy algorithm for channel assignment at each node. Raniwala et al. [105] presented a greedy load-aware channel assignment algorithm for 802.11-based wireless mesh networks which leads to orthogonal channel assignment. They later presented a complete set of experiment results for network settings in multi-channel 802.11 wireless mesh networks [104]. Kodialam et al. [70] investigated the necessary and sufficient conditions for orthogonal channel assignment and scheduling in such networks. They provided two algorithms for optimal assignment of orthogonal channels.

Given the tight coupling of different layers in wireless mesh networks, joint optimization across layers have attracted great interest [35] [78] [90]. Alicherry et al. [9] presented a joint orthogonal channel assignment and unicast throughput maximization framework. Rad et al. [90] investigated channel allocation, interface assignment and MAC design altogether. Merlin et al. [87] further provided a joint optimization framework for congestion control, channel allocation, interface binding and scheduling to enhance the throughput of multi-hop wireless meshes. Their framework
accommodates different channel assignments, but neighboring channel interference has yet to be addressed. A joint channel assignment and routing protocol [35] has also been proposed for 802.11-based multi-channel mobile ad hoc networks. While sharing many similarities with wireless meshes, the mobility concern and associated overheads are not critical in mesh networks given that the mesh routers and gateways are generally static.

In the work of Zeng et al. [139], two heuristics for multicast channel assignment were proposed, which also apply to multi-gateway configurations. They however did not explicitly address route optimization. In the work of Yuan et al. [138], routing and wireless medium contention were jointly considered. The impact of link interferences and power amplitude variations on each link were also closely examined, but were limited to single channel usage. Our work is motivated by these pioneer studies, yet our focus is mainly on throughput maximization in the multicast context.

2.2.2 Resources Management in Collaborative Wireless Local Access

The need to improve performance of 802.11-based wireless local area networks has been recognized in different applications and deployment scenarios [21] [88] [20]. We focus on research trends in access point association and further investigate solutions available in dense deployments.

Access Point Association

In existing wireless networks, a client is either associated to a dedicated access point, or to the best access point in terms of such metrics as the received signal strength index (RSSI), potential throughput estimates [122] [79], end-to-end airtime cost [59], or a combination of different metrics defined as association cost [92]. The decision is generally made by the client locally and selfishly, which has been analyzed through game theory in different network settings, e.g., wireless access networks [136] [48] [112], cognitive radio networks [60], and cellular networks with linear topologies (linear cellular networks) [71]. It has been shown that the selfish behavior does not necessarily converge to the optimal equilibrium [135].

Koutsopoulos et al. [71] suggested that a joint access point selection and channel assignment for linear cellular networks can achieve the optimum by minimizing the maximum clique (a set of cells that all interfere with each other) load. Their solution however cannot be extended to general cellular networks, in two dimensional area since the reuse constraints are not met and the clique loads cannot be balanced. Bejerano et al. [24] proposed an association control for global max-min service in wireless networks. Zhu et al. [143] proposed a user-centric management of wireless networks to
alleviate the collision problem. Liu et al. [79] also incorporated an access point association scheme for their practical spatial reuse antenna management solution.

There are also many decentralized association algorithms [72] [66] [135] [60] that try to achieve global optimum with local client measurements. Yet certain simplifications have been made in these pioneering studies, e.g., uniform throughput among all users or linear topology, which can hardly be extended to general network settings. For example, Xu et al. [135], Kumar and Kumar [72], and Kauffman et al. [66] assume all users on a single access point have the same throughput. Hong et al [60] limit the modelling to linear wireless networks.

Density and Interference

Limited number of orthogonal channels in 802.11 wireless networks results in overlapped channels among a number of access points, known as co-channel access points [53] [21]. These access points inevitably suffer from higher interference, higher collisions, and consequently sub-optimal throughput. Channel assignments mechanisms that we discussed earlier can be applied to partly alleviate this problem. There are also configuration-specific solutions that tackle the problem in enterprise settings.

DenseAP [91], self-managing architecture for thin access points (SMARTA) [3], and measurement driven guidelines for 802.11 design (MDG) [28] are a few of the available researches on supporting a dense deployment of access points to improve the performance of corporate wireless local area networks. The overall goal is to efficiently manage the data plane of an 802.11 deployment. All of the available solutions centrally optimize the client to access points associations, power, and channel assignment at the access points for corporate wireless local area networks. However, to our knowledge, there is no work addressing this problem in dense deployment of collaborative access networks.
Chapter 3

Resource Utilization in Cellular Multimedia Multicast

In this chapter, we discuss power-efficient high-throughput video multicast in LTE networks. Efficiently scheduling multicast traffic [128] and power efficiency in user device [113] and eNodeB [65] have been separately addressed in the literature. Transmission power in eNodeB is important for transmission quality throughout the cell and interference control for cell-edge users. The higher the transmission power, the better the reception of users with different channel conditions. However, high transmission power from different cells will increase interference for cell-edge users. Power consumption in user side is important for user device battery lifetime.

Our goal is to minimize transmission power in eNodeB (base station in an LTE cell) and user device power usage while enhancing the user-perceived quality of video multicast [14]. Multicast transmission in LTE systems requires using forward access channel (FACH) and allocation of power for full coverage of cell during the transmission of multicast subframes. Therefore, multicast transmission power might be as high as 43\% of an eNodeB’s total transmission power budget at a certain time [65].

To achieve power-efficient high-throughput multicast we should decide: how to group users with different multimedia needs in a few multicast groups, which layer of multicast video should be transmitted to the selected group, how many members of multicast group are involved in this transmission, which users can benefit from sleep mode power conservation, and what the transmission rates on each transmission should be. The higher the transmission rates, the lower the sleep mode power conservation chances, and the higher the eNodeB transmission power. Therefore, we try to
reduce redundant data transmission to enhance interference and power usage in end user devices. This reduction in volume of transmitted data should not hurt user-perceived video quality.

In this chapter, we provide our power-efficient scalable video multicast solution in three main steps. In the first step, we categorize the users in a cell into different groups. Then we decide on the base and enhancement layers that should be sent to each of the groups. Third, we calculate the burst transmission periods for all data packets including all layers of multimedia. We schedule transmissions to achieve maximum throughput as well as maximum user inactivity period.

3.1 System Settings and Problem Formulation

We assume an LTE cell \( A \) with \( n \) users within the cell. Each user \( i \) is within \( d_i \) meters from the eNodeB of cell \( A \). Users are moving in the cells, therefore, \( d_i \) is variable over time. Size of a cell, the farthest from eNodeB it can support a node, is indicated by \( D \), which is a constant value determined by eNodeB deployment and geographical shape of the area.

We are interested in scalable video multicast from eNodeBs to the user devices in the cell \( A \). We consider cell-border interference and cooperation contribution of cell \( A \) and possibility of users crossing the borders of cells and entering their neighbouring cells. Considering the problem for each cell independently, we do not consider single frequency multicast and synchronization complexities that arise with it.

We assume there are \( N_v \) different video resources and maximum \( L_v \) layers of multimedia used in our scalable video coding for each of the \( v \in V \) video resources. The number of users subscribed to video source \( v \) is indicated by \( S_v \). Layer \( j = 1 \) for each video source \( v \) indicates the base layer and \( j = 2, \ldots, L_v \) show the enhancement layers. Table 3.1 summarizes the parameters used in our system and problem statement.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v )</td>
<td>Video resource</td>
<td>( V )</td>
<td>Set of all video resources</td>
</tr>
<tr>
<td>( T )</td>
<td>Time frame</td>
<td>( K )</td>
<td>Number of resource blocks in a time frame</td>
</tr>
<tr>
<td>( j )</td>
<td>Video layer</td>
<td>( M_{T,r} )</td>
<td>Number of transmission antennas</td>
</tr>
<tr>
<td>( L_v )</td>
<td>Layers for resource ( v )</td>
<td>( N_v )</td>
<td>Number of video resources</td>
</tr>
<tr>
<td>( D )</td>
<td>Diameter of cell ( A )</td>
<td>( d_m )</td>
<td>Average distance for group ( m )</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Angle of user, from east</td>
<td>( \theta_m )</td>
<td>Average angle for group ( m )</td>
</tr>
<tr>
<td>( \theta_{max} )</td>
<td>Max angle</td>
<td>( C_{v,j}^{i} )</td>
<td>Importance of layer ( j ) of video ( v ) for node ( i )</td>
</tr>
<tr>
<td>( x_{i,k}^{v,j} )</td>
<td>Allocation of video ( v, j ), node ( i ) on slot ( k )</td>
<td>( \phi_j^v )</td>
<td>Encoding rate for video ( v ) layer ( j )</td>
</tr>
<tr>
<td>( p_i^v )</td>
<td>Preference for node ( i ) on slot ( k )</td>
<td>( r_{i}^{v,j} )</td>
<td>Rate for user ( i ) receiving video ( v ) layer ( j )</td>
</tr>
</tbody>
</table>
User devices are heterogeneous, therefore different in display resolution. These differences are reflected in their requested resolution of video sources. Our goal is to achieve power-efficient scalable video multicast for users in cell $A$. Meanwhile, we aim to conserve user-perceived quality in the heterogeneous devices.

To reflect the diverse resolutions required to fulfill user-perceived quality, we assign criticality values to layers of video for each user. We define criticality variable, $C_{i}^{v,j}$, as indicator of importance of each layer $j$ of video $v$ for a certain user $i$ in a cell. We use weighted average of received videos’ criticality values as a measure for user-perceived quality. We set $C_{i}^{v,j}$ values to meet the following conditions:

$$\sum_{j=1}^{L} C_{i}^{v,j} = 1 \quad \forall v \in V$$

$$C_{i}^{v,1} \geq 0.5 \quad \forall v \in V, \forall i = 1,..,N_n$$

The first condition states that the critically values should sum up to one. This ensures a common scale for the criticality values. If a user does not require a certain layer, $C_{i}^{v,j} = 0$ for that user and its criticality weight is divided among rest of the layers based on size of the user equipment screen and requested resolution of scalable video stream. The second condition implies that a user subscribed to a source at least needs the base layer. We compute criticality values based on application layer multicast video resolution requests.

We define $E(d_i^n, \phi_j^v, w_j^v)$ to indicate the transmission power in eNodeB used for transmission of layer $j$ of video $v$ to user with distance $d_i^n$ with modulation $w_j^v$. The higher the data rate, the higher the transmission power, and the lower the sleep mode chances at the user side. Transmission data rate is determined based on video encoding rate and modulation and coding scheme. In 3GPP LTE standard, a Resource Block(resource block) is the smallest allocation unit in LTE OFDM radio resource scheduling. An resource block can be independently modulated and assigned to a user. Each resource block contains 6 or 7 symbols, with 12 adjacent sub-carriers of $15KHz$ (total of $180KHz$) in the frequency domain. Two resource blocks form one time slots ($1ms$) in the time domain. Modulation and coding scheme decides the number of bits to be transmitted in each resource block. Therefore, it affects the transmission power as well.

We assume that the data capacity of a resource block is shown by $w$. $w$ is decided by modulation and coding scheme. For example, in QPSK $w$ is 2, for 16QAM modulation it is 4, and for 64QAM modulation $w$ is 6. Higher the number of bits to be transmitted in an resource block, higher the transmission power should be. We use variable $x_{i,k}^{v,j} = 1$ if the a resource block $k$ is assigned to
deliver the layer \( j \) data of video source \( v \) to user \( i \). We name \( x_{i,j}^{v,k} \) the resource allocation variable. To limit the boundaries for variable \( k \) and to formulate the OFDMA time or frequency resource block assignment as well as abstracting the technical details, we assume \( K \) is the number of resource blocks we can schedule on eNodeB on a specific period of time, \( T \) time steps. \( k = (f, t) \), where \( t \) can take values of \( t = 1, \ldots, T \), \( T=\)Length of Frame / Length of Subframe (e.g can be \( 10ms/0.5m = 20 \) in LTE) and \( f = 1, \ldots, F \), \( F = \frac{BW}{180KHz} \) and \( K = F \times T \).

While we achieve power efficiency in eNodeB by decreasing the transmission power and range, we target it in user device by scheduling the traffic in bursts so that user device can turn off its circuit more often to conserve battery. user device power usage minimization translates to user device discontinuous reception sleep time maximization. We achieved this by minimizing redundant video transmissions and scheduling video transmissions in bursts. Maximum rate each user can receive is shown by \( r_{i}^{v,j} \). If the transmission rate is higher than video encoding rate, we break down traffic into bursts of different sizes with transmission rate \( r_{i} \). \( p_{i}^{k} \) shows preferred resource block of user \( i \) for receiving requested multimedia. \( p_{i}^{k} = 1 \) if user \( i \) is awake at the time of resource block \( k \). We call \( p_{i}^{k} \) the scheduling preference. We assume \( T \) and \( K \) are constant design parameters. We also assume each eNodeB has \( M_{Tr} \) transmission antennas.

In the next section, we propose a simple practical solution for power-efficient scalable video multicast in LTE wireless networks.

### 3.2 Power-Efficient Multimedia Multicast

Our power-efficient scalable video multicast algorithm is provided in three steps: grouping, channel assignment, and scheduling. The proposed solution is trying to solve the problem by constantly limiting the power budget in grouping and channel assignment steps. Then, it provides a feasible scheduling for high throughput and low power consumption for user devices within the given power budget. We describe each part in detail in the following sections. All the variables used for our modelling and solution are summarized in table 3.1 for easier reference. Figure 3.1 shows the block diagram of different solution components and flow of data among them.

#### 3.2.1 Grouping

We first categorize the users in a cell in different groups. The grouping is based on four different factors: multicast sources a user is subscribed to, requested quality of video, user placing, and availability of LTE femto cells in a user’s area. Grouping based on multicast sources specifies the
groups of users that have requested a special multicast source. This is important because it may eliminate the need for multicast if the source request depends on user placement. Requested quality of video is considered as a factor in grouping, because it groups together the users with shared layer requests and shows us the opportunities to unicast the remaining layers which may not be requested by many users. User placing is important for beamforming as well as transmission power setting. Availability of femto-cells might help us in delegating the femto cell base stations to send the data to users in their area which may eliminate the need to multicast to that area.

Each group subscribed to multicast source $v$, requesting video quality enhancement layers up to layer $j$ is indicated by $g_m = g_{d,\theta}^{v,j}$. $m$ is the group ID, so each $g_m$ is unique. There are $N_G$ groups in each cell $n$. $N_m$ shows the number of users in the group $m$. Parameter $d_m$ shows average distance of members of group $m$ from base station. Starting from East as 0 in counter clockwise direction,
$\theta$ shows the angle of the middle geographic place of the group makes with the zero. While $d_{m}$ facilitates power adjustment for a multicast group, $\theta$ facilitates the beamforming decisions based on number of antennas available in the base station. Multicast beamforming increases spacial diversity and decrease power consumption in LTE networks. We do not discuss beamforming details in this chapter, but multicast beamforming is considered in grouping algorithm to facilitate incorporation of such solutions with the provided algorithm.

There are specific properties to the specified groups. These properties are:

$$N_n \leq \sum_{m=1}^{G_n} N_m \leq L_v \times N_n$$  \hspace{1cm} (3.1)

$$g_{d,\theta}^{v,j} \cap g_{d,\theta}^{v,j'} = g_{d,\theta}^{v,j} \forall j' \geq j$$  \hspace{1cm} (3.2)

The first property shows that users can be member of different layer groups. The second one shows that with the same $\theta$ and distance, the group with higher resolution request can contain the members of lower resolution groups. This may be used to merge and eliminate some groups later.

As the initialization of grouping algorithm, we group the nodes in $v$ different groups based on the different videocast sources they are subscribed to. Note that a node could be subscribed to more than one multicast source at the same time, therefore, a node may be in different groups with different $v$ values. The example of when a user might be subscribed to more than one videocast is watching one as well as tracking the other videocast source waiting for a specific moment of the show, or streaming one concurrent with buffering the other source for later use.

We apply position constraints of $d$ and $\theta$ for grouping on each video subscription group. In this step we bring into account the cooperative possibility and availability of femto cells. Finally, we distribute the nodes within each cluster among the groups of the same distance but with different requested layers of video. The number of users in a cell could be as high as 600 users. Grouping based on position of the nodes is coarse and we usually won’t have more than a certain number of popular videocast sources. Therefore, the probability that it will result in trivial one user grouping is low, regardless of rough clustering constraints.

On later steps of the algorithm we may merge different groups. The merge depends on granularity of enhancement layers, placing of the nodes and number of MIMO antennas. It usually happens in channel assignment and power allocation step. The merge may occur if the power consumption for unicast transmission of an enhancement layer to a groups members costs more than merging two or more groups and providing extra layers of multimedia for a portion of nodes in PTM transmission.

We will discuss the process further in the next section. Minor changes can be applied to the
grouping algorithm to consider the cooperative routing possibility. One change is adding a margin to $D_n$. This could be done by adding an extra 10% to $D_n$. Second, for nodes between $D_n - 10\%$ and $D_n + 10\%$ a flag for cooperative routing should be set.

### 3.2.2 Channel Assignment and Power Allocation

After we group the users, we assign bearers and channels for transmission of multicast data to each group of users. This assignment is based on the power efficiency calculations done separately for each group of users. We calculate the power needed to send a certain layer of video in point to multipoint (PTM) format over multicast channel (MCH), called $P_{PTM}(g_{d,j}^{v,j})$ and the power needed to send the same data using point to point (PTP) wireless transfer shared data channel among eNodeB and each of these nodes.

$$P_{P2P}(g_m = g_{d,\theta}^{v,j}) = N_m \times P_{SCH}(d, \theta, \phi_j^v)$$

$$P_{P2M}(g_m = g_{d,\theta}^{v,j}) = P_{MCH}(d, \theta, \phi_j^v)$$

The power consumption values are calculated separately for each layer $j \in L_v$ for each group. If eNodeB decides that multicast channel is the efficient medium to transmit data to an specific group, all the groups with the lower enhancement layers of the same videocast closer to the eNodeB with the same angle will be updated to use the multicast channel. In this case, channel assignment algorithm will merge those groups over $j \in L_v$ parameter if the power needed to send the extra layers of video is negligible for users not requesting it. Otherwise, the decisions are made for each group separately. This is an example of merge decision we discussed earlier.

If the PTM power to send a certain layer of multicast to a group of users is lower than PTP transmission to each of its members, all the groups in the same region, and closer to the base station will also receive the transmission in PTM. These means the two groups are merged.

### 3.2.3 Scheduling, Resource Allocation, and Energy Management

In previous sections we discussed how we group the users, decide on transmission and find the upper bound on transmission power. The next step in our algorithm is scheduling the video data and allocating resource blocks for burst transmissions for each layer of video over assigned channels.

Discontinuous Reception (DRX) is a mechanism in LTE networks that allows wireless user equipment to turn off its receiving circuit and go to a sleep mode when there is no reception scheduled. With hierarchical design of circuits for sleep mode, it is even possible to turn off the receiver
for time of a subframe. Sleep mode is not yet in practice for such small durations. To use DRX for minimizing user device power consumption, we require maximum user inactivity period so that users can switch off the receiving medium to save equipment battery usage. If the rate of data transmission is lower than the transmission rate the wireless medium allows, we can break down the traffic into bursts and use the gap among bursts for DRX. We minimize power consumption in user devices by maximizing data throughput in each transmission burst and the gap among the bursts. To achieve burst transmission, we require high data rate during each burst to achieve high utilization of wireless medium and longest possible gaps.

We know that higher data rates require higher transmission power. Therefore, there is a tradeoff among user device power consumption and eNodeB transmission power minimization. We address this tradeoff by staying within the transmission power budget constraint at eNodeB. So, we minimize user device power consumption within the transmission power maximum we calculated in the channel assignment step. In addition, we can delay transmission within specific deadlines to achieve higher gaps within data bursts. We decide on the transmission deadlines based on their timeliness and usability upon reception to produce the scalable video. We also need to be fair among the users in our scheduling. Another important factor is the readiness of the receiving nodes to turn on their radio to receive the scheduled transmission. Otherwise the transmitted packets might be dropped or lost.

Scheduling decides the video data to be transmitted next. We have to decide (1) which group of the users should be served next, (2) which video resource should be sent to the selected group, and (3) which layer of requested video should be sent first. We decide on the group to be served based on two factors: average Channel Quality Indicator (CQI) values for the users of the group, and fairness among the groups. This is to preserve fairness and achieve high throughput of the wireless medium in the mean time. Note that in our mechanism, there is no feedback from the users in the multicast mode of LTE. Therefore, the CQI values should be attained and registered from unicast transmissions of the same users. This is facilitated by coupled transmission of unicast and multicast. Note that since we have already selected the nodes with similar placing and therefore similar channel quality values in the grouping step of our algorithm, unlike similar rate decision algorithms, calculating the rate based on lowest CQI in the group will not significantly hurt the performance of the system.

Next, the videocast source for the group is selected for the group based on weighted round robin (WRR) on average criticality value for a source. This is to make sure all sources are being served, but the source with higher amounts of data to send gets the priority. The layer of video is also selected
based on the summation of criticality values. Priority is higher for base and lower enhancement layers in each RR period to enhance video quality. Note that significantly higher values of C for base layer will result in selection of base layer if there is any to be scheduled for the group. The scheduling and resource allocation algorithm consists of a two packet scheduler algorithms: unicast and multicast. Algorithm 1 shows our multicast scheduling algorithm. We use proportional fair scheduler for our unicast schedulers.

**Algorithm 1: Multicast Scheduling Algorithm**

\[
\text{for each scheduling duration do} \\
g = \arg \max_p \left( p_k^g \times (w_a \times \text{AvgCQI}_g + w_b \times D_g) \right); \\
\tilde{v}_s = \text{Max} \left( \text{Avg}_{j} (C_{v,j}^i) \right); \\
v_s = \text{Decide on WRR on} \tilde{v}_s; \\
j_s = \arg \max_j \frac{1}{N_g} \sum_{i=1}^{N_g} C_{v,j}^i; \\
\text{Queue Packets of } j_s; \\
\text{end for}
\]

After scheduling the packets, we perform resource block and subframe allocation. In LTE, multicast data is transmitted in specific dedicated subframes where the physical data shared channel that carries the unicast traffic is not transmitted. We use both point to point (PTP) and point-to-multipoint (PTM) bearers. Therefore, we should schedule packets for two different purposes, multicast and unicast. There are also some multicast packets that might be decided to be sent unicast. Those are also scheduled with unicast downlink scheduler.

The data reception, sleep and awake times of each user equipment is decided in scheduling and resource block allocation step in eNodeB, and transferred to the user. This is done within DRX management module with signaling the users to turn their receiver radio off when there is no transmission. We use current LTE DRX mechanism. We do not discuss DRX signalling and communication. We assume that it is carried out within DRX management module.

Transmission rate for each group is then calculated based on modulation and coding scheme. Higher coding rates are used for enhancement layers. In theory an LTE base station can separately modulate each sub-frame based on rate and channel quality of the receiver. However, changing modulation on each sub-frame is proved infeasible in practice. We start modulation and coding values with 4QAM for base layer and 16QAM for enhancement layers. We use historical modulation information, channel quality and source distortion for deciding on rate and modulation and coding scheme at each resource allocation period.

The last step of the algorithm assigns the transmission power to each multicast group. This
transmission power is within the constraint given but finalized after deciding all the parameters. After scheduling the packets for each group, assignment of transmission power to transmit the packet based on the $E_m$ value happens in physical layer. The summary of the steps are as follows:

- Compute $r^{v,j}$ based on $\phi_j^v$ and selected modulation and coding
- Update $p_k^g$ value based on new $r^{v,j}$
- Compute number of multicast channel subframes in a frame based $r^{v,j}$
- $E_{Transmit} = E_m(d_m, \phi_j^v, w_j^v)$

### 3.3 Performance Analysis

We compared our algorithm to a variety of state-of-the-art solutions. In this chapter, we present a set of results as compared to unicast-only transmission, basic MBMS, and counting algorithm. In MBMS, multicast packets are send via PTM bearers and the multicast channel. The counting algorithm decides on using the PTM or PTP bearers, i.e., multicast or shared data channels, for multicast based on the number of receivers of a multicast source. In figures, PTM shows the basic MBMS implementation and PTP shows the unicast-only transmission. This choice is based on the fact that the counting method is currently used in the MBMS standards for deciding to use PTP or PTM bearers. PTM and PTP transmission also show the two extreme sides of low power consumption in PTP versus low redundant data transmission and low delay in PTM. All the packet schedulers for the comparing methods are considered proportional. Scheduling algorithm for our solution is as proposed in previous section. We used coarse grained high definition scalable video as our video sources. We present single video source scenario for clarity. Our video sources is Sony
Demo coarse grained scalable video trace file in HD, 30 frames/sec from Video Trace Library [110]. Four layers of video are sent in separate streams to users. The criticality values are mapped to users’ request for these layers of video quality.

We implemented all of the algorithms for single-cell scenario. Therefore, there is no single frequency multicast and synchronization among different cell transmissions. This is because the solutions for single frequency multicast are completely different and out of scope of this chapter. However, we assume that interference from neighbouring cells affect the simulation. Thus, we simulated 7 cells in the evaluations. We implemented the counting method for different values of counting thresholds and we used the best value of such threshold in experiments, found using extensive simulations. During the first set of simulations, illustrated in figures 3.2 -3.5, 50 users are registered to the only multicast video source. This assumption seems to be fairly realistic assuming the number of users in an LTE cell, around 600. The video quality request of the users is randomly decided at the beginning of the simulation based on four layers of video available and the criticality values are assigned. We used same set of users with the same quality requests in all algorithm evaluations.

Figures 3.2 and 3.3 show DRX subframe potentials our algorithm creates for users in the network. Our algorithm outperforms all but PTM transmission in sleep potential. This is due to the fact that PTM transmission can only use MCH specific subframes and if there is no unicast traffic in the network, the rest of the subframes can be used for DRX. But our algorithm may use multicast channel and unicast subframes to decrease transmission power and preserve the quality of video. The enhancement is substantial compared to counting method. PTP transmission leaves very limited sleep potential as it has to serve all users by a different unicast transmission.

Figures 3.4 - 3.9, on the other hand, depict the costs for such gains in power consumption. Delay
in our algorithm, shown in figures 3.6 and 3.7, is considerably decreased compared to counting method and it is just a bit higher than PTM transmission in low bandwidth scenario.

Figures 3.6 and 3.7 contain the same data with 3.4 and 3.5 but the PTP curve is omitted so that we can compare the other curves. Delay in PTP transmission is significantly higher because it serves all users with separate unicast transmissions. Therefore, the transmission rate cannot afford 50 concurrent unicast HD video transmissions and it has to delay transmission of some packets. Some of the delayed packets are dropped before transmission. This degrades the quality of video at some of receivers. Grouping algorithm shows amazing results in higher bandwidth. It shows slightly higher delay compared to PTM in lower bandwidth due to the nodes it serves in unicast. The delay is negligible and does not affect video quality.

Figures 3.8 - 3.9 show that our algorithm adds minor data transmission overhead compared to the multicast-only transmission. It is considerably lower than the transmitted data volume in counting algorithm and PTP transmissions. It sometimes is even lower than PTM transmission. This is due to the fact that it can eliminate the need to transmit certain layers of video when groups of users don’t
need it. The reason this only happens in higher bandwidth is that in low bandwidth the share of multicast transmission decreases. Therefore, the rest of video data which cannot be fit into multicast subframes are dropped. These drops include low priority layers of video for all users.

Throughout the simulations, we use sum of criticality values of received and usable video packets in user devices to measure user-perceived quality. It degrades significantly in PTP transmission. But the values are over 0.98 for grouping and PTM methods over all simulations and for counting in higher bandwidth simulations (for 50 users, over 10 MHz). Our goal was to preserve the video quality as it is in PTM which is achieved in all, even enhanced in higher bandwidth scenarios.

3.4 Summary and Discussion

In this chapter we first discussed the need for efficient resource usage in LTE-based cellular wireless networks. Then, we focused on resource efficiency in the multimedia broadcast and multicast services in these networks. Evolved multimedia broadcast and multicast services are envisioned to provide multimedia streaming and download services in LTE networks. While high multicast throughput and low latency are the main objectives of these services, power efficiency in user and service planes are also essential design challenges. User device power efficiency is critical for longer lasting battery. Power efficiency in base stations is important for cost-effectiveness of services, performance, interference reduction, and environmental effects. Although current solutions for multimedia multicast over LTE networks offer major improvements, imposed by complexity of LTE systems, power and spectral efficiency of these services are still subject for ongoing research.

We examined power-efficient high-quality scalable video streaming in LTE networks through its eMBMS service. We considered scalable video streaming and download services offered by eMBMS service over LTE networks, and proposed a simple but effective and practical solution to enhance user experience and power consumption in both user device and eNodeB.

To perform power-efficient multimedia transmission in LTE networks, we face three key trade-offs: (1) maximizing energy saving vs. minimizing delay, (2) maximizing sleep time vs. minimizing lost packets, (3) maximizing quality of video vs. minimizing unnecessary video transmissions. We provide a balanced solution that addresses the trade-off by including user preference. Our simulation results indicated 5% to 18% improvement in base station power consumption and 13% to 25% improvement in user device power conservation chances. The provided solution also decreases the transmitted data in the network while preserving the user perceived quality of the video.
Chapter 4

Cellular Seamless Connectivity for High Speed Mobility

The recent advent of high speed trains introduces new mobility patterns in wireless environments. High speed trains are evolving and becoming popular as a new means of transportation. The Shinkansen in Japan (300 km/h), the Chinese high speed railway system (165km/h; expected to reach 350 km/h in the near future), and the Eurostar (175km/h to 334km/h) are extremely successful examples in use, with a number of others being under construction or preparation. Providing multimedia access for the passengers in these high speed trains thus becomes a critical demand when deploying wireless networks along the high speed railways [140].

The LTE-A (Long Term Evolution of 3GPP - Advanced) networks have largely tackled the Doppler effect problem in the physical layer and are able to keep wireless service with 100Mpbs throughput within a cell in speeds up to 350 km/h. Yet the much more frequent handovers across cells greatly increases the possibility of service interruptions, and the problem is prominent for multimedia communications that demand both high-throughput and continuous connections.

Since a railway system usually covers a vast geographical area, it is economically expensive to build a special network only for the passengers, not to mention that most of the users do not always stay on the trains. As such, general-purpose networking should be adopted, particularly LTE, the de facto standard for the new generation of wireless cellular networks. The LTE physical layer can support high throughput data delivery at speeds up to 350km/h and even 500km/h in rural areas. State-of-the-art implementations of LTE have the capability to support at least 200 active data users per 5MHz of bandwidth in a cell, which will further increase in LTE-Advanced. Yet a
high speed train remains a challenge environment for LTE networks. First, the wireless channel condition changes drastically [141], thus disturbing the data rate; Second, handover across cells becomes much more frequent, thus interrupting connections. Both of them are highly undesirable for multimedia communications given the stringent demand on data bandwidth and continuity.

To address these challenges, we present a novel LTE-based solution to support high throughput and continuous multimedia services for high speed train passengers [17]. Our solution is based on a Cell Array that smartly organizes the cells along a railway, together with a femto cell service that aggregates traffic demands within individual train cabins. Our Cell Array design is initially inspired by the moving extended cell [101] concept, though ours is based on LTE femto-cell [51]. It also works in the wide frequency band of LTE networks and it combines hard and soft handovers, as well as known information about high speed trains to improve the handover experience for high speed train passengers. Given that the movement direction and speed of a high-speed train are generally known, our Cell Array effectively predicts the upcoming LTE cells in service, and enables a seamless handover that will not interrupt multimedia streams. In this chapter, we detail the signalling and actions in such a predictive handover. To accommodate the extreme channel variations, we further propose a scheduling and resource allocation mechanism to maximize the service rate for femto cell base stations, according to periodical signals quality changes. This ensures the multimedia service quality within each femto cell.

4.1 System Overview

We investigate the problem of seamless high-throughput wireless access for High Speed Train (HST) passengers in an area covered by LTE cells and provide a seamless handover solution. Each LTE cell [130] has an outer radius of $r$, and a frequency reuse distance of $D$ meters. In practical implementations of LTE networks, $r$ is typically 250-500 $m$ for urban cells, and 1-10 $km$ for rural cells. A base station is the user access point of a cell, referred to as eNodeB or eNB in short. A user, connected to an eNB by its User Equipment (UE), can move freely across the cells, and a handover takes place when a UE changes the serving base station (eNB). The procedure of a hard handover includes the registration of the user to the new cell, deletion of its information from the previous eNB, and changing its routing path to the new eNB. An HST $H$ is a cascade of $M$ cabins. Each cabin is $l$ meters long and can accommodate up to $n$ passengers. For ease of exposition, we assume that all HST passengers are LTE users and are using Mobile User Equipments (MUEs). The train $H$ is moving at speed $0 \leq s \leq S_{Max}$, where $S_{Max}$ is up to 500 $kmph$, which has yet to be achieved...
in today’s systems. The users in an LTE cell, particularly the MUEs in the HST, are moving in the multi-cell arrangement. Therefore, $d_i$, the distance of each MUE $i$ to its currently serving eNB is changing over time. Figure 4.1 shows the settings and the key system parameters for the HST wireless access.

4.1.1 Seamless Handover with Cell Array

Our solution is completely based on LTE-Advanced networks and provides a novel vehicle-to-infrastructure communication solution to enhance the HST user connectivity. We introduce the concept of a moving LTE femto cell. Then we design a cell array architecture that ensures seamless handovers when the femto cell moves across the LTE cells.

Moving Femtocell

A femtocell is a small cellular base station, designed for limited coverage of 10 to 40 meters. The femtocells are traditionally used for extending wireless service coverage to the indoor spaces or for office or home usage. The typical length $l = 25m$ of an HST cabin fits well within the coverage range of an LTE femto cell. We propose using two LTE femto cells to provide vehicle-to-user communication within an HST cabin as illustrated in Figure 4.2. A femtocell base station is referred to as Moving eNB (MNB) throughout this chapter. Using femtocells, the HST users can access LTE wireless services through their local MNB.

An LTE femto cell uses the same frequency bands as an LTE macro cell. The number of allocated frequency subcarriers [142] can be decided based on the number of the wireless users within the femtocell. In our design, each femto cell covers half of an HST cabin and up to 50 users. A frequency band of 5MHz supports wireless access for the users in each femtocell. To accommodate the mobility of femtocells along the railway, the frequency band of the MNBs and their frequency
reuse distance $D$ have to be considered in frequency band selection.

On the other hand, an MNB should also receive the aggregate traffic for all MUEs through the LTE air interface. Therefore, each MNB includes two interfaces that work within two different frequency bands. One for cabin-to-infrastructure communications, i.e. receive the aggregate download traffic from the infrastructure LTE network, and send the requests and feedbacks. In this part of the communication, the MNB is an MUE for the infrastructure LTE cells. The other interface is used to communicate with the MUEs within the cabin, where the MNB functions as an LTE femtocell base station.

Figure 4.2: A Number of LTE Femtocells Can be Used to Cover Each Train Cabin

**Cell Array**

We define a Cell Array as an extended cell architecture composed of three cells in a row along the path of the high speed train railway. We call them cells A, B, and C. Cell A is the cell that train is partially or completely in at the current time. Cell B is the adjacent cell that is ahead of cell A along the railway path. Train can be partially in cell A and partially in cell B in any given time. As soon as the whole train is in cell B, cell A is no longer in the extended cell configuration. The cell array reconfigures when the train completely leaves cells A and enters cell B. Figure 4.3 shows the cell array re-configuration process as the train moves along the path. When cell D joins the extended cell, the original cell A is deleted and we rename the cells to A, B and C again.

We use the cell array structure to include the known railway path and the speed of train in the scheduling and handover mechanisms. It facilitates fast handover and helps us minimize the amount of data to be transferred among the eNBs in case of a handover. We always have the information of the MNBs registered in all of the three cells in the cell array. Only cell A and sometimes cell B are transmitting to the MNBs. Therefore, there is no need to transfer the whole downlink data to all of the eNBs in the cell array. Only the eNBs of the transmitting cell will receive the downlink data.

Inclusion of cell C in the cell array is to facilitate frequency spectrum assignment for MNBs of
Figure 4.3: Cell Array (CA) Reconfiguration Along the Railway Path

the HST without service degradation in an LTE cell. We call it a *soft handover*. It also provides scheduling and predictive buffering chances for non real-time traffic. It is worth noting that the reason we extend the cell array in three cells along the path is to smoothly clear the spectrum for vehicle to user communication without disturbing the currently active users in the cell, as well as to buffer non-real time data in case of delayed scheduling.

### 4.1.2 Predictive Handover Mechanisms

We now detail the two different types of predictive handovers for successfully reconfiguring the cell array as well as performing the hard handover for the MNB along the railway path. Our predictive handover mechanisms benefit from the cell array architecture to shorten the handover time and keep the service uninterrupted in high mobility conditions.

**Predictive Hard Handover (PHHO)**

A hard handover is the procedure of registering a UE, which is the MNB of the moving femtocell in our solution, in a cell and switching its data forwarding path to the new eNB. After a hard handover, the MNB will no longer be registered in the previous cell. Our hard handover mechanism is *predictive* because we predict two cells ahead by adding them to the cell array. Therefore, during the
CHAPTER 4. CELLULAR SEAMLESS CONNECTIVITY FOR HIGH SPEED MOBILITY

hard handover, there is no need for the current eNB to negotiate the registration of the MNB to the neighbouring eNBs, neither the MNB has to find the target cell by signal quality. This is because they both have the information of the target cell, which is reported by the MNB to the eNB, or the MNB initiates the handover as discussed in detail below.

Three different network elements can initiate a PHHO: the MNB crossing the cell boundaries, the front neighbouring MNB, or the current eNB (cell A). All these three elements are able to initiate a handover request so if any of them fails, others will be able to do the handover and initiate the cell array re-configuration. The first PHHO request is initiated by the MNB crossing the cell boundaries. When an MNB crosses the borders of cell A and enters the cell B, it sends a handover request to the eNB of cell A. The eNB of cell A informs eNB of cell B of the handover.

The second PHHO request mechanism is provided by the front neighbouring MNBs. Since the speed of the train is known for each MNB, when sending a handover request for itself, an MNB can send the handover request for the MNB following it. A high speed train moves in speeds of 50-100m/s. Therefore, the next MNB in the train will be reaching the handover point in fraction of a second. It will still be connected in cell A, but as soon as it can receive signal form target cell, handover is confirmed and its connection is established with cell B. Now, this MNB can request handover for its next in row MNB on the train. During the time it is registered in cell A and the handover request is initiated, new non real-time download traffic to this MNB is forwarded to cell B. The third PHHO request mechanism is initiated by the eNB of cell A. In case any of the previous two handover initiation mechanisms are not started before signal degradation for the MNB, the eNB of cell A can start it. The eNB of cell A knows the next cell in the cell array, therefore it can start the handover process without negotiating with neighbours. The MNB will remain connected to cell A, but as soon as it can receive signal form cell B, the handover is confirmed and its new connection will be established. All these handover mechanisms and their associated signalling are illustrated in Figure 4.4. Note that the handover responses are sent both to the current cell eNB and to the MNB itself. This is because the MNB may move during the handover and might already have entered the new cell. As such, sending the handover response along with the random access code directly to the MNB will speed up the handover process. When the eNB of the cell A receives the handover response, it tries to inform the eNB of the handover response. It sets a timer and sends the handover response to the MNB. After the timer expires, the eNB deletes the information of the MNB. Therefore, during the handover the MNB is still registered in cell A. The handover response sent through the eNB A is useful for the MNB if the reception of the MNB is still better through the eNB A. Otherwise, it can receive the handover response directly. This is possible, because
other needed information are exchanged among the eNBs A and B before the hard handover, when arranging the cell array.

In all the three handover initiation mechanisms, both cells B and C already have the registration information of the train MNBs. Buffering of traffic to cell B starts as soon as the handover request is initiated. As soon as all the train MNBs are transferred to cell B, cell A is deleted from the cell array configuration. The next cell along the railway path will be selected to be added to the cell array. This selection will be made in the Predictive Soft Handover Mechanism (PSHO).

**Predictive Soft Handover (PSHO)**

While in PHHO the cell array has the registration information of the train MNBs and the active cell change takes place, PSHO provides a predictive mechanism to register user information in a cell array. The MNB femto cells use the same frequency bands as the infrastructure LTE cells. If not chosen properly, the femto cell frequency bands will interfere the infrastructure LTE cells. To prevent this, and also to facilitate using the known path and speed information in the cell array
architecture, we now propose a soft handover mechanism.

We register cell C in the cell array to announce the femto cell frequency band. This is to prepare cell C for probable allocation of a large chunk of its frequency spectrum to an MNB femto cell traffic. Since the railway is on a known path, it could be considered in initial frequency band selection of LTE cells so that the frequency used by HST femto cells will not be used by the Infrastructure LTE cells. In the rare case of sharing the same spectrum, cell C should be aware not to use that frequency spectrum in the next scheduling and RB allocation cycle and send the scheduled UEs to their next preferred band with the highest Channel Quality Index (CQI) value. The eNB for cell C will be informed by the previous cell C’s (before reconfiguration) eNB. Cell C’s eNB approximates the time it can still use the spectrum according to the speed of the train and diameter of its region:

\[ T_{AHO} = \frac{2 \times D_{cellC}}{Speed} \]  

After frequency band selection, and registration of HST MNBs in the cells B and C, cell A receives the cell B and C’s information for the hard handover cell selection and handover response process. We call this procedure a Predictive Soft Handover (PSHO). The PSHO signalling will be done across the MNBs and the newly added cell to the CA, when the last MNB enters cell B. The signalling for this procedure is shown in Figure 4.4. Our PSHO only performs user registration and frequency spectrum allocation for the MNB femto cells. The MNBs will be registered in cell C after the PSHO. However, there will be no path switching or data forwarding to cell C.

Implementation of railway systems and an infrastructure network to provide connectivity to high speed train passengers are costly projects. Therefore, it is possible to consider an infrastructure cell alignment implementation next to the railway. However, if the railway crosses edges of multiple cells, the eNB C starts to negotiate selection of the eNB D during the PSHO process. The selection is reconsidered each time reconfiguring the cell array and will be available for the MNB before a PHHO. Therefore, cell A never has to negotiate cell selection, or wait for MNB cell selection during the PHHO. This allows more aggressive settings handover failure and re-entry timer which leads to faster handovers.

4.1.3 Scheduling

Although the moving femto cells and the cell array architecture along with the predictive handover mechanisms enable fast handover for seamless wireless connectivity, HST multimedia service users need continuous high throughput connectivity for seamless multimedia services. This is provided with a scheduling mechanism tuned with high speed movement and handover information. The
wireless coverage along the railway path is not uniformly distributed: as illustrated in Figure 4.5 [118] the signal is stronger near the eNBs and weaker as the HST gets further away towards cell edges. As such, when the MNBs handover to a new cell in the cell array, they send their speed and path information in the first communication to the target eNB. We assume the speed of the train in one cell is relatively constant. Each time the MNB sends a Channel Quality Index (CQI) feedback of the received signal quality, the eNB receives the position of the MNB as well. We now present a scheduling algorithm that uses the information of the HST provided in the cell array architecture for high throughput scheduling within an LTE infrastructure cell.

**Optimal Scheduling**

Let $N$ be the number of users in a cell, and $T$ be the scheduling period in which we target the maximum rate. A scheduling period is a duration $T$, during which we allocate $K = T \times F$ Resource Blocks (RB) to the LTE cell users. $F$ is the number of available frequency subcarriers. In the 3GPP LTE standard, a Resource Block (RB) is the smallest allocation unit in the LTE OFDM radio resource scheduling. Each RB can be independently modulated and assigned to a user. Each RB contains 6 or 7 symbols, with 12 adjacent sub-carriers of 15KHz (=180 kHz) in the frequency domain. Two resource blocks form one time slots (1ms) in the time domain. The Modulation and Coding Scheme (MCS) decides the number of bits to be transmitted in each RB. The scheduling period is 10 ms in an LTE network. Variable $x_{ift}$ is the scheduling variable, indicating the final scheduling decision. Value of the scheduling variable $x_{ift}$ is 1 if the RB on the frequency block $f$, and time slot $t$ is assigned to the user $i$, and 0 otherwise. The rate maximization problem in an LTE cell, containing
CHAPTER 4. CELLULAR SEAMLESS CONNECTIVITY FOR HIGH SPEED MOBILITY

36

where \( n_m \) is the number of passengers in cabin \( m \) of the HST. \( a_{ift} \) is the scheduling coefficient, which indicates the importance of each RB on time \( t \) and frequency subcarrier \( f \) to the user \( i \). \( a_{ift} \) can be defined as:

\[
a_{ift} = p_{ift} \times c_{ift} \times h_{it}
\]

where \( p_{ift} \) is the scheduling preference coefficient. The scheduling preference coefficient indicates the user feedback element in the scheduling coefficient. We use Channel Quality Index (CQI), speed and direction of mobile users to estimate the \( p_{ift} \) during the scheduling period \( T \). Every user \( i \) sends a \( \overrightarrow{CQI}_i = \{CQI_{ift}\}^{F \times T} \) feedback vector containing supported \( CQI \in \{0, ... , CQI_{max}\} \) values for \( F \times T \) RBs to the eNB after receiving a scheduled RB (in LTE \( CQI_{max} = 15 \)). The CQI value is an integer that represents effective Signal to Interference and Noise Ratio (SINR) as observed by UEs (the MUEs or the MBNs for the HST). The UE can provide the CQI values for the whole frequency band or a number for each selected frequency subcarrier. We assume frequency selective CQI, where the CQI values is reported for each frequency subcarrier. We also assume that an eNB is capable of calculating speed and direction of the movement of a user based on its previous and current position in the cell, which is available to the eNB. A user provides its speed information during the registration to a cell. We define speed vector \( \overrightarrow{S} = \{s_i\}^{1 \times N} \). Speed values can be up to the maximum speed a cell can provide connectivity to: \( 0 \leq s_i \leq S_{max} \). We define the direction vector with \( \overrightarrow{E} = \{e_i\}^{1 \times N} \). \( e_i \in \{-1, 0, 1\} \) shows if a user is moving toward or away from high signal area, or there is no change in its signal region.

\[
e_i = \left[ \frac{Avg(CQI_i) - Avg(\overrightarrow{CQI_i})}{avg(CQI_i)} \right]
\]

In normal mobility patterns, the changes in the channel conditions is not fast enough to change the channel and signal quality from the last CQI report to the transmission time. Therefore, scheduling
is based on the last reported CQI of the user. This CQI value which is based on the user’s last
received data is quite accurate for scheduling purposes in normal mobility patterns. In fast mobility
scenarios, although physical layer is able to combat Doppler effect and provide connection, but the
fact that the CQI might be changed due to fast movement can reduce throughput. To address this
problem, we try to predict the values of CQI on the transmission time:

\[ \overrightarrow{CQI}_i = (1 - |\alpha_i|) \times \overrightarrow{CQI}_i + \alpha_i \times \overrightarrow{CQI}_i^2 / \overrightarrow{CQI}_i \]  

(4.5)

where \( CQI_i \) are old CQI values. \( 0 \leq \alpha \leq 1 \) brings the speed and direction into account. It is
computed as follows:

\[ \alpha_i = 1/2 \times \frac{s_i}{S_{\text{max}}} \times e_i \]  

(4.6)

We do not want to completely ignore the new CQI values. Therefore, we use \( 1/2 \) coefficient in
calculating the \( \alpha \). At the fastest speed of the train, the newly received CQI values will weight for
the half of the forward CQI value. We use these CQI values as well as Quality of Service (QoS)
parameters of each user to calculate \( p_{i_{ft}} \) values:

\[ p_{i_{ft}} = CQI_{i_{ft}} \times q_i^s \]  

(4.7)

where \( q_i^s \) is the set by the quality of service parameters. \( c_{i_{ft}} \) shows the potential capacity of an RB
if assigned to user \( i \). \( c_{i_{ft}} \) is the MCS value which will be acquired based on the CQI feedback of the
user. \( c_{i_{ft}} \in \{0, ..., 64\} \).

Finally, \( h_{it} \) provides the handover probability of the user \( i \) in time \( t \). \( 0 \leq h_{it} \leq 1 \) shows if user \( i \)
is currently in the cell or performing a handover to or from the cell. If it is completely in the cell at
time \( t \), the probability should be 1 (or higher than a preset threshold).

The presented scheduling formulation for high speed trains is a 0-1 integer programming prob-
lem, therefore, NP-hard. It can hardly be implemented in realtime because scheduling decisions
must be carried out in every subframe, that is, every 10\( ms \) of LTE scheduling time. Therefore, it is
necessary to further simplify the formulation model.
Real-time Scheduling

We convert the scheduling problem presented in equation (4.2) to a Linear Programming (LP) formulation presented in equation (4.8).

\[
\text{Maximize} \quad \sum_{i=1}^{N} \sum_{f=1}^{F} \sum_{t=1}^{T} a_{ift}x_{ift} \quad (4.8)
\]

Subject to

\[
\sum_{i=1}^{N} \sum_{f=1}^{F} \sum_{t=1}^{T} x_{ift} \leq F \times T
\]

\[
\sum_{f=1}^{F} \sum_{t=1}^{T} x_{ift} \geq 1, \forall i \in \{0, ..., n_m\}
\]

\[
0 \leq x_{ift} \leq 1
\]

This formulation does not exactly specify the download RB allocated to each. Since \(x_{ift}\) can take values between 0 and 1, each block may be assigned to more than one user with fractional values. To avoid this, in algorithm 2, we run a Weighted Round Robin (WRR) on users to select the user to be scheduled next. The LP-based scheduling is then solved for each user \(W_i\) times and the highest values of variables found for user \(i\) are selected as the RBs assigned to that user. If the highest \(x_{ift}\) is on the RB already assigned, the next best values \(x_{ift}\) for that user is used. The decided variables are omitted from the LP and their \(x_{ift}\) values is assumed as 1. The LP is solved again for the remaining variables and the same process continues. Since the scheduling in the LTE-based systems takes place every 10 ms and the subcarrier placing is 15 KHz, all LTE users in the cell may not be allocated an RB in one scheduling duration. That is why if we have all the \(F \times K\) RBs already allocated to users, we finish the scheduling cycle and start a new one. To have even a better time complexity, we can select all \(W_i\) variables for user \(i\) in one run of the LP.
Note that we use the same algorithm to schedule both HST passengers in the cell and the normal LTE users. The only difference is in how we compute the $a_{ijt}$ values for these two types of users. For the normal users, $a_{ijt}$ can simply be a CQI report without update with speed and direction information.

Algorithm 2: Scheduling Algorithm

| Inputs: $M =$ Number of MNBs = Number of Cabins; $N =$ Number of users in the cell, including $M$; $W_i =$ Aggregate weight of user $i$; $\overset{\scriptstyle\text{CQI}}{\text{avg}}$ values for $N$ users; |
| for each scheduling duration do |
| Reset Counter; |
| Sort users based on their $CQI_{\text{avg}}$ values; |
| for $i=0$ to $N$ do |
| for $w=1$ to $W_i$ do |
| Solve updated scheduling LP 4.8; |
| Choose highest $x_{ijt}$ value for user $i$ with available RB; |
| Assign $x_{ijt} = 1$ to the selected value; |
| Select that RB for user $i$; |
| Mark $RB_{ijt}$ as used; |
| Counter++; |
| Omit that $x_{ijt}$ from LP and user $i$ from sorted list; |
| if Counter = $F \times K$ then |
| Break; |

4.2 Performance Evaluation

In this section we evaluate our solution and also compare it with the state-of-the-art LTE solutions. There are a number of open source simulation software available for LTE networks in the link and system levels. However, to our knowledge, none of them is capable of simulating LTE networks for the high mobility patterns, that is, higher than 120 kmph. The high speed mobility patterns need simulation of a large number of cells to examine the effect of frequent handovers. A detailed implementation of the handover process and an accurate yet abstract system-level simulation of the cellular environment for HST is therefore expected.

To this end, we developed our own simulator for HST in LTE in C++ and MATLAB, providing system-level simulation of HST wireless access in the LTE networks with speeds up to 500 kmph. The first part of our simulator consists of a detailed yet simplified model of the LTE physical layer. This part simulates train movement, and geographical placement of cells. It computes physical
layer SINR, path loss, and frequency selective CQI feedback values. Programmed in C++, at a higher layer, it also provides detailed implementation of the LTE handover mechanism that monitors handover request times and granted times. We use the time and frequency selective CQI feedback values for HST and individual LTE users from the C++ code output and run the scheduling part with MATLAB. This way, scheduling is not always performed in real-time during the simulation, but generally finished in an acceptable timeframe. Therefore, our simulation is composed of two steps: running the C++ code and getting LTE parameters on each part of the map, and then providing the output file from the C++ simulation to the scheduling provided in MATLAB where the scheduling for the video trace file is done using the LTE information and trace file data using the proposed algorithm. The movement paths across the LTE cell map is defined with a line or a curve that crosses multiple cells along a railway. This path can be input using the parametric representation of the line or the curve. Both the MATLAB and C++ parts of our simulation codes are publicly available on http://www.sfu.ca/~oba2/dls.

We used Variable Bit Rate (VBR) High Definition (HD) video streams as the source traffic, which are randomly chosen for each user and downloaded for each user in unicast. Our video sources is the single layer Sony Demo coarse grained SVC trace file in HD, 30 frames/sec from the Video Trace Library [13] [110].

4.2.1 Simulation Settings

Throughout our simulation, each high speed train cabin is 25 meters, accommodating 100 users, and is equipped with two LTE femto cell MNBs. We run simulations for 1 to 10 cabin combinations, and in both urban and rural settings, where size of an urban cell is 500 meters, and that of a rural cell is 10km. Table 4.2 shows these simulation settings, which are based on practical values available in the current HST implementations. Figure 4.6 shows the geographical map of three typical paths we used.
for simulation of HST movement to cover different cell crossing scenarios. We also run simulations for other random movement patterns to find out the average latency, delay and throughput and their relation to movement patterns. All of the simulations are carried our during a 5 min simulation time, within which 10-200 handovers takes place in different scenarios.

### 4.2.2 Results and Analysis

#### Handover Latency

Our solution provides significantly lower handover latency compared to the normal LTE with seamless re-entry. Figures 4.9 - 4.12 show a detailed comparison of the handover latency values in six different simulation settings, which cover speeds of 200-500, paths 1, 2, and 3 in both urban and.
Figure 4.8: Urban Cell, Speed: 350kmph, Path:3. (a) Delay (ms) (b) Throughput (Mbps) (c) Handover Latency (ms) (d) SINR (dBm) (e) Handover Requests

Figure 4.9: Urban Cell, Speed: 200kmph, Path:2. (a) Delay (ms) (b) Throughput (Mbps) (c) Handover Latency (ms) (d) SINR (dBm) (e) Handover Requests
CHAPTER 4. CELLULAR SEAMLESS CONNECTIVITY FOR HIGH SPEED MOBILITY

Figure 4.10: Urban Cell, Speed: 350kmph, Path:2. (a) Delay (ms) (b) Throughput (Mbps) (c) Handover Latency (ms) (d) SINR (dBm) (e) Handover Requests

Figure 4.11: Urban Cell, Speed: 500kmph, Path:3. (a) Delay (ms) (b) Throughput (Mbps) (c) Handover Latency (ms) (d) SINR (dBm) (e) Handover Requests
rural cells. The throughput and delay values are shown for transmission of the video trace file that simulates a video with high transmission volume at times 15-47. We can see that parameter variations have less impact on latency value in our solution. This is because the cell array architecture ensures the selection of the target cell well ahead of the hard handover time, eliminating the need for searching for the target cell in the MNB, and that for negotiating the hard handover time in the eNB.

A handover latency is a summation of different delays in the process of registering to a new cell:

\[ T_{HOLatency} = T_{HO} + T_{re-entry} + T_{re-establish} \] (4.9)

\( T_{HO} \) is the handover time which is the sum of the cell search and negotiation, registration and random access times. If the handover fails, the user should go through a re-entry process. \( T_{re-entry} \) is the time it takes for the re-entry. \( T_{re-establish} \) is the time it takes to re-establish a connection if it was dropped because of a long handover latency. Since the handover process in our cell array simplifies the search for a new target cell, all of these times are considerably shorter. This also increases the probability of success in each step, and hence possibility of our HST to successfully register to a cell within the \( T_{HO} \) deadline. Although the handover latency is not highly variable in different speed patterns, users with higher mobility experience longer handovers. Figure 4.13 illustrates the handover latency changes with increasing the HST speed. The handover latency noticeably increases in the normal LTE access. This is because of the need for frequency handovers and the need for the
The handover latency in our solution is lower and increases more smoothly as the speed increases. Besides the better prediction, the MNB acting as a single user also contributes to this.

**Handover Failure Rate**

Figure 4.14 illustrates the handover failure rate in different movement speeds. Without surprise, the failure rate increases in higher speeds. This is because of the short duration a user is residing in a certain cell area. In the normal LTE, this usually happens when a handover initiated by the eNB fails and the UE has to start a re-entry process. While a user have to search and find a new eNB during a re-entry process, it might reach the cell edges already. In this case, the user may not be able to establish the connection to the found cell before it enters the new cell, thus breaking the connection.

As we discussed before, the predictive handover mechanism in the cell array eliminates the need for the search for a new eNB when the HST is in the cell. Thus, the handover latency is considerably shorter and the probability that the handover is successful in the target cell increases.

**Throughput**

Figure 4.15 shows the throughput of our proposed solution, as compared to those of the individual HST passengers. With increasing the speed, the enhancement of the proposed algorithm experiences an improvement of 10% to 35%. This is because of two main reasons: (1) the decreased latency (as discussed above), and (2) the handover process does not take place for each user. The latter, decreases the registration overhead that affects the RB allocation for a new user in a cell. For the
HST in our algorithm, the handover only happens once for each MNB and an aggregate traffic is requested for the RB allocation.

![Average Throughput](image1)

![Average Delay](image2)

**Figure 4.15: Average Throughput**

**Figure 4.16: Average Delay**

Figures 4.7 - 4.12 also provide a detailed comparison of the throughput in our six simulation settings. The throughput of the proposed solution experiences lower fluctuations with changing path and speed. There is an obvious drop in throughput in the normal LTE performance in the urban area compared to that of the rural cells. Our solution only experiences a slight drop only. This is due to the more frequent handovers while moving across smaller urban cells. This minor drop in throughput in our solution is largely caused by the lower SINR values in the urban areas due to higher interferences.

**Delay**

Figure 4.16 compares the delays of the proposed solution. It experiences 5%-18% lower delays mainly due to the lower handover latency for each user. However, we see some sudden increases in delay when a handover takes place. These delays are for data forwarding to the new cell. Since the delay is mostly on the non-delay sensitive data, it is negligible.

More detailed delay values for different simulation settings are also provided in figures 4.9 - 4.12.

**4.3 Summary and Discussion**

In this chapter, we reviewed the state of the art technologies for wireless access of high speed vehicle passengers. We presented a novel LTE-based solution to support high throughput and continuous
multimedia services for high speed train passengers. Our solution is based on a Cell Array that smartly organizes the cells along a railway, together with a femto cell service that aggregates traffic demands within individual train cabins. Given that the movement direction and speed of a high-speed train are generally known, our Cell Array effectively predicts the upcoming LTE cells in service, and enables a seamless handover that will not interrupt multimedia streams. To accommodate the extreme channel variations, we further propose a scheduling and resource allocation mechanism to maximize the service rate based on periodical signal quality changes. Our simulation under diverse network and railway/train configurations demonstrates that the proposed solution achieves much lower handover latency and higher data throughput, as compared to existing solutions. It also well resists to network and traffic dynamics, thus enabling uninterrupted quality multimedia services for passengers in high speed trains.

This chapter concludes our discussion on resource efficiency on cellular wireless networks. In the next chapter, we will address the problems of spectrum shortage for cellular networks by introducing target networks as offloading solutions that should be incorporated with cellular wireless networks, in addition to enhancing resource utilization in these networks, for high-throughput pervasive wireless access.
Chapter 5

Cooperative Access in Dense WMNs

Large number of users in urban areas and their increasing demand for high bandwidth data applications on the move increases the demand for mobile data, resulting in an 13-fold increase by the end of 2017 [40]. Although the 3G and 4G cellular networks are adopting optimized technologies to maximize spectrum usage efficiency and also can support a higher number of users in each cell, the limited spectrum cannot accommodate the number of current high bandwidth users. Furthermore, spectrum usage improvement in the new technologies is not happening as fast as the increase in the mobile data volume.

Mobile data users utilize 802.11(WiFi) interfaces of their mobile devices to offload some of their high volume data usage off the cellular networks to inexpensive, or sometimes free, WiFi resources. Mobile devices have the ability to switch to 802.11 based networks whenever available using vertical handover methods [97] [8]. 802.11-based wireless mesh networks are one of the target networks for offloading. However, offloading requires availability and effective management of resources for high-throughput applications.

In this chapter, we study high-throughput multicast that combats the interference and bandwidth limitation of wireless channels, which are particularly severe with wireless mesh networks. We suggest that they can be addressed by introducing multiple cooperative mesh gateways and exploiting the diversity of wireless channels. We present a cross-layer design that jointly selects appropriate channels for each mesh node to use at judiciously tuned power, and computes the optimal multicast flows from multiple cooperative gateways. We show that this design can be iteratively optimized through Lagrange relaxation and primal-dual decomposition. A progressive channel assignment and power level adjustment heuristic is introduced in the MAC/PHY layer, together with a smart link capacity allocation for cooperative gateways in the network layer. Through extensive simulations, we
demonstrate the effectiveness of the proposed solution framework and the sub-problem heuristics. In particular, a throughput improvement of up to 100% is observed compared to straightforward approaches of utilizing multiple wireless channels for multicast routing [15] [16].

Our work was motivated by many pioneer studies in channel assignment and multicast routing; yet our focus is mainly on throughput maximization in the multicast context, completing the previous researches [15] [70]. For multicast routing, Nguyen and Xu [93] systematically compared the conventional minimum spanning trees or shortest path trees in wireless meshes. Novel approaches customized for wireless meshes have also been proposed [107] [139] [138]. Our work is closely related to two of them. In the work of Zeng et al. [139], two heuristics for multicast channel assignment were proposed, which also applies to a multi-gateway configuration. They, however, did not explicitly address route optimization. In the work of Yuan et al. [138], routing and wireless medium contention were jointly considered. The impact of link interferences and power amplitude variations on each link were also closely examined, but were limited to single channel usage. Our work differs from them in that we examine both multicast routing and channel assignment in a coherent cross-layer framework, and present effective solutions. We also explicitly explore the potentials of multi-gateway configurations.

We consider two techniques for addressing the high-throughput requirement of multicast applications in wireless mesh networks. The first is to use multiple gateways. A gateway is directly connected to the Internet, and hence serves as the data source for users in a wireless mesh. A single gateway design makes the gateway node a bottleneck, and is prone to congestion during high network activities. Having multiple gateways can dramatically improve the network performance at a reasonable cost. These gateways can collaboratively serve their clients using minimal signalling among the gateways. The second is to exploit the diversity in wireless channels, and provide a multi-channel multicast solution. Wireless interference is a critical limitation on throughput of wireless mesh applications [75]. Utilizing distinct channels at neighbouring nodes for transmission can help reduce interference to minimum. For example, the IEEE 802.11b/g protocol defines 13 channels within a 2.4GHz frequency band [35]. The further apart two channels are, the lower interference exists between them; in particular, channels 1, 6 and 13 are totally orthogonal. However, only considering the orthogonal channels will result in collision and reduced throughput, especially in more dense networks like urban areas. We consider using non-orthogonal channels and show that judiciously assigned, this selection will result in lower interference and enhanced performance, especially in dense deployments. The applications for dense wireless meshes include mesh networks for broadband wireless connection in dense urban areas with no wired network infrastructure.
We first formulate the multi-gateway multi-channel multicast problem in wireless meshes as a mathematical programming problem, which jointly considers channel assignment and transmission power tuning at the MAC/PHY layer, as well as multicast routing at the network layer. Two important regions that the formulation is based on, the channel capacity region and the routing region, are both convex. Furthermore, the objective function that models the utility of multicast throughput is strictly concave. Therefore, the entire optimization model we obtain is a convex program, if we can freely select the frequency band for a channels. However, with pre-defined channels such as in IEEE 802.11, the optimization model contains discrete variables, which complicates the solution design.

In order to provide an efficient and practical solution to the optimization model, we apply the classic Lagrange relaxation technique [26] [138], and derive an iterative primal-dual optimization algorithm that leads to a cross-layer multicast solution. Towards this direction, we first relax the link capacity constraints that couple the channel region and the routing region, and decompose the overall optimization into two smaller sub-problems, one for channel assignment at the MAC/PHY layer, and one for multicast routing at the network layer. Our primal-dual solution framework then iteratively refines the primal solution, with help of the Lagrange dual that signalizes capacity demand at each wireless link. The dual is updated during each iteration based on the latest primal solutions.

To complete the solution defined by the primal-dual framework, we need to precisely define the channel region and the routing region, and design a solution algorithm for each of the channel assignment and routing sub-problems. We formulate the channel assignment problem as a mathematical program, in which channel capacities are computed from their signal-to-noise-and-interference ratio (SINR), and the computation of SINR in turn appropriately takes into account the separation between different wireless channels used at neighbouring mesh nodes. The main challenge in solving this mathematical program lies in the presence of discrete channel assignment variables. We design an efficient heuristic, progressive channel assignment, for overcoming this difficulty. Finally, we discuss both multicast tree based and network coding based solutions for the multicast routing sub-problem. Extensive simulations, with various network sizes, were conducted for evaluating the effectiveness of both the overall primal-dual optimization framework and the sub-problem solutions. Throughput improvement of up to 100% were observed, when the proposed solution is compared to straightforward channel assignment schemes such as orthogonal channel assignment and consecutive channel assignment.
5.1 Motivating Example

In this section, we present an example to illustrate the importance of multicast channel assignment in multi-channel wireless networks. Figure 5.1 illustrates a mesh network consisting of 25 nodes within a $150m \times 150m$ area. The network contains four gateways providing the connectivity to the same multicast source. Multicast receivers, D1-D7, each wishes to connect to one of the available gateways to access multicast data. We assume using 802.11g in an indoor environment. The effective communication range of each wireless link in 802.11g indoor mode is about $35m$, and the interference range is about three times the communication range.

A simple solution to this multi-gateway problem is hop-count based multicast, which builds a multicast forest in the wireless network by connecting each receiver to its closest gateway. This results in connections from D1, D5 and D6 to G1, from D2 to G2, from D3 to G3 and from D7 and D4 to G4. Orthogonal channel assignment on such multicast routes is shown in Figure 5.1(a). Numbers on the active wireless links represent the channel assignment on each link.

Orthogonal channel assignment will assign channel 1, 6, 11 with lowest interference detection on each, starting from the source, extending all the way to the destinations. The throughput on the links to the multicast receivers D5 to D7 will be rather low, imposing the multicast rate limit on all nodes on the same multicast tree. The low rate is due to the fact that the same channel is re-used in the interference range. If the scanning and scheduling techniques are implemented, the throughput will be as low as half of average 802.11g throughput.

Given the same flow routing scheme in the wireless network, our proposed channel assignment algorithm, discussed in detail later, will instead produce a solution shown in Figure 5.1(b). We note that this solution does not insist on using orthogonal channels only, and systematically selects one of the available channels for each link in the flow, for minimizing global interference. As we will discuss later, further coupled with power adjustment module proposed in this work, our channel assignment algorithm can enhance the multicast throughput up to $3/4$ of average 802.11g throughput.

Similar network topology and scenario may exist in crowded, dense urban area wireless networks. The number of wireless users and the interference levels will also be substantially higher in such scenarios due to the fact that usually no node in a wireless meshes is idle. Higher interference levels might also be a result of other working wireless devices in $2.4GHz$ range. Our proposed algorithm increases the multicast throughput up to 100% in such scenarios. Later, we will discuss in detail the improvements each part of our algorithm will make on these specific examples.
5.2 The Multi-Channel Multicast Problem Formulation

We first construct mathematical programming formulations of the optimal multicast problem in wireless mesh networks, with multi-gateways and multi-channels. Envisioning two different physical layer technologies for selecting a frequency band for a channel, we present two corresponding optimization models. The first one is based on flexible frequency bands enabled by variable frequency oscillators, such as assumed in software-defined radios. This ideal radio model leads to optimal multicast throughput that can be computed precisely, through the classic primal-dual optimization framework. The second model is rather similar, but makes a more realistic assumption on frequency bands based on the state-of-the-art IEEE 802.11 standard: each transmission has to use one of the 13 pre-defined channels.

5.2.1 Network Model and Notations

We model a wireless mesh networks as a graph $G = (V, E)$, with nodes $V$ and links $E$. Assume $T \subseteq V$ is the set of collaborative gateways. Each gateway has a high-bandwidth connection to the Internet, and can be viewed as a data source. Let $S$ be the set of data transmission sessions. We define five vectors of variables. The first four are: the vector of data flows $f = < f_i^e | i \in S, e \in E >$; the vector of multicast throughput $r = < r^i | i \in S >$; the vector of link bandwidth capacities
Table 5.1: Variables of Channel Assignment and Routing Sub-problems

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_{ee}$</td>
<td>Gain associated with link $e$</td>
<td>$\gamma$</td>
<td>Interference factor $= \sum_{\gamma' \in \Gamma} I_{\gamma'}/d_{\gamma}$</td>
</tr>
<tr>
<td>$p_e$</td>
<td>Power on link $e$</td>
<td>$\Gamma(v)$</td>
<td>Set of channel activated on node $v$</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>Noise on line $e$</td>
<td>$\Gamma$</td>
<td>Set of all available channels</td>
</tr>
<tr>
<td>$l_{le}$</td>
<td>Correlation coefficient from $l$ on $e$</td>
<td>$r^s$</td>
<td>Multicast rate from source $s$</td>
</tr>
<tr>
<td>$G_{le}$</td>
<td>Interference coefficient from $l$ on $e$</td>
<td>$G$</td>
<td>Set of all gateways in the network</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Dual variable vector</td>
<td>$f_e^s$</td>
<td>Flow on link $e$ from source $s$</td>
</tr>
<tr>
<td>$c_e$</td>
<td>Capacity on link $e$</td>
<td>$e_{s,j}^c$</td>
<td>Conceptual flow on link $e$ from $s$ to $j$</td>
</tr>
<tr>
<td>$b$</td>
<td>Bandwidth</td>
<td>$O(v)$</td>
<td>Set of outgoing links from node $v$</td>
</tr>
<tr>
<td>$p_{v,max}$</td>
<td>Total power budget on node $v$</td>
<td>$I(v)$</td>
<td>Set of incoming links on node $v$</td>
</tr>
</tbody>
</table>

$c = < c_e | e \in E >$; and the power assignment vector $p = < p_u \leq p_{u,max} | u \in V >$. The last one is on channel assignment. We assume that each node is equipped with one radio with capacity $b$, which can transmit at different frequencies with adjustable power. $b$ is the frequency bandwidth the radio has available for transmission. In the flexible channel model, we have the vector of centre frequencies $\mu = < \mu_u | u \in V >$. The frequency band of the channel used by $u$ is then $[\mu_u - b/2, \mu_u + b/2]$. In the case of fixed channels, we have the vector $\gamma = < \gamma_u \in \Gamma | u \in V >$ to represent the channel assignment at each node. Here $\Gamma$ represents the set of pre-defined channels, such as the 13 in the IEEE 802.11 standard. Table 5.1 summarizes the variables used in the formulation.

5.2.2 The Flexible Channel Model

Two capacity regions are fundamental to our multicast problem formulation: the channel region and the routing region, at the MAC/PHY layer and the network layer, respectively. The channel region $H$ defines a set of $(c, h)$ such that channel assignment in $h$ can support link capacity vector $c$. The routing region $R$ defines a set of $(r, f)$ such that the throughput vector $r$ can be supported by flow rates in $f$. Detailed characterization of the two regions are not immediately relevant to the overall optimization structure, and are postponed to Sec 5.3.4 and Sec 5.3.5 respectively, where we select optimal solutions from each region. The multicast throughput for each session is measured as the data receiving rates at the receivers, which are equal for receivers across the same session. A basic physical rule that establishes a connection between the routing region and the channel region is that the aggregated data flow rates have to be bounded by the corresponding link capacities. Furthermore, we follow the convention [138] in modeling throughput utility, and adopt the concave utility function $\log(1+r_i)$ for session throughput $r_i$. Then, the throughput maximization problem can be formulated...
as:

\[
\begin{align*}
\text{Maximize} & \quad U(r) = \sum_{i \in S} U(r_i) = \sum_{i \in S} \log(1 + r_i) \\
\text{Subject to} & \quad (c, \mu, p) \in H \\
& \quad (r, f) \in R \\
& \quad \sum_{i \in S} f^i_e \leq c_e, \forall e \in E
\end{align*}
\]  

The first constraint \((c, \mu, p) \in H\) models the dependence of effective channel bandwidth on channel assignment and power assignment at each node. The second constraint \((r, f) \in R\) models the dependence of multicast throughput \(r\) on the routing scheme \(f\). \(\sum_{i \in S} f^i_e \leq c_e, \forall e \in E\) model link capacity constraints. The objective function \(U(r)\) is concave, and both the routing and channel regions are convex regions. Therefore, convex optimization methods [26] can be used to compute the optimal solution \((\mu^*, p^*, f^*)\). In section 5.3 and section 5.3.3, we present a primal-dual solution based on Lagrange relaxation and iterative primal-dual optimization.

If nodes can transmit using pre-defined channels only, we can modify the mathematical program in 5.2, by replacing the frequency vector \(\mu\) with the channel assignment vector \(\gamma\). Since \(\gamma\) is an integer vector, the mathematical program can not be directly solved to optimal using conventional convex optimization methods, in polynomial time. Nonetheless, the solutions in section 5.3 and section 5.3.3 will be flexible enough to compute approximate solutions, based on a heuristic channel assignment method.

## 5.3 The Primal-Dual Solution Framework

The overall solution framework we propose for solving optimization problem 5.2 is an iterative primal-dual schema, which switches between solving primal sub-problems and updating dual variables. We describe in section 5.3.1 how to decompose the primal problem while introducing dual variables, and then present the primal-dual solution framework in section 5.3.2.

### 5.3.1 The Routing vs. Channel Assignment Decomposition

A critical observation of the optimization problem 5.2 is that, the channel region \(H\) and the routing region \(R\) characterize variables from the MAC/PHY layer and the network layer respectively, and are relatively independent. The only coupling constraint between them is \(f \leq c\). We can apply the
Lagrange relaxation technique \cite{26} \cite{77} to remove $f \leq c$ from the constraint set, and add a corresponding price term into the objective function: $L = U(r) + \sum_{e \in E} \alpha_e [c_e - \sum_{i \in S} f_{i e}]$. Here $\alpha$ is a vector of Lagrange multipliers, which can be viewed as prices governing the link capacity supply - the larger $\alpha_e$ is, the tighter bandwidth supply at link $e$ is. After the relaxation, the resulting optimization problem is naturally decomposed into two smaller, easier-to-solve sub-problems, including the Channel Assignment Sub-problem at the MAC/PHY layer:

$$\text{Maximize} \quad \sum_{e \in E} \alpha_e c_e$$

$$\text{Subject to} \quad (c, \gamma, p) \in H$$

and the Routing Sub-problem at the network layer:

$$\text{Maximize} \quad U(r) - \sum_{e \in E} (\alpha_e \sum_{i \in S} f_{i e})$$

$$\text{Subject to} \quad (r, f) \in R$$

It is interesting to observe that, given a link $e$ with high price $\alpha_e$, the routing sub-problem will automatically attempt to reduce the amount of flow $f_e$ through $e$ during the next round, since its objective function implies minimizing $\sum_{e \in E} \alpha_e \sum_{i \in S} f_{i e}$. On the other hand, the channel assignment sub-problem will automatically attempt to create more capacity for $e$, since its objective function is to maximize $\sum_{e \in E} \alpha_e c_e$.

5.3.2 The Primal-Dual Solution Schema

The primal-dual approach iteratively updates the primal ($f$, $\mu$, $p$) and dual ($\alpha$) solutions. During each iteration, we solve the two primal sub-problems given the current dual vector $\alpha$, and subsequently update $\alpha$ with the newly computed primal vectors as below. Here $t$ is the round number, and $\beta$ is the step size vector.

i. Set $t = 1$; initialize $\alpha(0)$, e.g., set $\alpha_e(0) = 0, \forall e \in E$

ii. Solve primal sub-problems (2) and (3).

iii. Update the dual domain variables as below:

$$\alpha(t) = \max(0, [\alpha(t - 1) + \beta(t) (\sum_{s \in G} \sum_{t \in T_s} f_{t e} - c_e)])$$
iv. Set \( t = t + 1 \) and return to step ii, until convergence.

The primal-dual algorithm above converges to an optimum primal solution \((f^*, \mu^*, p^*)\). of the optimization problem (5.2), as long as the regions \( R \) and \( H \) are convex and the step sizes \( \beta(t) \) are appropriately chosen.

Proof: The constraint \( f < c \) is linear, the objective function in optimization 5.2 is strictly concave. The convexity of the capacity regions \( R \) and \( H \) then ensures that the update in the dual domain (iii) is a sub-gradient for the dual variables in \( \alpha \). Therefore as long as the step sizes are appropriately chosen, the dual update converges [77] [138]. Strong duality further assures that the convergence point of the primal-dual algorithm corresponds to a global optimum of the network optimization problem in (5.2).

\[
\beta[t] \geq 0, \lim_{t \to \infty} \beta[t] = 0, \text{ and } \sum_{t=1}^{\infty} \beta[t] = \infty. \text{ A simple sequence that satisfies the conditions above, is } \beta[k] = a/(mk + n), \text{ for some positive constants } a, m \text{ and } n.
\]

### 5.3.3 Solving Channel Assignment and Routing Sub-Problems

In order to obtain a complete solution under the primal-dual schema, we need to design algorithms for solving each of the two primal sub-problems. We next discuss how to solve the channel assignment sub-problem in section 5.3.4, and the routing sub-problem in section 5.3.5.

#### 5.3.4 The Channel Assignment Sub-problem

We now construct a detailed model for the channel capacity region \( H \), and discuss how the resulting channel assignment problem from (2) can be solved. The effective link bandwidth capacity are determined by the signal-to-noise-and-interference ratio (SINR) of the transmission; following the Gaussian channel capacity model [138]:

\[
ce_e = b \log_2(1 + SINR_e), \quad SINR_e = \frac{G_{ee}P_e}{(\sum_{l \neq e} I_{le}P_l + G_{te}) + \sigma^2}
\]

Here \( G_{ee}, P_e \) and \( \sigma^2 \) are gain, power and noise associated with a link respectively. \( G_{le} \) and \( \sigma^2 \) denote the interference coefficient and noise from link \( l \) to link \( e \) respectively. \( I_{le} \) is the channel correlation coefficient, which depends on the separation between channels used by \( l \) and \( e \), e.g., the separation between channels 1 and 4 is 3.

The correlation \( I_{\gamma \gamma'} \) between two channels \( \gamma \) and \( \gamma' \) are known for all possible channel separations, as shown in figure 5.2 [12]. From this figure, the correlation between any two channels,
either flexibly selected or pre-defined, can be found. For example, for the 13 IEEE 802.11 channels, $I_{\gamma_1 \gamma_1} = 1.0$, $I_{\gamma_1 \gamma_2} = 0.7906$, $I_{\gamma_1 \gamma_3} = 0.5267$, and $I_{\gamma_1 \gamma_7} = 0$. Furthermore, we assume the total budget at each node $v$ is $p_{v, max}$ and $O(v)$ is the set of outgoing links from node $v$. Then, the channel assignment problem for capacity maximization can be formulated as:

$$\text{Maximize } \sum_{e \in E} \alpha_e c_e$$

subject to

$$c_e = b \log_2 (1 + SINR_e), \forall e \in E$$

$$SINR_e = \frac{G_{ee} p_e}{(\sum_{l \neq e} I_{\gamma_l \gamma_e} p_l G_{le}) + \sigma^2}, \forall e \in E$$

$$\sum_{e \in O(v)} p_e \leq p_{v, max}, \forall v \in V$$

$$\gamma_e \in \Gamma, \forall e \in E$$

Without the discrete variables in $\gamma$, (4) can be solved using known techniques, such as using geometric programming or through a power control game [138]. The new challenge in (4) is to compute good channels for vector $\gamma$. We present a heuristic solution for efficiently solving $\gamma$, and evaluate its performance later in section 5.4.

**Heuristic Channel Assignment Algorithm**

We design a heuristic channel assignment algorithm based on the interference factor $\varphi_\gamma = \sum_{\gamma' \in \Gamma} I_{\gamma \gamma'}/d_{\gamma'}$, for a given candidate channel $\gamma$. Here $d_{\gamma'}$ is the distance to the nearest node transmitting at channel $\gamma'$. In measurement-based systems, $d_{\gamma \gamma'}$ is not necessary because the node can sense if a channel
is in use in the range of this node and signal strength could be used instead. In coordination-based systems, \( d_{\gamma'} \) could be found based on the coordination information. We assume \( d_{\gamma'} = \infty \) if channel \( \gamma' \) is not in use in the network. Our heuristic solution, performs a breadth-first-traversal of the wireless mesh networks. At each node, candidate channels are sorted by the interference factor to already assigned channels at other nodes. Different options are possible in selecting the channel. A greedy algorithm selects the channel \( \gamma \) with the smallest \( \varphi_{\gamma} \) value, i.e., as apart from neighboring channels in use as possible. We propose a progressive channel assignment approach instead, and select a channel \( \gamma \) with the highest \( \varphi_{\gamma} \) below an acceptable threshold \( \varphi_{th} \). The rational here is to look beyond channel assignment at the current node, and to leave good candidate channels for neighbour nodes.

**Algorithm 3: Progressive Channel Assignment**

**Initialization:**
\[
\begin{align*}
\Gamma(v) &:= \emptyset, \forall v \in V; \\
\varphi_{Set} &:= \emptyset; \\
\text{forall the } v \in V \text{ do} & \\
\text{forall the } \gamma \in \Gamma \text{ do} & \\
\text{Compute } \varphi_{\gamma}^v; \\
\text{if } \varphi_{\gamma}^v \leq \varphi_{th} \text{ then} & \\
\varphi_{Set} &:= \varphi_{Set} \cup \varphi_{\gamma}^v; \\
\text{if } \varphi_{Set} = \emptyset \text{ then} & \\
\text{Choose } \gamma_v \text{ with smallest } \varphi_{\gamma_v} \text{ and activate it on node } v : \Gamma(v) = \gamma_v; \\
\text{else} & \\
\text{Choose } \gamma_v \text{ from } \varphi_{Set} \text{ with largest } \varphi_{\gamma_v} \text{ and activate it on node } v : \Gamma(v) = \gamma_v; \\
\varphi_{Set} &:= \emptyset
\end{align*}
\]

Progressive channel assignment algorithm consists of a double loop. The outer loop iterates through nodes in the network, and the inner loop iterates through all possible channels. The number of channels is 13 and therefore the total number of iterations is \( 13|V| \).

**Power Adjustment Algorithm**

In the channel assignment heuristic, we assumed that all nodes use the same power levels to activate their links. In this section, we provide an algorithm to incorporate power adjustment to our channel assignment algorithm.

We propose a power adjustment algorithm based on \( \alpha_e \) values computed in the dual domain.
CHAPTER 5. COOPERATIVE ACCESS IN DENSE WMNS

When $\alpha_e < 0$ (for the power adjustment algorithm, we use the $\alpha_e$ values before resetting the negative $\alpha$ values to zero), link $e$ has its capacity higher than its assigned flow rate from the routing sub-module. For nodes with such outgoing links, we find their interfering nodes, and adjust their power. The adjustment is conducted based on the $\alpha$ value of outgoing links at each node. Therefore, interfering nodes with negative $\alpha$ values will decrease their power, while those with positive $\alpha$ values will increase their link activation power.

Algorithm 4: Power Adjustment Algorithm

Initialization:
\[
\alpha_{\text{avg}} = \frac{\sum_{e \in E} \alpha_e}{|E|};
\]
\[
P_{\text{avg}} = \frac{\sum_{v \in V} P_v}{|V|};
\]
for all the $u \in V$ do
\[
\text{if } \alpha_{O(u)} < 0 \text{ then}
\]
forall the $v \in NI(v)$ do
\[
P_v = P_v + \frac{\alpha_{O(v)}}{\alpha_{\text{avg}}} \times P_{\text{avg}};
\]
\]

The main part of power adjustment algorithm consists of a double loop. The outer loop iterates through nodes in the network, and the inner loop iterates through its interfering neighbours. The number of interfering nodes is most $|V|^2/4$, therefore the total number of iterations is upper-bounded by $|V|^2/4$. The initialization parts consists of two independent loops of $|V|$ and $|E|$ iterations. Therefore, it consists of at most $|V|^2/4 + |V| + |E|$ iterations in total.

Note that since power adjustment on a node may change the interference effect of it on other wireless links, channel assignments of a network may change during the next iteration of the primal-dual algorithm based on power adjustment updates.

5.3.5 The Routing Sub-problem

The multicast flow routing problem at the network layer has been extensively studied in the literature during the past decade. Two classes of solutions have been proposed. The first class includes multicast tree based solutions. Since achieving optimal multicast throughput using multicast trees corresponds to the NP-hard problem of Steiner tree packing, one needs to resort to efficient approximation algorithms, such as the KMB algorithm. The second class includes network coding based solutions. By assuming information coding capabilities for nodes in the network, the complexity of the optimal multicast problem decreases from NP-hard to polynomial time solvable [77]. In particular, conceptual flow based linear programming models have been successfully developed
for multicast in various network models [77, 138]. In this section, we apply similar techniques and formulate our routing sub-problem into a convex program with all-linear constraints, which can be solved using general convex optimization algorithms such as the interior-point algorithm [26], or tailored sub gradient algorithms [77]. We model flows from each of the gateways to different destinations as conceptual flows that do not compete for link bandwidth. $e_{i,j}^{s,l}$ denotes the conceptual flow rate on link $l$ in $i$th multicast session to its $j$th destination. This assumption, brings network coding assumption into the problem and helps us achieve polynomial time solvable problem. $I(v)$ is set of incoming links to node $v$ and $O(v)$ is the set of outgoing links from node $v$. The multi-gateway multicast routing sub-problem with network coding can be stated as a convex optimization problem.

Max $\ U(r) - \sum_{e \in E} \alpha_e \sum_{s \in G} f_s^e$

s.t. $r^s \leq \sum_{l \in I(j)} e_{l}^{s,j}, \forall s, \forall j, \in V$

$e_{l}^{s,j} \leq f_l^s, \forall s \in G, \forall j \in V, \forall l \in E$

$\sum_{l \in O(v)} e_{l}^{s,j} = \sum_{l' \in I(v)} e_{l'}^{s,j}, \forall s \in G, \forall v, j \in V$

$f_l^s \geq 0, e_{l}^{s,j} \geq 0, r^s \geq 0$ \hspace{1cm} (5.4)

Each source $s \in G$ models a gateway in our network. Note that the term source is used as gateway in this part and does not necessarily mean that multicast contents are different. Specifically, we assume that all the gateways transmit the same multicast content. First constraint assures that the multicast rate transmitted from each source $s$ is less than or equal to the rate of the destination with minimum conceptual multicast flow from that source. Second constraint imposes the maximum flow rate limit. Conceptual flows do not compete for link bandwidth. Therefore, their maximum, rather than sum, will put a constraint on actual flow rate on the links. Third constraint imposes the flow conservation constraint on the problem. In other words, sum of input flows from one source to a node are equal to sum of the output flows of the same source out of that node. For compact LP formulation, the convention of assuming a virtual feedback link from multicast receivers to sources is followed. Finally, the last constraint imposes the lower bounds on each optimization variable.

The utility function $U(r) = \log(1 + r)$ is a concave function and the constraints are all linear. Therefore, the routing sub-problem could be solved as a convex optimization problem in polynomial time. There exist efficient numerical algorithms to solve convex optimization problems.
The solution to this convex optimization problem utilizes multiple gateways in the network, to cooperatively multicast data to each destination. In other words, may or may not receive multicast data from more than one gateway node in the network. It does not mean that all receiver nodes will receive multicast data from all gateways; they may or may not be connected to a certain gateway. This limit is imposed by the capacity feedback from our channel assignment algorithm (update on $\alpha_e$ by primal-dual algorithm) or non-multicast traffic flows in the network. The latter could be brought into formulation by an update factor in dual variable.

Note that in wireless mesh networks nodes are not usually idle. In other words, when a node is part of a wireless mesh network, it usually has a traffic to transmit. Therefore, we are not increasing the number of nodes working in the network by involving them in the multicast. This means, involvement of more or less nodes in the multicast will not dramatically change the network interference levels. This is one of the practical reasons that the two problems of routing and channel assignment could be looked upon independently in different layers.

We have implemented the primal-dual framework and the sub-problem solutions to examine the performance of the proposed algorithm. We present our performance analysis in two simulation scenarios. In the first part we present the throughput and dual variable convergence analysis for a simple network. In the second set of simulations we analyze the primal-dual solution in different combination of nodes and network density, discuss scalability of the solution, and compare it to other routing and channel assignment algorithms for multi-gateway multi-channel wireless mesh networks.

5.4 Performance Analysis

We used CVX [42], a package for specifying and solving convex programs [42] [54], to solve the routing sub-problem. We first placed the nodes randomly in the given area (ranging from $300 \times 300 m^2$ to $1000 \times 1000 m^2$ in different experiments). Then, we build a two dimensional mesh where each node is connected to at most four neighbours. The number of nodes within each area are chosen to simulate possible mesh deployments in dense urban areas. We used random deployments and find the average of results in 10 different deployments. Simulations are conducted for both indoor and outdoor environments with transmission rages of 45 and 90 and interference range of 135 and 270 respectively. Each node is equipped with radios running IEEE 802.11 protocol and works with initial transmission power of $100mW$ tuned during the simulation. 30% of nodes in each scenario are randomly chosen as receiver nodes and max(4,10% of nodes) are randomly assigned as gateway
nodes. The link capacities are computed using \( c_e = b \log_2(1 + \text{SINR}_e) \)

In this part of simulations we first compare our channel assignment solution with greedy orthogonal and consecutive channel assignment methods. We decide on channel assignment on each node rather than analytically computing the solution centrally. In other words, the channel assignment is performed in a distributed way.

![Graph showing Channel Assignment and Multicast Throughput with Different Number of Nodes](image1)

Figure 5.3: Channel Assignment and Multicast Throughput with Different Number of Nodes

![Graph showing Channel Assignment and Multicast Throughput in Different Deployment Densities](image2)

Figure 5.4: Channel Assignment and Multicast Throughput in Different Deployment Densities

To be able to compare our solution, we also have implemented two other algorithms: consecutive channel assignment and orthogonal channel assignment. In consecutive assignment, the 13 different 802.11 channels are assigned to mesh nodes in a consecutive fashion (from channel 1 to 13, then back to 1), during a BFS traversal. In orthogonal assignment, the greedy approach of selecting a
chapter with maximum separation is adopted. In this approach, even thought we have 13 channels available, only channels 1, 6, and 11 are selected and used based on their complete orthogonality.

Figures 5.3a and 5.3b present the network throughput in a wireless mesh networks including $6 - 80$ nodes in a $1000 \times 1000 \, m^2$ area. This throughput is the maximum throughput in the whole network and is computed based on maximum link capacities in the network (min cut/max flow). In fig.5.3a a random network is generated for a given number of nodes, then the output of all three algorithms is examined on the given single network. Figure 5.3a, on the other hand, compares the average throughput obtained by each algorithm over 10 random networks of a given number of nodes. The variance of the number of nodes is intended for observing the performance of the solutions with different levels of interference.

Figures 5.4a and 5.4b present the network throughput in a wireless mesh networks including 60 nodes in an area of $300 \times 300 \, m^2$ to $1000 \times 1000 \, m^2$. Again, figure 5.4a compares the throughput over a single random network and figure 5.4b presents the average values over 10 random networks. Shaky behaviour of the curve in figure 5.4a is a result of random node placement in the network which may lead to bottleneck links in a random network. Lower values in smaller areas are because of higher interference and more often collisions. An overall observation in figures 5.3a-5.4b is that our solution leads both orthogonal and consecutive channel assignment methods, with a largest margin of up to 100%. The improvement increases as the network size increases. Note that the throughput of orthogonal channel assignment and our proposed algorithm could be improved by scheduling mechanisms.

![Figure 5.5: Maximum Multicast Throughput with Different Number of Nodes](image)

Then, we examine the performance of the primal-dual algorithm. Figures 5.5a and 5.6b present
the maximum multicast throughput using our algorithm as well as consecutive and orthogonal channel assignment algorithms. The multicast routing is assumed to be hop count-based multi-gateway multicast routing used with orthogonal and consecutive channel assignment algorithms. Again, the results are shown for a random network and average over 10 random networks. Results for the single random network is presented to show that for some networks the results of our algorithm might be the same as the orthogonal channel assignment. This is especially true for lower interference levels. Intuitively, when interference is low and not a serious concern, a judiciously designed multi-channel transmission scheme becomes less important. We conclude that our proposed solution is more beneficial when applied in networks with high transmission activities and high interference.

Over figures 5.5a-5.6b we can also discuss the scalability of our algorithm. The performance of the algorithm is affected by the density of the nodes in the network rather than overall number of nodes. Thus, for any large given number of nodes throughput is dependent on the density of the nodes in an area where the interference range of the already included nodes expands. Therefore, the network considered as combination of smaller mesh networks of different densities given in figure 5.6b will result in the same throughput with central or distributed solution implemented.

5.5 Summary

In this chapter, we discussed the challenges of providing high-throughput multicast applications in dense wireless mesh networks, including simultaneous use of wireless resource for relaying and end nodes. We further presented our multicast solution for high-throughput multicast with efficient
resource utilization in multi-channel wireless mesh networks. Two techniques in system design were assumed: introducing multiple mesh gateways for mitigating the gateway bottleneck problem and utilizing multiple wireless channels for combating wireless interference.

Our overall solution framework is a primal-dual schema based on a mathematical programming formulation of the optimal multicast problem. The framework iteratively switches between solving primal sub-problems for channel allocation and routing, and dual variable update, and gradually progresses towards optimal or approximately optimal solutions. We further presented precise models for each primal sub-problem, and discussed solutions for each of them. Simulation results confirmed the proposed solutions, in considerate throughput gains of up to 100% observed over currently used approaches of multi-channel multicast.
Chapter 6

Collaborative Wireless Local Access

The popularity of wireless local area networks for home internet access has led to a dramatic increase in the density of access points, especially in urban areas. While a user is connecting to an access point, other access points may be active within its receiving or sensing range. Unfortunately, many of these access points operate on the same channel, given the limited number of orthogonal channels in 802.11 wireless networks. This results in higher interference, higher collisions, and consequently sub-optimal throughput.

Figure 6.1 shows a real-world example of the received signal strength, channel usage, and rate of available access points for a representative user that we have tracked in a residential building. There are 54 available access points within the receiving range of the user, with 20% of them having a fairly high signal strength level as illustrated in 6.1a. Figure 6.1b further illustrates the high number of active access points on each of the three orthogonal channels. Apparently, there are overlapped channels among a number of access points, known as co-channel access points [53][21], which inevitably suffer from collisions and interference. The interference, on the other hand, depends on the usage of the channels, which further depends on the particular users associated to individual access points. To date, naive fixed association or simple signal strength based association remain dominating, which in general fail to maximize the overall network throughput, not to mention ensuring fairness among the users.

Earlier studies have suggested that a proportional fair access point association ensures fairness with maximized aggregate throughput [76]. Unfortunately, it requires that all the access points be working on orthogonal channels, which is simply not practical for dense networks as we have illustrated. In this chapter, we closely examine the fair access point association with densely deployed wireless networks. We enable sharing and collaboration among individual networks, and present an
optimized access solution. Sharing is enabled by defining explicit usage limits that are different for local and external users of an access point.

Inspired by the advances in wireless firmware virtualization [30] [29] and traffic aggregation [64] [50], which enable a client to connect to more than one access point, we further expand our model to a multi-point association that fully utilizes the available network resources. We present optimized solutions for multiple access points to collaboratively serve wireless users within a set of networks that share the same upstream provider. Unlike previous studies, our solution focuses on maximizing wireless throughput rather than backhaul throughput. The upstream provider can collect the user signal reception data from individual wireless networks, centrally find the optimal association, and provide it to the access points.

To our knowledge, this is the first work that enables the access points to effectively collaborate with explicit sharing bounds, so as to mitigate the impact of interference and poor node deployment in dense wireless networks. Our performance evaluations, based on real data traces for both 2D and 3D scenarios, have demonstrated the superiority of the proposed solution, which generally outperform the dedicated access point association by up to 140% in certain node deployments. We further discuss the practical implications of our solution; specifically, we show that only firmware updates are expected at wireless client, while no change is required at the access points.

The densely deployed access points, on the other hand, offer great opportunities for wireless users to smartly choose the access point to associate with. However, choosing the best association needs availability of a number of access points to the users. This is not true for home wireless networks. In home wireless networks, users can only use the resources of their dedicated access point. Fortunately, most of these networks share the same upstream provider. In our measurements,
four different upstream providers offered residential access, two of them providing the majority of the connections. This is an opportunity to enable collaboration, and central management for optimization of access point association. The provider can control sharing and usage limits of the collaborative networks to consistently provide the bandwidth requirements of local users. This way the users can take advantage of access points available to them in an area covered with multiple dense wireless networks.

We are particularly interested in proportional fair association, which yields a good balance between system throughput and fairness. Li et al. [76] proposed a $2+\epsilon$ approximation of optimal proportional-fair access point association in multi-rate wireless local area networks. Baid et al. [21] extended the formulation to enable cooperation among the access points from different networks.

The proportional fair access point association framework assumes there are no two access points sharing the same channel within the same sensing range. This requirement can hardly be met in dense local wireless networks and ignoring it in collaborative access will lead to a sub-optimal solution. We extend the proportional fair association to model collaboration among the access points with explicitly defined local and external usage limits in individually owned local wireless networks. Inspired by the advances in wireless firmware virtualization to enable a client to connect to more than one access point [30] [29] and measurements confirming the gains of the method [120], we further extend the model to include association decisions to connect a user to multiple access points.

6.1 Motivating Example

Consider three access points $a_1$, $a_2$, and $a_3$ and four wireless clients, $w_1$-$w_4$, where the rate of the wireless connection between each access point and wireless client is as indicated on the dotted lines on figure 6.2. Since they are part of a dense deployment, all three access points are on the same frequency channel in the frequency reuse map of the global network. The association to the predefined dedicated access point results in connections as illustrated in 6.4, i.e., each user connects to its own local access point which is $w_1$ and $w_2$ to $a_1$, $w_3$ to $a_2$, and $w_4$ to $a_3$. The access points serve their users with equal serving time. The access points working on the same channel should also share the channel with equal share times. The dedicated access is also a proportional fair association. In such an association, the throughput for user $w_1$ will be 2. This is because it shares the access point with another user and therefore gets $1/2$ of serving time, and the access point shares channel with 3 co-channel access points and therefore gets $1/3$ of channel air-time. The rest of the users' throughputs will be 8, 3, and 16, respectively, and the overall throughput of the network is 29.
Alternatively, assume each access point allocates up to 30% of its bandwidth to serve other users. As shown in figure 6.4, $a_1$ serves $w_1$ and $w_2$ and $a_3$ serves $w_3$ and $w_4$, the throughput for users is 3, 12, 6, and 16, respectively, resulting in the total throughput 37. This is the association decision of the optimal collaborative access point association.

If we allow multiple access point associations in collaborative association optimization, as illustrated in figure 6.5, $a_1$ serves $w_1$ to $w_3$ and $a_3$ serves $w_4$ and $w_3$. In this case, the throughput for users is 2, 8, 4 + 6, and 16, respectively, resulting in a total throughput of 36. The total throughput is less than the collaborative association in 6.4, however, the throughput for user $w_3$ is noticeably higher with all other throughputs equal to those in the network with no collaboration, illustrated in 6.3. This case is desirable when user $w_3$ is more important for the provider (e.g., pays more and expects higher throughput).

In addition to the increase in the system throughput, disabling the access point $a_2$ results in decreased interference in the network. This will improve the throughput of all users within its sensing range.

### 6.2 Collaborative Wireless Access

There are three different entities within collaborative networks: access points, wireless clients, and upstream providers. Collaborative access does not impose any changes on the access points. Local
users of a wireless network deploy the access points. Therefore, the placement, power setting, or channel assignment of the access points have already been decided (e.g., optimized) during deployment. The deployment parameters can be changed by the users and these changes trigger association decisions at the upstream provider.

Wireless clients connect to the access points based on association decisions provided by the upstream provider. However, while initially joining the network, the wireless clients do not know the association decisions. Therefore, the association process starts by wireless client connecting to the upstream provider by the local access point or the strongest signal access point prior to having the association decision from the upstream provider. This is the initial step in association, and association between the wireless client and access point is not considered completely established. Therefore no data connections are made in this step, before the decision is announced by the upstream provider. Association decisions are made by the upstream provider and announced to the wireless clients through the initial association.

A software update at the wireless client enables the wireless client to decide the connection based on the association variable announced by the upstream provider rather than the strongest signal value. It also provides firmware virtualization for multiple access point connections at the same time.

The upstream provider centrally optimizes the access point associations within the collaborative network. This is performed by clustering, and access point association optimization. Clustering is performed on the access points and users to decide the scope of each centralized optimization. Since the wireless local area networks have a limited coverage area, the provider performs clustering based on location, i.e., access points and users within the same residential building form a cluster. This ensures the scalability of the centralized optimization by limited the optimization variables within the scope of each cluster. Access point association is then centrally performed at the upstream provider for each cluster.

All the access points ensure secure connection to the local and external users. Security settings for each access point is performed by the local users of that access point, and shared with the upstream provider for external user access setting. This is also performed by the wireless client software update, enabling it to configure the local access point for such settings.
Table 6.1: Summary of the Variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>Set of all available access points</td>
<td>$\alpha_{ij}$</td>
<td>Max share % of user $j$ on AP $i$</td>
</tr>
<tr>
<td>$m$</td>
<td>Number of access points $m =</td>
<td>A</td>
<td>$</td>
</tr>
<tr>
<td>$U$</td>
<td>Set of all wireless clients</td>
<td>$d_{ij}$</td>
<td>$1$ if AP $i$ is in receiving range of user $j$</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of clients $n =</td>
<td>U</td>
<td>$</td>
</tr>
<tr>
<td>$C$</td>
<td>Set of available wireless channels</td>
<td>$w_{ij}$</td>
<td>Priority of user $j$ for access point $i$</td>
</tr>
<tr>
<td>$c$</td>
<td>Number of available channels $c =</td>
<td>C</td>
<td>$</td>
</tr>
<tr>
<td>$p_i$</td>
<td>Transmission power of AP $i$</td>
<td>$r_{ij}^k$</td>
<td>Link rate between AP $i$ &amp; user $j$ on channel $k$</td>
</tr>
<tr>
<td>$R_i$</td>
<td>Downlink rate limit of access pint $i$</td>
<td>$t_{ij}^k$</td>
<td>Air-time of AP $i$ for user $j$ on channel $k$</td>
</tr>
<tr>
<td>$L_i$</td>
<td>Downlink usage limit of access point $i$</td>
<td>$b_{ij}^k$</td>
<td>Bandwidth share of user $j$ on AP $i$ on channel $k$</td>
</tr>
<tr>
<td>$n_i$</td>
<td>Number of clients in network $i$</td>
<td>$z_{ij}^k$</td>
<td>Association of user $j$ to AP $i$ on channel $k$</td>
</tr>
</tbody>
</table>

6.3 Access Model and Notations

In this section, we provide the model and notations for the mathematical programming formulation of centralized access point association optimization at the upstream provider. We consider a service provider that manages $m$ wireless local area networks, each of one access point. There are $n$ wireless clients served by the same upstream provider, and the provider’s goal is to maximize the total downlink throughput for all the networks, while ensuring fairness across individual users. Table 6.1 summarizes the notations we use throughout the chapter.

![Figure 6.6: Access Points and Wireless Clients in a Single Provider’s Networks](image)

Existing solutions generally associate a client to its dedicated access point or to the nearest access point, often inferred by signal strength. Such simple solutions inevitably cause strong interferences and unbalanced load in dense networks, and eventually lead to suboptimal network throughput. With our collaborative association, an access point however can provide a share of its resources to serve any user of the provider.

To provide collaborative association, we assume that each access point can allocate a certain portion of its bandwidth $\alpha_{ij}$ to serve other (non-local) users to increase the global provider network throughput. $\alpha_{ij} = 1$ for the local users. A user gets different share on different non-local access points based on each access point’s sharing settings.
In practice, each access point \( i \) has a downlink rate limit of \( R_i \), to provide local wireless network access to \( n_i \) local users. This limit is based on their contract with the provider. Also, each network (access point) \( i \) has a downlink usage limit \( L_i \), also dictated by the provider. To motivate collaborative access, the provider can increase the usage limit by the usage of external users supported on each access point and decrease it by the users’ usage on other networks, so that the total downlink usage limit of a local network does not shrink for its own users.

If two access points are within each others’ sensing range or have a receiving user within each others’ receiving range, they should use orthogonal channels for transmission, or their simultaneous transmissions will interfere. In a densely deployed network, frequency planning can hardly guarantee orthogonal channel assignment on all access points within each other’s sensing range, and careful association is therefore needed. We assume each access point on channel \( k \in C \). The frequency band of the network can accommodate a total of \( c \) channels. We assume multi-rate access points that serve different users with different wireless link rates. \( r_{ij}^k \) indicates the downlink rate for user \( j \) on channel \( k \) of access point \( i \). The \( r_{ij}^k \) value is non-zero over only one channel if the network only uses omni-directional single input, single output access points. It can have multiple non-zero values over different channels for multiple-input multiple-output or directional antenna access points. The rates for different users depend on different parameters like interference, power, and modulation scheme. \( t_{ij}^k \) denotes the fraction of time the user \( j \) is served on access point \( i \) on channel \( k \). \( b_{ij}^k \) is the bandwidth share of user \( j \) on access point \( i \) on channel \( k \) \( (t_{ij}^k r_{ij}^k = b_{ij}^k) \).

The ultimate goal of the service provider is to maximize the total downlink throughput for all the networks, while ensuring fairness across individual users for customer satisfaction. Association variable \( x_{ij}^k \) indicates the state of the association. Access point \( i \) is associated to a wireless client node \( j \), over channel \( k \) on wireless link with capacity \( r_{ij}^k \) if \( x_{ij}^k \) is non-zero. We consider both single and multiple access point associations. For single access point association \( x_{ij}^k \) is a binary variable.

### 6.4 Collaborative Access Point Association

The objective of wireless access network management is to optimize the \( x_{ij} \) associations for maximum network utility while providing fair service to all of its users. It is known that proportional fair association provides a good balance between fairness and throughput maximization as it uses a weighted sum of logs objective function [76]. However, the proportional fair association cannot be used in local area networks where the access points only serve their local users. In this section, we provide a solution to enable sharing access to non-local users and present the optimal collaborative
association with proportional fairness.

6.4.1 Single Access Point Association

We start from the single access point association, where a client can connect to only one access point, and present the access point association optimization at the upstream provider. The baseline proportional fair access on orthogonal channel access points, with no co-channel access points and no sharing at the access points, can be formulated as the following non-linear program [76] [21]:

$$\max \sum_{j \in U} \log (\sum_{i \in A} w_{ij} x_{ij} b_{ij})$$

subject to

$$\sum_{i \in A} x_{ij} = 1 \quad \forall j \in U$$

$$b_{ij} = \frac{r_{ij}}{\sum_{j' \in U} x_{ij'}} \quad \forall i \in A, \forall j \in U$$

$$x_{ij} \in \{0, 1\} \quad \forall j \in U, \forall i \in A$$

The objective is to maximize the total throughput of the network over $x_{ij}$, the association variable. In the proportional fair framework a user can connect to any access point as opposed to local wireless networks. The priority of serving a user $j$ on access point $i$ is defined by weight of the user in throughput maximization $w_{ij}$. The basic proportional fair formulation assumes the channels are all orthogonal and there are no co-channel access points. Therefore, $r_{ij}$ is used for the rate notation instead of $r_{ij}^k$. The first constraint dictates the client connection to only one access point; the second defines the throughput of a connected user based on the number of connections on the serving access point and the connection rate.

The above problem is NP-hard and Li et al. [76] have proposed a $2 + \epsilon$ approximation. It relaxes the problem to a discrete linear program, and the rounding process by Shmoys and Tardos for the generalized assignment problem [114] is then applied to provide binary values of the association variable $x_{ij}$. Note that there is no collaboration among the access points. If sharing is allowed, the above solution can result in even worse throughput than the dedicated access point association, given that the channel orthogonality assumption does not hold, particularly for dense local area wireless networks.

To address this problem, we introduce access point sharing constraints to the formulation to enable collaboration with explicit sharing limits. We also add multi-channel rate specification, and frequency re-use constraints to the proportional fair association formulation to model the dense...
residential areas with multiple access points working on each channel. The resultant optimal collaborative access point association thus becomes:

$$\max \sum_{j \in U} \log \left( \sum_{i \in A} w_{ij} x_{ij}^k b_{ij}^k \right)$$ (6.2)

subject to

$$\sum_{i \in A} x_{ij}^k = 1, \forall j \in U, \forall k \in C$$

$$\sum_{c \in K} x_{ij}^k = 1, \forall j \in U, \forall i \in A$$

$$b_{ij}^k = \frac{d_{ij} \alpha_{ij} \gamma_{ij}^k}{\sum_{j' \in U} x_{ij'}^k - \sum_{i' \in A} s_{ij} \gamma_{i'j}^k x_{i'j'}^k} \cdot \frac{1}{R_{ij}}$$

$$x_{ij}^k \in \{0, 1\}, \forall j \in U, \forall i \in A, \forall k \in C$$

Again, the objective is to maximize the weighted bandwidth shares of all of the users connected to an access point. The weights are computed based on association costs, as discussed in section 6.2. The provider whose goal is to provide high-throughput access to the customers (local network/access point owners) solves the optimization. Given a node placement, throughput is then determined by the association strategy.

The weights of the users in collaborative association are based on their contract with provider for using network resources. We define $\sigma_{ij}$, as the cost of association between user $j$ and access point $i$, to specify this priority. We define $w_{ij} = \frac{1}{1+\sigma_{ij}}$, where $\sigma_{ij}$ is calculated based on the provider pricing. The cost $\sigma_{ij} = 0$ if user $j$ is the local user of the access network $i$. If user $j$ is not a local user for network $i$, the value of $\sigma_{ij}$ depends on its contract. This cost is lower for users with higher usage and bandwidth limits.

With the carrier sense multiple access (CSMA), each wireless node (access point or client) has a sensing range and a receiving range. A wireless client (end user) can sense the channel conditions of all access points within its sensing range and can associate with any access point within its receiving range. If a user is in the sensing range of an access point, the access point interferes with other wireless signals the user receives. Let $d_{ij}$ indicate if user $j$ is in the receiving range of access point $i$, and $s_{ij}$ indicate if user $j$ is in sensing range of user $i$, both of which are binary values and are available to the provider through feedbacks.

Since the $x_{ij}^k$'s are binary values, the first constraint limits the user’s connection to only one access point. The second constraint makes sure the connection of user $j$ to access point $i$ is only over a single channel. We do not need to impose any further limits on the access point channel
selection. This is because each access point decides the transmission channel, and the channels used in the formulation are only for consideration of already configured co-channel access points. The third constraint defines the throughput of user \( j \) connected to access point \( i \) on channel \( k \). The calculation is based on the number of users to share the same access point, and the number of co-channel access points transmitting data on each others’ receiving range. 

\[ d_{ij} \alpha_{ij} r_{ij}^{k} \]

ensures access point \( i \) is in the receiving range of user \( j \) and user \( j \) is within its sharing limits \( \alpha_{ij} \) on access point \( i \). \( \alpha_{ij} \) is one for local users of an access point. The term, \( \frac{1}{\sum_{j' \in U} x_{ij'}^{k}} \), takes care of time sharing among the users connected to the same access point. The term

\[
\frac{1}{\sum_{j' \in A} s_{ij}^{j'} \frac{\sum_{j' \in U} \gamma_{ij}^{j'} x_{ij'}^{k}}{R_{ij'}}}
\]

ensures sharing the channel among the co-channel access points within the sensing range of user \( j \) in the network. This also ensures that the association decision on the channel is based on user’s perspective. i.e. the access points received or sensed by the user \( j \) are not working on the same channel at the same time. \( \frac{\sum_{j' \in U} \gamma_{ij}^{j'} x_{ij'}^{k}}{R_{ij'}} \) is zero if the access point does not work on channel \( k \), where the user \( j \) receives on channel \( k \). This term is equal to one if all the access point’s users are connected to it over channel \( k \). Therefore, if the working channel on access point \( i' \) is channel \( k \) but there is no user connected to it, we will not consider the access point as a co-channel access point in sharing the spectrum. We will further discuss this in section 6.4.3. Modulation \( \gamma_{ij} \) indicates the modulation coefficient for user \( j \) on access point \( i \). The modulation schemes and transmission rates are decided by the MAC layer. For any association combination in the system

\[
\sum_{k \in C} \sum_{j \in U} t_{ij}^{j} r_{ij}^{k} \gamma_{ij} = R_{i}
\]

holds for all access points.

The nonlinear optimization problem (6.2) can be relaxed to a discretized linear program without the integrality constraint on \( x_{ij}^{k} \) to allow multiple associations. We can then apply rounding techniques to obtain the single access point association.

### 6.4.2 Multiple Access Point Association

In practice, an access point may not be willing to provide more than a certain amount of its available bandwidth (indicated by \( \alpha_{i} \)’s in the optimization) to a certain user. If the user demands even higher throughput, we suggest that it can connect to multiple access points simultaneously (see Figure 6.4). Association to multiple access points is possible by time sharing among the connections in the user [30].

This enables the provider to keep the connections for the user, which could use multiple TCP
or TCP-friendly rate controlled connections [32] to be enabled on the multiple access points. The challenge however is to increase its throughput without increasing the interference or contention level in the network.

To this end, we model the multiple access point association using convex optimization, as below.

$$\max \sum_{j \in U} \log \left( \sum_{i \in A} (w_{ij} d_{ij} \sum_{k \in C} t_{ij}^k r_{ij}^k) \right)$$

s. t. 
$$\sum_{i \in A} t_{ij}^k \leq 1 \quad \forall j \in U, \forall k \in C$$
$$\sum_{j \in U} (t_{ij}^k + \sum_{i' \in A} s_{i'j} \sum_{j' \in U} t_{i'j'}^k) \leq 1 \quad \forall i \in A, \forall k \in C$$
$$\sum_{k \in C} t_{ij}^k \leq \alpha_{ij} \quad \forall i \in A, \forall j \in U$$
$$0 \leq t_{ij}^k \leq 1, \forall k \in C \quad \forall j \in U, \forall i \in A$$

The first constraint limits the total allocated time fraction of each user $j$ over a channel $k$ to 1. The second constraint says that the fraction of time a user $j$ is connected (to any access point) cannot be more than 1. It also ensures the time sharing of co-channel access points within the sensing range of user $j$ over channel $k$. The third constraint says that a user $j$'s fraction of time being served on access point $i$ over all channels cannot be more than $\alpha_{ij}$ of the access point’s air-time.

The convex optimization does not determine the number of associations. It may assign a serving time over all access points in the receiving region of a user. Hence, we limit the maximum number of simultaneous access points associations for a user to $c$, the number of channels. This is because of the limitation on the frequency channels available for each wireless access point. If a wireless client is connected to more than $c$ access points at a time, the $(c + 1)th$ access point channel will interfere at least with one of the current associations and therefore the connection will not work at the desired rate. The number is different for directional antennas. We discuss this in further detail in section 6.6.

The provider can solve the convex program in polynomial time. Then, it applies rounding to omit the associations with connection rates lower than $\beta$ times of the link rate to have at most $c$ connections at each access point by setting the air-time share of the rest of the users to zero. Assume $t_{ij}^k$ denotes the transmission time shares after dropping extra connections. We compute the new bandwidth and the final association variable and on each client:
\[ b'_j = \sum_{i \in A} \sum_{k \in C} r^{k}_{ij} f^{k}_{ij} d_{ij} \quad \forall j \in U \quad (6.5) \]

\[ x^{nk}_{ij} = \frac{r^{k}_{ij} f^{k}_{ij}}{b'_j} \quad \forall j \in U, \forall i \in A, \forall k \in C \quad (6.6) \]

We recompute the new transmission times using the new association variables \( x^{nk}_{ij} \) and the ratio of the used bandwidth \( b'_j / r^{k}_{ij} \). Then, we round up the non-zero association variables to 1 and recompute the air-time share. The approximation factor for the rounding algorithm is \( \frac{\beta^2}{\epsilon} \).

The provider can also use the convex optimization (6.4) and the rounding algorithm above for wireless clients connecting only to one access point. It provides a looser approximation factor than that of the discrete approximation for non-linear program (6.2). Due to convexity of the optimization equation, it is fast enough even for large networks. For providers of ultra large networks, we suggest that the network can be partitioned and focus can be put on areas of dense co-channel access points, e.g., high-rise buildings.

As mentioned earlier, the provider sets the \( R_i, L_i, \) and \( \sigma_{ij} \) values for the access points, based on service contracts. The provider also announces the optimal association \( x^{nk}_{ij} \)'s to the access points, which further distribute them to the users for their actions.

### 6.4.3 Channel Assignment

We use multi-channel formulation in the collaborative access point association formulation. However, the channel in the optimization formulation is used for the one or multiple channels that the MAC layer has already selected for transmission and have provided the rates over it. In other words, the channels selected for transmission from the MAC layer are used to formulate the multi-channel collaborative access point association. Therefore, the \( r^{k}_{ij} \) values are provided on the channels that the access point has already selected and decided the modulation scheme for, and the \( r^{k}_{ij} \) values over the rest of the channels are zero. We assume that the conventional channel selection scheme is used in the MAC layer, where each access point selects the least contended orthogonal channel in the channel re-use map of the network by sensing the other access points in the vicinity. If the access points are single input single output, as in 802.11a/b/g, they can only have non-zero \( r^{k}_{ij} \) values for one \( k \in C \). If the access points are multiple input multiple output, as in 802.11n, they can have non-zero \( r^{k}_{ij} \) values over \( c \) different channels. The term

\[ \frac{\sum_{j' \in U} \gamma_{ij} r^{k}_{ij} x^{nk}_{ij'}}{R_{ij'}} \quad (6.7) \]
only takes 0 or 1 values for single input single output access points. It can have values between 0 and 1 for multiple input multiple output access points that sum up to 1 over all channels for a single user. However, connection to a user over a single channel is the common choice based on the limited number of orthogonal channels.

The access points also set the values of $d_{ij}$, $s_{ij}$, and $\gamma_{ij}$ after sensing the access points in their vicinity and report these values along with the user feedback on $r_{ij}^k$ values from the user to the provider for the optimization.

The provider can compute the optimal values of the non-local $\alpha_{ij}$ and recommend these values to the wireless clients to configure the access points locally according to these values. In this chapter, we do not consider details of setting $\alpha_{ij}$ values and consider them as predefined values.

### 6.5 Performance Evaluation

#### 6.5.1 Data Collections

To accurately emulate real-world WiFi deployment, we have collected residential WiFi data in more than 100 locations in different residential buildings and high-rises in Vancouver, Canada. We used open source Wi-Fi scanning software, including InSSIDer [62], and WiFiExplorer for scanning the wireless access points and their signal strength in these locations. For each location, we have collected MAC address, SSID, BSSID, modem vendor, network type, operation mode, RSSI value, maximum, average, and minimum signal values and percentages, security type, encryption, band,
channel, signal to noise ratio (SNR), minimum, maximum and average noise values and percentages, maximum rate, and stability in RSSI values in five-minute time intervals at different times of a day. Our collected data suggested that in many residential buildings there are more than a handful of access points that share the same provider, and have strong enough signal to be considered for association for a single wireless user. We also observed that home sharing devices (e.g., gaming consoles, phones with WiFi tethering), although working in the WiFi frequency, usually use lower transmission powers and are usually not discoverable by the devices more than a few meters away. Therefore, the devices using WiFi for home sharing can continue working within the provided scheme.

We also collected application and traffic measurements in a few residential buildings. We observed that downlink transmissions from the access points account for the majority of the transmission in residential wireless networks. We also found out that the wireless capacity, reduced by high contention and suboptimal associations, is the bottleneck in providing a high throughput internet downlink for applications like streaming in high density areas. This could be solved by controlled sharing and collaborative access.

We used the collected data in setting up the evaluation scenarios in our numerical analysis and simulations, and the dataset can be found at:

http://www.sfu.ca/~oba2/dls.
6.5.2 System Configuration

We used CVX [42], a package for specifying and solving convex programs [42] [54], to solve the convex problem in multiple access point connection and lpsolve to solve the discrete LP problem.

We set all the access points to work on the 2.4 GHz band on 11 channels of 802.11b/g/n. The default transmission power level for each node in the network is 10dBm, with a sensing range of 125m and a receiving range of 35m for indoor access points. The interference coefficient for completely overlapping channels is one, and the coefficients for one, two, three, four, and five channel separations are 0.7906, 0.5976, 0.2651, 0.00627, and 0.0012, respectively [11]. There is no interference between orthogonal channels.

Our proposed system solves a problem which is more evident for complex systems with a large number of access points at each location. We simulate this environment using our measurements from residential networks. We deployed the access points and wireless clients with clustered distributions and in 2D and 3D scenarios, respectively modelling flat residential areas and residential high-rises. In the 2D scenario, we divided the given area into smaller, non-overlapping square-shaped areas and randomly placed an access point within the borders of each small area. We then placed a random number of users (up to $n_{max}$, the maximum number of local users for each access point) for each access point within the area to resemble a residential wireless local area network.

For 3D configurations, we set the height of each floor to 3m, and the above 2D placement is then applied for each floor.

![Figure 6.11: 3D: Wireless Client Rates](image1)

![Figure 6.12: 3D: CDF of Client Rates](image2)
6.5.3 Evaluation Results and Analysis

For comparison, we also implemented dedicated access point association (default access in residential wireless networks), and the basic proportional fair access point association [76]. We enabled a share portion of $\alpha$ on each access point, for the proportional fair association as well. Therefore, the bandwidth sharing available for proportional fair will be the same as for the collaborative access. However, proportional fair does not consider co-channel access points in making association decisions.

We sorted the user rates in the figures based on the client rates in the dedicated access. The collaborative access rates and basic proportional fair association access rates of the users are shown with the same order to illustrate the gains of each client using different associations.

Figure 6.9 illustrates the user rates in a sample configuration of a wireless networks in a 150$m \times 150$m area. The access point and the user placement are depicted in Figure 6.7, where all the access points are one meters above the ground. $\alpha_{ij}$ is 30% for all of the foreign users on an access point $i$ and 100% for all of its local users. Each access point has at most 6 local users. The cost of association is considered to be $\sigma_{ij} = 0$ for local users, and 0.5 for other users. In this network, we assumed $R_i$, the downlink rate limit from the provider, is equal to the maximum achievable wireless rate at the access point. This assumption implies that the bottleneck is the wireless link capacity.

Figure 6.11 shows the user rates in a 3D scenario, which consists of 10 floors, each of 25$m \times 25$m. The access point and user placement is depicted in Figure 6.8. We have deployed 4 access points in each floor and each access point has at most 6 users connected. The sharing parameter is $\alpha_{ij} = 30\%$ for non-local users. Because of the high number of nodes in this configuration, we have used the convex optimization for this scenario. The number of multiple connections allowed is 3, but most of the users connect to only one access point. We can see that the higher the number of access points and users per access point, the better the collaborative access performs in comparison to dedicated access point users.

The average wireless client rate in this scenario is improved by about 140%. This increase is due to three main reasons: (a) The collaborative association uses fewer access points, and therefore fewer collisions occur for co-channel access points, resulting in less interference for non-orthogonal channel users; (b) Majority of the users are connected to the access points providing higher rates; (c) The load is better balanced across access points, instead of constrained to each access point serving its local users.

We also find that the highest increase is for low-rate users. We believe that this is mostly due to
the elimination of co-channel access points. In our experiments, collaborative association has eliminated up to 6 access points in the 25 access point scenario in different node deployments to enhance throughput. Note that some high-rate users in dedicated access point association may experience lower rates in collaborative access, because of the sharing of the access point capacity to provide access for other users. This however does not happen frequently, and the users still have acceptable rates, e.g., for the network in Figure 6.7, only two users have experienced decreased rates.

Another interesting observation is that the basic proportional fair association works very similar (even worse for some users) to the dedicated access in this scenario. This is because of neglecting the co-channel access points and using all of the available access points for load balancing among them, which indeed increases the number of co-channel access points.

Figure 6.13 illustrates the impact of the number of deployed access points on the collaborative access algorithm. The access points are deployed in a $150m \times 150m$ area. As illustrated in the figure, our solution works better when more access points are deployed. This is because there are more access points with different rates and connection qualities to choose from in creating the associations. Furthermore, our solution adaptively associates users to the access points and each access point may not necessarily have associated users. Therefore, the rates will not drop with excessive co-channel access points.

On the other hand, the dedicated access point, and the basic proportional fair associations experience a mild increase with increasing the number of access points to 20. However, any further increase leads to a decrease of the throughput, showing the impact interference among the co-channel access points.

Figure 6.14 illustrates the impact of the number of users for 15 access points in a $150 \times 150m$ area. If we have one or two users for each access points, the gains of collaborative access over the
dedicated access point association are negligible, unless the $\alpha$ for foreign nodes is high. Figure 6.14 illustrates it for $\alpha_{ij} = 30\%$ for non-local users. The gain is the most for 6-12 users per access point, which makes it ideal for residential wireless access control.

![Figure 6.15: Impact of Association Costs](image1)

The cost of association is also an important factor. For our collaborative association, it defines the priority of users in association decisions. Figure 6.16 shows the impact of different association costs on the user rates. If the cost is very low for non-local users, it may sacrifice the throughput of local users with high-rates for providing access to non-local users as illustrated by $\sigma_{ij} = 0.5$ in figure 6.15. We can see the decrease in the throughput of high-rate users in the $\sigma_{ij} = 0.5$ and completely preserving them in $\sigma_{ij} = 2$. We defined the weight as $\frac{1}{1+\sigma_{ij}}$. In this definition $\sigma_{ij} = 2$ gives us the best performance when the share of non-local users on an access point is 30%.

### 6.6 Summary and Discussion

With wireless local area networks being the main end-point access method in homes and offices, we are expecting higher densities in access point deployments in the next couple of years. The high deployment density, given there is no collaboration among these networks, increases interference and co-channel access points. This results in reduced throughput in a dense area.

In this chapter, we studied the challenges in achieving high-throughput wireless access in dense wireless networks. We addressed effective access point association to maximize user throughput. Our overall solution is a centralized optimization based on a proportional fair access point association mathematical formulation. It is solved at the common provider of the local networks, to maximize the provider network throughput as well as the experience for individual end users. Our
simulation results confirm the proposed solution, in considerable throughput gains that were observed over predefined access point association.

There are some practical challenges that should be considered in implementation of the proposed access point association service:

**Access point settings for sharing:** To provide access for users on non-local access points, the access points can configure a guest service set identifier (SSID) for their non-local users. Configuring guest networks on the access points is straightforward and an access point can have multiple guest networks. The provider can choose to have a single or multiple guest networks, e.g., for different clusters covering different residential buildings, to control the usage in different regions of its network. This way, each access point can have a network for its local users, that has the rate limit \( r_{ij}^k \) on the channel the MAC layer selects. It also will have one or multiple guest networks with total of \( \alpha_{ij} r_{ij}^k \) rate limit for the non-local user \( j \). These settings are performed by the owner of each access point.

**Access point feedback:** Most of the parameters needed for the collaborative association optimization are available at each access point after scanning the users and the access points in their sensing regions. A software update on the access points enables them to send the required parameters to the provider and get the updated \( x_{ij} \) and provide it to the wireless clients.

**Changes in the wireless client equipments:** The proposed optimal collaborative association needs minor changes to the wireless client equipments. To connect to the preferred access point, we need a software update on the client devices to choose the access point based on the \( x_{ij} \) values it collects instead of connecting to the access point with the strongest signal. Also, this integrates well with the service access layer, i.e., using Serval [95] on the wireless client, adds a service access layer and enables the user to seamlessly move from one wireless network to another based on the \( x_{ij} \) feedback without service interruptions.

For connection of a user to multiple access points, the client device needs multiple wireless interfaces or a software update for allowing the wireless network card virtualization for multiple connections [30].

**Omni-directional versus directional antennas:** Omni-directional antennas are the most common type of access points used for home and office wireless access. If the access points use omni-directional antennas, transmission for a single user is sent on the wireless medium and all of the users in the sensing range of the access point can sense the wireless transmission signal. All of the access points within the sensing range of the sending access point can also send the signal and will not send anything during its transmission time.
Using directional antennas at the access points increases the number of the connections an access point can support by spatial reuse and decreases the interference level in the network. Our optimization formulation can be used without any alterations for access points with directional antennas as well. Using directional antennas limits the collision possibility by directing the wireless signal to the user and avoiding the transmission on other directions. Therefore, multiple associations for a user is most practical in a network with access points using directional antennas. Assume $D_{\text{max}}$ is the maximum number of directional antennas on each access point and wireless client. The maximum number of associations for each wireless client will be $\left\lfloor \frac{f_{\text{band}}}{f_{\text{channelwidth}}} \times D_{\text{max}} \right\rfloor$ which is $D_{\text{max}}$ times higher than the omni-directional antennas.
Chapter 7

Conclusion and Outlook

We conclude this dissertation by summarizing its contributions. We follow with a discussion of how the pieces of our contributions relate to each other, and make a big picture of how it relates to the advances in wireless network technologies. We then provide insight into the future of advanced wireless networks, infrastructure, and applications.

7.1 Contributions

We discussed three different advanced wireless network technologies: LTE-based cellular wireless networks, wireless mesh networks, and collaborative wireless local area networks.

For LTE-advanced networks, we first proposed a simple solution, yet effective and practical, to enhance user experience and power consumption in both user devices and eNodeB. Our balanced solution addresses the trade-off by including user preference. Our simulation results indicated 5% to 18% improvement in base station power consumption and 13% to 25% improvement in user device power conservation chances. The provided solution also decreases the transmitted data in the network while preserving the user perceived quality of the video.

We further explored resource utilization of LTE networks for high mobility users. We proposed a Cell Array design and a predictive handover mechanism that decreased handover latency by up to 18%. The reduced latency in this design, together with our scheduling algorithm, ensure seamless delivery of multimedia services with reducing the buffered packets for handover time and improve user throughput by up to 35%.

Then we discussed offloading cellular traffic to other types of networks. We first focused on wireless mesh networks and presented our high-throughput multicast with cooperative gateways in
multi-channel wireless mesh networks. Two techniques in system design were assumed: introducing multiple mesh gateways for mitigating the gateway bottleneck problem and utilizing multiple wireless channels to combat the wireless interference.

Our overall solution framework was a primal-dual schema based on a mathematical programming formulation of the optimal multicast problem. The framework iteratively switches between solving primal sub-problems for channel allocation and routing, and dual variable update, and gradually progresses towards optimal or approximately optimal solutions. We further presented precise models for each primal sub-problem, and discussed solutions for each of them. Simulation results confirmed the proposed solutions, in considerable throughput gains of up to 100% that were observed over currently used approaches of multi-channel multicast, i.e., hop-count routing, and orthogonal channel assignment.

Then we focused on wireless local area networks as the best cellular offloading option. For 802.11 based wireless local area networks, we designed a collaborative network with centralized access point association. We modelled co-channel access points, sharing, and collaborative access in network utility maximization framework and provided optimal solutions in this system. We presented measurement data from residential buildings to verify our design choices and show that our solution can improve the overall throughput by up to 140%.

7.2 The Big Picture and Future Outlook

We studied the challenges in achieving high-throughput wireless access in advanced cellular wireless networks. We addressed effective resource utilization to maximize user throughput in these networks. We learned that with increase in the number of users and the volume of mobile data demands, cellular wireless networks are facing the scarce spectrum problem.

On the other hand, with wireless local area networks being the main end-point access method in homes and offices, and their usage in wireless meshes, we are expecting higher densities in access point deployments in the next couple of years. The high deployment density, given there is no collaboration among these networks, increases interference and co-channel access points. This results in reduced throughput in a dense area.

We addressed the problems separately with efficient resource usage within each type of network. However, the problem is not yet fully solved from a user’s perspective. While each network might be achieving better efficiency, we are neglecting the cooperation and offloading among these networks that can drastically improve a user’s experience.
Availability of different wireless interfaces on user devices offers opportunities for offloading high bandwidth applications like video streaming, video upload, and multimedia multicast services to more abundant wireless networks. Therefore, effective resource usage can be achieved by offloading some services from more expensive spectrum cellular networks to more abundant local area and mesh networks. We believe an efficient social local area network design is the best way to increase offloading access from sporadic availability of these networks, changing the offloading from opportunistic in nature towards a ubiquitous wireless access. We are extending our collaborative wireless local access solution towards a social WiFi design as a step towards providing such stable offloading resources.
Bibliography


[38] Cisco visual networking index: Forecast and methodology, 2011-2016.


