Channel Time Allocations and Handoff Management for Fair Throughput in Wireless Mesh Networks

CHANNEL TIME ALLOCATIONS AND HANDOFF MANAGEMENT FOR FAIR THROUGHPUT IN WIRELESS MESH NETWORKS

BY

LEI QIN, B.Eng., (Electronic Engineering)

Beijing Institute of Technology, Beijing, China

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

Master of Applied Science (2013) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE: Channel Time Allocations and Handoff Management for
Fair Throughput in Wireless Mesh Networks

AUTHOR: Lei Qin

B.Eng., (Electronic Engineering)

Beijing Institute of Technology, Beijing, China

SUPERVISOR: Dr. Dongmei Zhao

NUMBER OF PAGES: xvii, 108

This thesis is dedicated to my beloved parents, husband and friends.

Abstract

Wireless mesh networks (WMNs) have been widely deployed for various applications. In contrast to a pure ad hoc network, a WMN having an infrastructure formed by fixed access points (APs) can provide relatively reliable transmissions for forwarding data to and from mobile stations (MSs). The fixed infrastructure also provides chances for effectively supporting various services, better utilizing the radio resources, and improving other network performance. On the other hand, in a WMN with a large number of APs and MSs, maintaining the service fairness among the MSs located in different geographical areas and having different mobility patterns can be difficult. In this thesis, we study how to provide fair throughput for MSs in a WMN by joint channel time allocations and handoff management, where the channel time is allocated at two levels, one among the APs, and another among the MSs associated to the same AP.

In the first part of the thesis, we assume that the channel time allocations at the AP level are given, and jointly consider the handoff management of the MSs and the channel time allocations at the MS level. An optimization problem is formulated based on long-term proportional fairness (PF), so that all the MSs in the entire WMN can receive fair average throughput, while the total throughput of the MSs is maximized. The optimum solution is based on the assumption of having global information

about the channel conditions and mobility information of all the MSs, and cannot be easily implemented in a practical system. A heuristic scheme is then proposed, which allows each AP to allocate its channel time based on local information only, and the handoff decisions of the MSs are based on information exchanged between neighboring APs. Numerical results show that performance of the proposed heuristic scheme is very close to the optimum in terms of both fairness and throughput.

In the second part of thesis, we consider that different APs can share the same frequency channel and should coordinate for their transmission time, so that neighboring APs sharing the same frequency channel do not transmit at the same time in order to avoid strong co-channel interference. We jointly study the channel time allocations at the AP level and the MS level together with the MS handoff management. An optimization problem is first formulated and solved, and the objective is to provide fair average throughput for the MSs while efficiently utilizing the channel resources. Two distributed schemes are proposed by decoupling the handoff management and time allocations. The HO-CA scheme performs heuristic handoff decisions for the MSs based on their link gains to nearby APs, and then optimizes the channel time allocations. The CA-HO scheme allocates the channel time to individual APs based on the neighboring relationship of the APs, and then allows the MSs to make handoff decisions based on possible utilities from nearby APs. In both schemes, individual APs solve a local optimization problem to allocate channel time for their associated MSs. Numerical results indicate that both the distributed schemes can achieve close-to-optimum fairness; and with higher computational complexity, the HO-CA scheme achieves higher throughput.

Acknowledgements

To begin with, I would like to express my sincere gratitude towards my supervisor, Dr. Dongmei Zhao, for her continuous support and insightful advice throughout my entire research work and thesis writing. I truly appreciate for her expertise in identifying the problems and formulating the solutions in wireless networks, patience in keeping me on track of my work, and attention in providing me with timely feedbacks. Without her guidance and encouragement, I would not have come even close to what I have achieved today.

In the meanwhile, I would like to thank Dr. Terence D. Todd and Dr. Rong Zheng for being on my thesis committee, for their contribution of precious time in providing suggestions, comments and questions that considerably helped clarifying issues and improving the quality of the thesis.

My special thanks go to Abdulelah Alganas, Naby Nikookaran, Hadi Meshgi, and Mohammad Sheikh Zefreh in their invaluable suggestion and help over my research. The informal support and encouragement of my colleagues in Wireless Network Laboratory and many friends has been indispensable, and I would like particularly to acknowledge the contribution of Maryam Mohseni, Morteza Azimifar, Amir Khezrian, Shokouh Mostofi, Abdulla Hammad, Hanan Hassanein, and my dear friends Thu Zar Kyaw and Xia Chen.

I would also like to thank my husband, Quan Zhou, for his assistance in improving my code efficiency, and unconditional support in my life and study. Finally, I would like to show my gratitude to my parents for their consistent support and love.

Notations

$\mathcal{M}$ Set of all MSs

$\mathcal{I}$ Set of all APs

$\mathcal{M}_i(t)$ Set of MSs associated to AP i in frame t

$A_{im}(t)$ binary variable, equals 1 if MS m is associated to AP i in frame t and 0 otherwise

T_i Total amount of time that AP i can transmit in one frame

T_{frame} Duration of one frame

P_{AP} Transmission power of an AP

$T_{im}(t)$ Amount of time that AP i transmits to MS m in frame t

$R_{im}(t)$ Transmission rate from AP i to MS m

$\eta_m(t)$ Throughput (total amount of received data) of MS m in frame t

$\overline{\eta}_m$ Long-term average throughput of link m

$\overline{X}_m(t)$ Weighted average rate for link m up to time t

γ Weight used for calculating $\overline{X}_m(t)$

$g_{im}(t)$ Link gain between MS m and AP i at in frame t

$|\mathcal{M}_i(t)|$ Total number of MSs associated to AP i

$\overline{U}_{AP,i}(t)$ Average utility of MSs associated to AP i in frame t

$\mathcal{N}_i$ Set of neighboring APs of AP i

$|\mathcal{N}_i|$ Number of neighboring APs of AP i

$\mathcal{G}_n$	Set of APs in nth GNB
K_i	Total number of GNBs that AP i belongs to
$\mathcal{F}_i$	set of FAPs replacing AP i
Z_j	Channel time for FAP j
$h_j(Z_j)$	Total utility of FAP j
$\mathcal{J}^{net}$	Set of all net-FAPs
$\mathcal{J}^{loc}$	Set of all local FAPs
p_n	Total number of net-FAPs in nth GNB
M^{loc}	Total number of net-APs
$\mathbf{E}_n$	A matrix mapping net variables to the public variables in the nth GNB

Abbreviations

APA   Adaptive Power Allocation

APs   Access Points

BE   Best Effort

BSs   Base Stations

CAC   Call Admission Control

CDMA   Code Division Multiple Access

CW   Contention Window

DARPA   Defense Advanced Research Projects Agency

DCF   Distributed Coordination Function

DOD   Department of Defense

ESS   Extended Service Set

GloMo   Global Mobile Information Systems

HLR   Home Location Register

IETF	Internet Engineering Task Force
IGWs	Internet Gateways
ISPs	Internet Service Providers
MAC	Medium Access Control
MANETs	Mobile Ad-hoc Networks
MANs	Metropolitan Area Networks
Max-Min	Maximizing the Minimum
MM	Mobility Management
MRs	Mesh Routers
MT	Mobile Terminal
NICs	Network Interface Cards
OFDM	Orthogonal Frequency Division Multiplexing
PANs	Personal Area Networks
PF	Proportional Fairness
QoS	Quality of Service
RRM	Radio Resource Management
SIR	Signal-to-Interference Ratio
UWB	Ultra Wide Band

VLR	Visitor Location Register
Wi-Fi	Wireless-Fidelity
WiMAX	Worldwide Inter-operability for Microwave Access
WMNs	Wireless Mesh Netowrks
WSNs	Wireless Sensor Networks

Contents

List of Figures

Chapter 1

Introduction

This chapter gives an introduction to WMNs, some research issues, and related works. The motivation and overview of the thesis are given at the end.

1.1 WMNs - An Evolution from MANETs

Mobile ad-hoc networking is one of the most active technologies in communication networks in past several decades. It has been widely applied initially in military and subsequently in civilian services. Since 1990s, Internet and wireless communication technologies have given rise to tremendous revolution in people's lives, and more attentions have been drawn to design low-cost and high-speed wireless environment that allows Internet access anytime and anywhere. WMNs were later proposed as a solution to address the limitations of conventional mobile ad-hoc networks (MANETs). In this section, we give an overview of traditional MANETs and WMNs, and compare their architectures as well as some characteristics.

1.1.1   Overview of MANETs

A traditional MANET is composed of wireless mobile devices and established spontaneously without the support of existing infrastructure. In other words, a MANET has no centralized entity like cellular base stations (BSs) to manage and control wireless communications. The main characteristics of MANETs are:

- Self-configuring, self-forming and self-healing. Without pre-configuration, the mobile nodes in MANETs are free to move randomly and organize themselves arbitrarily, which cuts down the cost to set up the infrastructure.

- Dynamic network topology. The mobile nodes can join and leave a MANET without notifying each other in advance, hence the network topology can change in an unpredictable manner.

- MANETs are multi-hop networks, where nodes also act as routers and forward messages for their peers. This extends the network coverage of single-hop networks such as wireless local area networks (WLANs), and allows a large number of nodes to communicate with each other within the same MANET.

- Constrained resources. MANETs often suffer from energy shortage due to the limited battery power supply of mobile devices. In addition, MANETs usually operate under a shared wireless channel, which results in limited bandwidth for MSs to compete with each other.

The initial driving force for the development of MANETs was military applications launched by Department of Defense (DOD) of United States in order to provide quick deployable and fast healing communication systems. Defense Advanced Research Projects Agency (DARPA) started the research of Packet Radio Networks

(PRnets) (Jubin and Tornow, 1987) in 1973 to enable communications among MSs by multi-hop wireless links, and launched the Global Mobile Information Systems (GloMo) (Leiner *et al.*, 1996) program in 1994, aiming to develop new wireless ad-hoc networking technology compatible with the Internet technology. Since then, MANETs received massive attentions from both academia and industry. The MANET Working Group was established by Internet Engineering Task Force (IETF) in 1997 to design new routing protocols, since applying the traditional routing algorithms in MANETs will generate many routing updates due to the highly dynamic network topology, which makes the networks unstable.

1.1.2   Evolution from MANETs to WMNs

Working in a self-configuring, self-forming and self-healing manner, the performance of MANETs is significantly affected by many factors, such as link breakage due to frequent topology variations, which, together with channel contentions, further make quality of service (QoS) provisioning in MANETs a challenge. In late 1990s, the prosperity of the Internet and the popularity of personal wireless devices strengthened people's desire to design a low-cost and stable wireless environment to access Internet whenever and wherever they want. WMNs were therefore proposed to satisfy civilian application requirements, whereas MANETs in early stage benefit more for the military networks rather than civilian networks.

Unlike MANETs, WMNs are infrastructure based. A WMN consists of Internet gateways (IGWs) with wired connections to the Internet, mesh routers (MRs) forming the backbone and connected to IGWs mostly through wireless links, and mesh clients. WMNs not only keep some advantages of MANETs like the ability of self-configuring

and self-healing, but also provide high-speed wireless communications and better network coverage to support commercial uses. Some major improvements of WMNs over MANETs include:

- In contrast to the infrastructure-less nature of traditional MANETs, the IGWs of WMNs allow MSs to access Internet services such as email, web browser, interactive multimedia, etc. As MSs are moving under the coverage of a WMN, they can handoff from one MR to another while maintaining their services.

- Contrary with the highly-dynamic network topology of MANETs, the topology of WMNs is relatively stable because MRs in WMNs are mostly fixed. This reduces the probability of link breakage, and consequently makes it easier to manage network resources and improves network performance.

- With MRs that are equipped with different air interfaces, mesh clients with different radio technologies can coexist in the same WMN and communicate with each other through the MRs.

- Compared to battery-supplied devices in MANETs, MRs in WMNs have a better chance to access more reliable power supply, which allows extended network coverage area and longer network lifetime.

- MSs in conventional MANETs mostly work at a shared channel, which limits the network capacity, while the network capacity in WMNs can be magnificently extended by employing multi-radio and multi-channel techniques.

All those benefits make WMNs a cost-efficient solution for civilian applications and widely deployed in office buildings, campus and communities. Because of their

popularity in commercial market, researchers have been redesigning the protocols of some existing networks from the perspective of WMNs, especially of IEEE 802.11 networks, ad hoc networks, and wireless sensor networks (WSNs). In addition, several sub-groups have been established in IEEE standard groups to define specifications for WMN techniques. For instance, IEEE 802.15.5 (Group, 2002b)-(Kinney, 2003) were established to meet the demands of wireless personal area networks (WPANs) in 2003. In the same year, IEEE 802.11s (Group, 2002a)-(Jou and Eastlake, 2004) were specified to accommodate the WLAN devices. Besides, both IEEE 802.16a (Group, 2001a) (Piggin *et al.*, 2003) (Whitehead *et al.*, 2003) and IEEE 802.20 (Group, 2001b) working groups for mobile broadband wireless access networks included WMNs techniques into their standards in 2002.

1.2   Features and Applications of WMNs

The unique architecture and features of WMNs not only resolved some limitations of traditional MANETs, but also made it possible to integrate WMNs with existing wireless network technologies, such as cellular, wireless-fidelity (Wi-Fi) (Alliance, 1999), worldwide inter-operability for microwave access (WiMAX) (Forum, 2001), and WiMedia (Alliance, 2005) networks. In the following subsections, the network architecture and benefits of WMNs will be introduced along with some examples of testbeds and commercial networks.

1.2.1   Network Architecture and Benefits

There are mainly two types of nodes comprising a WMN, MRs and mesh clients. Compared to MANETs, the routing capability of the MRs allows WMNs to achieve the same coverage with much lower transmission power through multi-hop communications. Mesh clients communicate with each other via MRs, and in some systems mesh clients can also function as MRs. The architecture of WMNs can be classified into three groups according to nodes' functionality (Akyildiz *et al.*, 2005):

- Infrastructure/Backbone WMNs. As shown in Fig. 1.1, this WMN architecture consists of mesh clients and MRs. The MRs form a backbone that is connected via IGWs to the Internet. Under this architecture, MRs are a group of special nodes performing routing and resource allocations for mesh clients, and mesh clients only function as traffic sources and destinations. The backbone can be built with various radio technologies such as IEEE 802.11. This enables the integration of WMNs with existing wireless networks as well as supports most updated network technologies.

- Infrastructure-less client WMNs. Client meshing provides peer-to-peer networks among mesh clients. Here the mesh clients can operate as routers as well as end-user devices. Therefore, special MRs are no longer needed in this type of architecture. This type of WMNs is not often clearly distinguished from MANETs in the literature.

- Hybrid WMNs. This is a combination of infrastructure based and infrastructure-less client meshing architecture. While MRs provide connectivity to other types of networks like Internet, mesh clients can forward data inside a WMN, and this

considerably improves the coverage and connectivity within the WMN.

Wireless Mesh Network
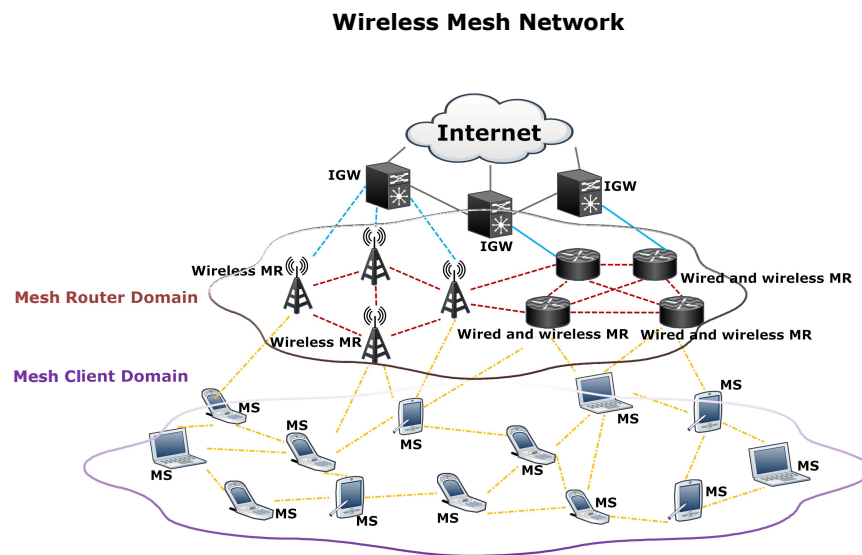
Figure 1.1: Infrastructure/Backbone WMNs

With flexible architecture, WMNs offer many benefits apart from those mentioned in subsection 1.1.2:

- High-speed data transmissions. While IEEE 802.11g (Group, 2002a) works in the 2.4-GHz band with a maximum raw data rate of 54 Mbps, the data rate of IEEE 802.11n can reach between 100 and 200 Mbps. This makes it possible to support more real-time applications and bandwidth-consuming communications.

- Extended coverage and connectivity. Multi-hop working mode and high transmission power of MRs can significantly extend the coverage and connectivity of

WMNs. Moreover, fast and seamless handoffs can be realized for a mesh client to hand over from one MR to another (Amir *et al.*, 2006).

- Flexible and cost-efficient deployment. MRs can be controlled and managed by a single or several Internet Service Providers (ISPs), or installed/removed anytime and anywhere due to their self-configuration and self-forming features. In addition, the wireless connection of MRs cuts down the expense on wired links, and in turn reduces the deployment cost of WMNs.

1.2.2   Testbeds and Commercial WMNs

At the early stage, research works on WMNs (Mahmud *et al.*, 2006) were mostly implemented based on open source share, such as IEEE 802.11b/g based Roofnets (Bicket *et al.*, 2005), which was developed by MIT to provide broadband Internet access to Cambridge, MA. The major feature of Roofnet is that it does not rely on a prior configuration. BWN-Mesh (BWN-Mesh, 2005) developed by Broadband and Wireless Network (BWN) Lab at Georgia Institute of Technology is composed of 15 IEEE 802.11b/g MRs. Some of them are connected to other wireless networks such as WSNs via IGWs, others provide access to mesh clients within WMNs. A 4-node multi-channel IEEE 802.11b testbed called Net-X (Net-X, 2007) was proposed by University of Illinois at Urbana-Champaign, where each node is equipped with two wireless interfaces, and the network channels were determined based on a load-aware channel assignment algorithm. These testbeds provided chances for a wide range of institutes and groups to further design architectures, protocols, and algorithms for commercial applications of WMNs. A lot of industrial companies (Churchill, 2007) have put their interests in providing WMNs services.

Cisco has developed its Aironet products (Cisco, 2007) to allow private places, government, transportation systems, industry, health medical systems and security surveillance systems to deploy their own low-cost WMNs. Tropos Networks employed a cellular Wi-Fi network architecture to support WMNs (Tropos, 2006) with an operating system called Tropos Sphere, which runs on standard IEEE 802.11 hardware and software. Kiyon provided solutions for building automation and "small office home office (SOHO)" customers with its broadband wireless mesh routers (Kiyon, 2006) employing hybrid CSMA/CA and distributed TDMA MAC protocol atop an IEEE 802.11g/a physical layer.

WMNs have demonstrated to be a promising solution for most civilian applications with low cost. However, there are still a great many challenging issues pending to be solved before the advantages of WMNs can fully take effect. In Section 1.3, we will discuss some open issues in WMNs, and specifically focus on the MAC layer issues, critical problems of radio resource management (RRM) and mobility management (MM) in WMNs.

1.3   Some Major Research Issues in WMNs

Although WMNs have emerged as a promising technology for broadband wireless access in a cost-efficient fashion, there are still a number of open issues to be resolved, which can be key factors affecting the network performance. Some of these issues are listed below:

- RF techniques. A great many RF solutions have emerged to increase the capacity and flexibility of wireless systems, including directional and smart

antennas (Ramanathan, 2001)-(Spyropoulos and Raghavendra, 2003), MIMO systems (Xiang *et al.*, 2004)-(Sundaresan *et al.*, 2004), and multi-radio/multi-channel systems (So and Vaidya, 2004)-(Adya *et al.*, 2004). All those advanced technologies are expected to be applied on future platforms, which further require to reinvent the upper layer protocols, especially MAC and routing protocols. Hence, MAC protocols in legacy IEEE 802.11 are no longer applicable.

- Scalability. Multi-hop communications are common in WMNs. As the number of hops increases, throughput decreases accordingly in the network, which severely impacts the scalability of WMNs in terms of both geographical coverage and total number of nodes in the network. Scalability in WMNs is very much dependent on the design of MAC and routing protocols, thus hybrid multiple access schemes and enhanced routing schemes should be further developed.

- Fairness. To better support various applications in WMNs, it is necessary to optimize the spectrum utilization as well as provide fair services to users. In order to fullfill such requirements, new MAC protocols should be developed to implement local fairness and enhanced transport protocol should be proposed to accomplish end-to-end fairness. Besides, RRM and MM also play vital roles, which should be considered jointly.

- Apart from fairness, other performance metrics, such as end-to-end transmission delay, delay jitter, aggregate and per-node throughput, and packet loss rate should be considered in routing protocols and RRM schemes to provide QoS support.

- Topology control. One of the advantages that WMNs inherit from MANETs

is self-configuring, which requires network devices to have updated knowledge of the dynamic network topology. Topology-aware MAC and routing protocols should be proposed to support self-configuration and self-healing in WMNs.

- Security. The architecture of WMNs determines that their security mechanism needs to be redesigned, ranging from encryption algorithms to security key distribution. Without centralized trusted authority, the security schemes of a WLAN cannot be directly applied to WMNs.

It can be seen from above that not all the factors are mutually exclusive, instead, most of them are interrelated. Hence, the development of protocols or schemes at one layer in WMNs should not be isolated from other layers. From Subsection 1.3.1 to Subsection 1.3.3, the key problems of MAC layer, RRM and MM in WMNs are addressed.

1.3.1   MAC Layer

There are some major differences separating WMNs from the conventional wireless networks such as MANETs, including the network architecture and the node functionality. In general, the traditional MAC protocols for MANETs were mainly designed for nodes with similar structure and functions. In WMNs, MRs and mesh clients hold distinct characteristics. For instance, MRs usually have much higher transmission power than mesh clients, and MRs have low or no mobility while mesh clients can move much more dynamically. Due to such heterogeneous nature, MAC schemes proposed for traditional wireless networks with homogeneous network devices cannot be directly employed in WMNs. In multi-channel WMNs, simultaneous transmissions

are allowed without collisions or interference. Distributed multi-channel MAC protocols should be proposed so that MRs and mesh clients can intelligently switch among different channels. In addition, MAC protocols that were designed for traditional single-hop networks do not work efficiently in the multi-hop environment of WMNs. The hidden-node issue will happen in each hop from source to destination, thus the multi-hop issue should not be left to routing protocols alone to deal with.

Apart from the main task of a MAC protocol to provide fair and efficient resource sharing, the above differences should be taken into account as well in developing new MAC protocols for WMNs to ensure scalability. At the MAC layer, the scalability of WMNs can be addressed from two different perspectives. One is to enhance or reinvent the current MAC protocols to increase end-to-end throughput as well as guarantee user's QoS in single channel case, another is to develop multi-channel MAC. Regarding the single channel case, several MAC schemes have been proposed by enhancing the CSMA/CA protocol (Calì *et al.*, 2000)- (Qiao and Shin, 2003). By adjusting the contention window size and modifying the backoff procedure, the single-hop throughput can be considerably improved, but the multi-hop throughput decreases as the number of hops and length of path increase. Since random access protocols such as CSMA/CA are not an efficient solution for multi-hop environment, TDMA-based MAC protocols are proposed (Cicconetti *et al.*, 2008)- (Trung and Mo, 2010). In IEEE 802.16, the original MAC protocol is a centralized TDMA-based scheme, while a distributed TDMA MAC for IEEE 802.16 mesh is still under investigation. In IEEE 802.11-based WMNs, how to design a distributed TDMA-based MAC protocol overlaying CSMA/CA is still an open issue.

By applying multiple-channel techniques, multipoint-to-multipoint communications can be conducted at the same time so as to improve the transmission throughput. In multi-channel single-transceiver scenario, only one channel is active at a time in each network node but different nodes may operate on different channels simultaneously in order to improve system capacity. The work in (So and Vaidya, 2004) belongs to this case. Without restriction of cost, multi-channel multi-transceiver per node can be another option, but there is only one MAC layer to coordinate the functions of multiple channels on top of the physical layer. Little work has been done in this field (Bononi *et al.*, 2009), but multi-channel multi-transceiver MAC will be a promising solution for WMNs since it can maximize both spatial and spectral reuse to offer better throughput, resource utilization and scalability. In addition to above two single-radio solutions, multi-radio MAC can also be designed with little hardware modifications. By using multiple legacy IEEE 802.11 network interface cards (NICs) per node, each node has multiple radios with its own MAC and physical layers. Communications in these radios are totally independent. Therefore, a virtual MAC protocol such as the multi-radio unification protocol (MUP) proposed in (Adya *et al.*, 2004) is required on top of MAC to coordinate communications in all channels.

In addition to achieving high throughput and good fairness, a number of applications in WMNs need to support broadband multimedia communications. Thereby, the development of MAC protocols with multiple QoS metrics such as delay, packet loss rate, and delay jitter is important.

1.3.2   Radio Resource Management

RRM is a policy that determines how network resources are assigned to users with the objective of optimizing some functions such as total network throughput or average data transmission delay, subject to some constraints like maximum data transmission delay or packet loss rate requirements. Both entire network performance and individual user's satisfaction can highly depend on RRM techniques. Generally speaking, RRM techniques should consider the following resources (Ahmed, 2005):

- Access of end nodes to other network devices, such as APs, MRs or relay stations, which are usually shared by surrounding nodes.

- Frequency channel allocations and time slots scheduling. This determines how much bandwidth and when this portion of bandwidth is allocated to individual links, and therefore affects the data transmission performance, such as transmission rate and delay.

- Transmission power. This not only determines the signal quality of individual transmissions but also affects mutual interference level among users.

1.3.2.1   Admission Control

Among all techniques in RRM, call admission control (CAC) normally acts as the long-term (the whole duration of a connection or a session) guarantee for various QoS requirements, such as call dropping probability, packet delay, delay jitter, and throughput. CAC is essential to ensure the total amount of admitted traffic below the network capacity. The heavier the network load is, the more deteriorated is the service quality for users. As a result, CAC accepts users only if it can maintain satisfactory

service quality to admitted users (including both existing users and new users). In this case, the admission criterion can be the number of users, signal-to-interference ratio (SIR), or transmitted/received power. From users' point of view, dropping an ongoing call usually appears to be more troublesome than blocking a new call, thus a CAC mechanism is mainly applied on restricting new calls in favour of ongoing calls and handoff calls. Sometimes, an overwhelming number of packets waiting to be transmitted in a wireless network can cause excessive packet delay and delay jitter, and the throughput may also drop to an unacceptable level. Therefore, CAC should be able to limit the amount of admitted traffic to guarantee the packet-level QoS parameters.

1.3.2.2   Scheduling

After being admitted to the network, users are controlled by another technique in RRM, which is scheduling. Scheduling is part of the MAC layer protocol, and decides the best moment to grant bandwidth to admitted users. By principle, wireless channel is a broadcast medium, where all users can share and compete the radio resource simultaneously. One major problem with this is that, if concurrent transmissions happen in the same carrier frequency at the same time, from the pure MAC layer's perspective, it may induce packet collisions. Even if some techniques such as the spread spectrum technology in code division multiple access (CDMA) can be used to reduce such effect, mutual interference can still affect signal transmission quality. Therefore, a scheduling mechanism plays a central role in organizing transmissions to decrease or avoid co-channel interference while efficiently utilizing the radio resources.

One of the objectives for scheduling is to provide services with different priorities.

For instance, real-time voice traffic should be granted channel resources with a higher priority than elastic data traffic. Users in the same traffic class can be prioritized according to subscription fees or urgency. Fairness is another objective that scheduling aims to achieve among different users in the same class, so that no one can dominate the entire resources. First, a well-designed scheduling scheme should achieve high throughput. Second, it should also prioritize the traffic with different classes, such as giving higher priority to traffic that is more urgent or represents better revenues for the network operators. Third, it should allow users to fairly share the network resources, especially in saturated networks, starvation of nodes with the same QoS level should be avoided.

Utility-based frameworks are widely applied in scheduling schemes to balance the bandwidth utilization and the users' fairness. A scheduling problem can be formulated to maximize a certain utility function (for example, a concave function of throughput) subject to some constraints. The utility function can be set from different points of view. For best effort (BE) services, the basic objective can be maximizing the total throughput of all users without considering service fairness. This sacrifices the benefit of users with poor link conditions. For instance, by giving scheduling priority to the MSs close to the AP, MSs far away from the AP are likely to be starved and get little chance to access channel. Opportunistic algorithms such as the one proposed in (Liu *et al.*, 2001) optimize aggregate throughput, but always favour privileged users and penalize disadvantaged users, which degrades the fairness among users. One type of the scheduling algorithms that can solve such issue is max-min fairness such as the one proposed in (Tassiulas and Sarkar, 2002), which tries to fully satisfy the users with poor transmission conditions and evenly split the rest of resources for users with

good link quality. It was proved to achieve higher throughput than that the equal fairness scheduling does but still lower than the maximum throughput scheduling does. In order to achieve the best tradeoff between effective resource utilization and users' satisfaction, PF (Kelly, 1997)-(Kelly *et al.*, 1998) was introduced so that multiple users can be served either simultaneously at a time or through a long-term iterative scheduling. The PF is in the form of logarithmic utility function, which is a specific case of Nash bargaining solution (Nash Jr, 1950)-(Zhou, 1997). It was proved to be a compromise-based scheduling method, which maintains a balance between two competing interests: maximizing total wireless network throughput and allowing each user at least a minimal level of service. In a real system, PF scheduling has been employed at the BSs to schedule the downlink flows among different users for 3G wireless data networks (Jalali *et al.*, 2000).

Some techniques approach the scheduling problems using simple heuristics. However, with the systematic mathematical framework, utility-based optimization gives a better alternative to address fairness than some heuristic approaches. By relaxing certain optimization constraints, some approximate results can be close to the theoretical optimum, hence providing performance bounds for the scheduling design of WMNs.

1.3.2.3   Implementation of CAC and Scheduling

Both CAC and scheduling schemes for wireless networks can be classified as centralized and distributed. Centralized CAC/scheduling is mainly designed for infrastructure-based networks, where a central controller can collect channel and other necessary

information to determine the network resource allocations. Acquiring global knowledge of the entire network can ensure a collision-free wireless environment. In distributed CAC/scheduling, each node determines its own access to the channel on the basis of local sense of the medium, which can easily cause transmission collisions or interference (Gummalla and Limb, 2000). Compared to centralized CAC/scheduling, distributed CAC/scheduling is more realistic for networks without a central controller, but its design can be more challenging. In many cases, a distributed approach can turn a complicated network into a simple structure that allows a collection of subsystems to coordinate their actions to achieve global optimality or close-to-global optimal solutions.

Sometimes both centralized and distributed approaches may coexist in the same network, and global information can also be obtained in a distributed manner with information exchanged between correlated nodes. The nature of WMNs determines that neither pure centralized schemes designed for cellular networks (Viswanathan and Mukherjee, 2005) nor existing distributed schemes (Jiang *et al.*, 2002) (Jiang *et al.*, 2007) proposed for MANETs can be directly applied. Considering the different networking characteristics between the mesh backbone and mesh clients in WMNs, hybrid CAC/scheduling schemes are more desirable.

1.3.2.4   Transmission Power Control

Transmission power control is another concern in RRM. In WMNs, there's usually no restrict to power consumption of MRs, and the aims of power management are to control connectivity, interference, spectrum spatial reuse and topology (Li *et al.*, 2005). If a single channel is shared among MRs, co-channel interference will further

impact the signal transmission quality. Thus, transmission power control is crucial in reducing the interference. Most recently, as the growing popularity of vehicular wireless communications, WMNs have been made to provide services for fast-moving vehicles. In order to accommodate the vehicular communications in WMNs, MRs tend to transmit much higher power compared to that in the legacy IEEE 802.11-based WMNs, so that the coverage areas for high-speed moving vehicles are magnificently extended. Under such circumstance, the transmission distance of an MR can be several kilometers or even longer (Jones, 2005), which also makes it possible to have larger overlapping coverage areas among the MRs. However, higher power subsequently results in spectral interference. Therefore, adaptive power allocation (APA) has been applied as one of the solutions on decreasing the interference in the network as well as minimizing the MR energy consumption (Yang and Zhao, 2011). For MSs, energy is mainly spent on computation and communications, wherein the communication energy consumption takes the major proportion of the whole. If a collision occurs, either a retransmission should be scheduled which consumes more energy, or the data is lost. Thereby, power control and scheduling should be jointly considered to effectively optimize the energy consumption while satisfying the user's requirements.

1.3.3   Mobility Management (MM)

MM comprises two components: location management and handoff management. Location management includes address management and movement detection. The former is to identify each mobile terminal (MT) when it moves around, and the latter is to determine the current location of the MT based on the information available

at the system databases when a communication for the MT is initiated (Xie and Wang, 2008). Handoff management is responsible for initiating handoff requests, establishing connectivity to new APs, switching connections from old APs to new APs, and keeping smooth transmissions for connections during and after handoffs.

There are a collection of MM schemes developed for cellular (Akyildiz *et al.*, 1999) and mobile IP networks (Akyildiz *et al.*, 2004) that mainly rely on wired infrastructure for mobility-related signaling traffic delivery. However, it might not be useful for WMNs lacking of infrastructure such as home location register (HLR) and visitor location register (VLR). In general, MM schemes and routing protocols are designed independently in cellular and mobile IP networks, whereas MM and routing should be jointly considered in multi-hop WMNs to reduce the transmission and handoff delay and failure rate of location update. Regarding MANETs, MM schemes are mainly proposed in two types: distributed (Haas and Liang, 1999) and hierarchical (Sucec and Marsic, 2002)-(Chen and Chen, 2003). However, these schemes for MANETs may not perform well for WMNs due to the specific features of WMNs. Specifically, APs in WMNs do not have high mobility as mobile nodes in MANETs while the MSs in WMNs may constantly roam across different APs, thus the distributed MM proposed in MANETs cannot be directly applied to WMNs. On the other hand, the focus of handoffs in a traditional MANET is rerouting. That is, to find a feasible multi-hop path quickly so that a flow can be handed off to the new path upon link breakage (Wu and Harms, 2006), which is not the same concept and main focus as in WMNs. Mobility may affect the performance of the MAC layer as well, since the highly dynamic changes in network topology may significantly impact the performance of the MAC protocol. In order to be adaptive to mobility or even to make use of the

mobility (Grossglauser and Tse, 2002), the network nodes need to exchange topology information. Considering the service fairness among highly mobile MSs, schemes designed for cellular networks or MANETs cannot be employed in WMNs. As a result, new handoff management schemes need to be developed for WMNs according to their own nature. Mobility management is closely related to multiple layers of network protocols, thus the development of multi-layer mobility management schemes as in (Dutta *et al.*, 2002) is an interesting topic.

1.4   Related Works

In this section, we summarize some related works on WMNs, including both industry standards and research works.

1.4.1   Industry Standards

IEEE standard groups have played a key role in the development of WMNs standards with network coverage ranging from personal area networks (PANs) to metropolitan area networks (MANs). The IEEE 802.11 family of standards (Group, 2002a) have been activating since the end of 20th century, and are proved to be the most successful wireless networking standards for WLANs. From IEEE 802.11a/b/g to IEEE 802.11e/n, the primary objective is to implement one-hop communications, which will result in throughput degradation and unfairness in multi-hop environment. Motivated by the issues and limitations, a separate task group called "IEEE 802.11s extended service set (ESS) Mesh" was founded in May 2004 under the 802.11 work group to address the demand for WMNs (Hauser, 2003). The IEEE 802.11s with

support of the original IEEE 802.11 DCF protocol and the IEEE 802.11e EDCA protocol aims to create a wireless distribution system (WDS) that supports both broadcast and unicast delivery at the MAC layer using radio-aware metrics over self-configuring multi-hop topology. In addition, new MAC features were proposed to enhance QoS, including an intra-mesh congestion control scheme to alleviate congestion at the mesh nodes, an optional multi-channel MAC protocol to increase link capacity, and an optional mesh deterministic access (MDA) scheme. Based on ultra wide band (UWB) orthogonal frequency division multiplexing (OFDM) technique, IEEE 802.15.3a standard (Group, 2002b) intends for high throughput PANs with communication distances of around 10m (or less), whereas IEEE 802.15.4 specifies for low rate WPANs (LR WPANs) with low data rate, long battery life and low device cost. The ZigBee Alliance (Alliance, 2003) has been developing the application layer, the network layer, and the security layer protocols based on IEEE 802.15.4 MAC and PHY layers that operate in unlicensed bands worldwide. Afterwards, IEEE 802.15.5 TG, kicked off in May 2004, is currently working to provide an architectural framework for inter-operable, stable, and scalable WMNs for WPAN devices. On the basis of meshed tree approach, it proposes the meshed tree routing, multicasting, and key pre-distribution. IEEE 802.16 (Group, 2001a) focuses on serving the broadband wireless access in MANs, supporting both point-to-multipoint (PMP) mode and mesh (multipoint-to-multipoint) mode to extend fiber optic backbones. The 802.16a extension established in January 2003, uses a lower frequency of 2-11 GHz that allows non-line-of-sight (NLOS) communications, ranging up to 50km. To further support the broadband connection to Internet for customers in moving vehicles, the IEEE 802.16e was developed. One of the commercial applications based on IEEE 802.16 is

WiMax (Forum, 2001), which is a good solution for places where the cost of laying or upgrading landlines to broadband capacity is expensive.

In addition to industrial standards, a great number of academic works have emerged to address various issues coming up with different protocol layers.

1.4.2   MAC Layer and Scheduling

Since scheduling is considered as part the MAC layer protocol, a lot of works focus on designing and improving the MAC layer algorithm to attain fair and efficient scheduling. An interesting scheduling mechanism is proposed for WLANs in (Luo et al., 2000) based on the notion of a flow contention graph that takes into account the topology of the network and the objective is to achieve weighted fairness while maximizing throughput. A priority-based fair medium access control protocol is proposed in (Qiao and Shin, 2002) by modifying the DCF of the IEEE 802.11 MAC to achieve weighted fairness and maximize aggregate throughput, where the key idea is to choose a contention window size for each wireless station according to its relative weight among data traffic flows. In (Heusse et al., 2005), an access method named "Idle Sense" is proposed. The method is derived from IEEE 802.11 Distributed Coordination Function (DCF), in which all hosts estimate the collision rate and use similar values of the contention window (CW) to benefit from good short-term access fairness as well as offer high throughput, low collision and contention overhead. A distributed MAC protocol is proposed in (Dugar et al., 2001) for wireless networks to support prioritized scheduling along with a weighted fair sharing of the bandwidth among the users belonging to the same priority level. A distributed scheduling algorithm is proposed for supporting real-time traffic in a WLAN (Sobrinho and Krishnakumar,

1996). In (Vaidya *et al.*, 2005), a fully distributed algorithm is derived from the DCF in the IEEE 802.11 standard for weighted fair scheduling in a WLAN, which was proved to allocate bandwidth to different flows proportional to their weights. Another scheme of dynamic adaptation of the Scaling Factor is also designed to achieve good performance regardless of the choice of the initial Scaling Factor or the assignment of weights to the flows. A fair scheduling mechanism called distributed elastic round robin (DERR) is also proposed in (Ferng *et al.*, 2005) for IEEE 802.11 WLANs operated in the ad-hoc mode, which outperforms IEEE 802.11e in terms of delay and throughput. (Mingyi Hong, 2013) studied a joint BS assignment and resource allocation problem in a multi-cell downlink OFDMA network. They proposed a strategy-proof mechanism, where each user dynamically selects a BS to maximize its own utility, and the BSs implement the celebrated Vickrey-Clark-Groves mechanism (VCG) to allocate resource under the current user-BS assignment profile, which achieves at least $\frac{1}{2}$ of the optimal throughput. In (Salem and Hubaux, 2005), a fair scheduling mechanism is proposed to assign transmission rights to the links in WMNs that maximizes the spatial reuse.

Among all the criteria for making scheduling decisions, maximizing the minimum (max-min) throughput seems to be one of the most popular approaches in early generation because it can be easily implemented by studying contention graph through MAC layer. An online prioritized max-min scheduling scheme is proposed in (Tassiulas and Sarkar, 2002) for wireless ad-hoc networks, which assigns dynamic weights to the flows to attain both short term and long term fairness. An algorithm is proposed in (Bejerano *et al.*, 2004) to determine the user-AP associations obtained a near optimal max-min fair bandwidth allocation in a WLAN. In (Huang and Bensaou,

2001), distributed algorithms are designed based on IEEE 802.11 MAC, that allow each node to determine its max-min per-link fair share in a global ad-hoc network without knowledge of the global topology of the network. Bandwidth allocation is studied in (Tang *et al.*, 2006) in multi-channel multi-hop WMNs, and an LP-based polynomial time algorithm on the basis of max-min algorithm is designed for throughput maximization and fairness enhancement. A centralized algorithm based on the IEEE 802.11 MAC protocol for max-min capacity calculation is designed in (Aoun and Boutaba, 2006) for WMNs to ensure fair and efficient use of network resources. (Staehle *et al.*, 2009) investigated the potentials of link rate selection for increasing the max-min fair throughput of WMNs. As the development of utility optimization framework, PF algorithm has attracted more researchers in recent years. (Banchs *et al.*, 2007) proposed a more efficient throughput allocation criterion based on PF for IEEE 802.11e WLAN standard. (Boche *et al.*, 2007) investigated the problem of weighted proportional fairness for log-convex interference in WLANs. (Bensaou and Fang, 2007) studied a bandwidth sharing problem in wireless multi-hop networks as a general utility maximization problem with link bandwidth constraints to guarantee fairness, and derived a gradient-based iterative algorithm to solve the problem.

1.4.3   End-to-End Performance and Cross-Layer Resource Management

Beside single-hop scheduling, some works paid attention to the end-to-end allocation and moved their eyes on upper layer protocols. In (Li, 2005), author analyzed the issue of increasing spatial reuse of bandwidth from an end-to-end perspective of multi-hop flows. The proposed algorithm can maximize end-to-end throughput in

wireless ad-hoc networks, while still maintaining basic fairness across the multi-hop flows. (Raniwala *et al.*, 2007) designed and implemented a centralized coordinated congestion control algorithm (C3L) for transport layer that performs global bandwidth allocations and thus provides end-to-end flow-level max-min fairness over the unmodified IEEE 802.11 MAC layer. It proves that a well-designed transport protocol does yield significantly better fairness and should be able to achieve fairness across flows and a simultaneous high overall network throughput even on top of an unfair MAC layer such as IEEE 802.11. (Jun and Sichitiu, 2003) proposed several network layer solutions to the fairness problem for wireless multi-hop networks, which can provide prioritized QoS and improve network efficiency.

Due to interrelationship of different factors in RRM, a great many joint design schemes for WMNs have been proposed. Some of the recent work include joint channel allocation and routing (Meng *et al.*, 2006), (Chen *et al.*, 2006), (Alicherry *et al.*, 2005), joint routing and interface assignment (Kyasanur and Vaidya, 2005), joint topology control and interface assignment (Mohsenian-Rad and Wong, 2007), joint routing and media access control (Zhang *et al.*, 2005), joint topology control and routing (Tang *et al.*, 2005), joint power control and channel allocation (Zhang and Qu, 2006), and joint channel allocation and congestion control (Mohsenian Rad and Wong, 2006). Apart from those joint studies over two factors, the scheduling, power control and rate control problem in WMNs have been investigated in (Korkmaz *et al.*, 2006) to provide bandwidth guarantees to traffic flows and to optimize transmission resources utilization. (Rad and Wong, 2007) formulated the joint channel allocation, interface assignment, and MAC problem as a cross-layer non-linear mixed-integer network utility maximization problem, and solved with global and dual decomposition techniques

to maximize the aggregate network throughput. (Luo *et al.*, 2010) jointly studied the routing, scheduling, power control, and rate adaptation problems, and developed two computational tools to obtain the theoretically near-to-optimal solutions for WMNs.

1.4.4   Mobility and Handoff

In traditional IEEE 802.11-based WMNs, the handoffs and RRM problems are usually studied separately. One of the main reasons is that performing handoffs in traditional IEEE 802.11-based WMNs requires considerable signaling overhead, which can cause a long delay. Therefore, the main objective for handoff management in IEEE 802.11-based WMNs has been either to reduce handoff delay to achieve fast handoffs, or to minimize packet losses during handoffs. (Zhao and Xie, 2010) proposed an XMesh architecture enabling fast handoffs for supporting real-time applications, in conjunction with a Xcast-based caching mechanism that builds on top of mesh routing protocols to guarantee minimum packet loss during handoffs. (Chi *et al.*, 2006) introduced an approach to ameliorating cache effectiveness in APs for mobility support in an IEEE 802.1X framework, which significantly reduces the handoff delay. To prevent high signaling cost and long handoff latency (Huang *et al.*, 2007) presented a new mobility management scheme for WMNs, Mesh Mobility Management (M^3). It utilizes some WMN's features and combines the per-host routing and tunneling techniques to better support IP micro-mobility in WMNs. A network level mechanism for supporting seamless handoff of users between APs in WMNs is proposed to minimize packet losses during handoffs in (Wei *et al.*, 2006).

Handoff management can also be employed for load balance. Approaches are proposed in (Velayos *et al.*, 2004), (Zhao, 2007) to adjust the MS-AP associations to

balance the traffic load among the APs. (Xie *et al.*, 2006) proposed a scheme for domain partition to achieve the tradeoff between load-balancing and inter-domain mobility to reduce the negative impact of the host mobility. (Elshenawy *et al.*, 2009) proposed several MAC address caching schemes for handing off vehicles between WMN APs with the objective of balancing traffic loads among APs, which reduces handoff latencies and improves the overall network throughput at different vehicle speeds. Although these works address the issues by balancing traffic load with mobility control, fair services to individual MSs or end-to-end flow fairness with handoff management has not been considered.

With the growing interests in providing Internet access to vehicles and enabling communications with outside vehicular networks, some works in the literature (Ahmed and Yanikomeroglu, 2009), (Yang *et al.*, 2007) studied the service fairness in vehicular communication networks by considering the effect of different factors (such as user mobility and link adaptation) on the service fairness. (Chen *et al.*, 2008) proposed a new fast handoff scheme (the DeuceScan scheme) to reduce the probe delay for IEEE-802.11-based WLANs, which can improve wireless communication qualities on vehicles and efficiently reduce the layer-2 handoff latency. In (Amir *et al.*, 2006), the authors presented the architecture and protocols of SMesh, a completely transparent wireless mesh system that supports seamless and fast handoff, VoIP and other real-time applications based on IEEE 802.11 MAC. But not much work is available in the literature to jointly consider the AP resource allocations and MS handoffs to achieve both fair and efficient resource allocations in a distributed manner.

1.5   Motivation and Overview of the Thesis

1.5.1   Motivation of the Thesis

In a WMN with a large number of APs and MSs, maintaining the service fairness among the MSs located in different geographical areas and having different mobility patterns can be difficult. In general, MSs associated to heavily loaded APs or having poor channel conditions to the associated APs tend to receive poorer QoS than those associated to lightly loaded APs and having good channel conditions. Although balancing the traffic loads among different APs can help achieve fair QoS among the MSs (Zhao, 2007), such an objective cannot be easily achieved without a central controller, which should coordinate both the association relationship between the MSs and the APs and the amount of resources allocated to the MSs associated to the same and different APs. The problem can be more challenging if the MSs follow different mobility patterns. For example, some MSs may move in a small area around heavily loaded APs, while others can frequently handoff between both heavily and lightly loaded APs. Furthermore, as the MSs move around in the network service area, a heavily loaded AP may become lightly loaded later, and vise versa. Therefore, the handoff decisions, or the MS-AP associations, and the network resource allocations should be updated based on the current network conditions.

The performance of load balancing to achieve fair QoS among the MSs is strongly dependent on the overlapping coverage areas between the APs. When a large number of MSs are located in the vicinity of one AP (or a small number of APs) but not of other APs, these MSs can receive much lower throughput than other MSs. With the applications of WMNs for vehicular communications, the APs can transmit

much higher power in order to extend the coverage areas for accommodating the vehicles with high moving speeds. The transmission distance of an AP can be several kilometers or even larger, which provides more possibility to have large overlapping coverage areas among the APs. Adaptive transmission rate also allows the APs to extend transmission range at a price of reduced data transmission rate. As a result, the MSs have more flexibility to choose their associated APs, and the network can achieve better performance for balancing the QoS provided to the MSs associated to both the same and different APs. This also brings good potential to achieve fair QoS among the MSs and efficient utilization of network resources. Meanwhile, the resource availability of neighboring APs is more dependent on each other, making the resource allocations even more complicated.

In traditional IEEE 802.11-based WMNs, the handoff and resource management problems are usually studied separately. Performing handoffs in traditional IEEE 802.11-based WMNs requires considerable signaling overhead, which can cause a long delay to data transmissions. Therefore, the main objective for handoff management in IEEE 802.11-based WMNs has been either reducing handoff delay to achieve fast handoffs, e.g., (Zhao and Xie, 2010), (Chen *et al.*, 2008), (Fashandi and Todd, 2005), (Chi *et al.*, 2006), or minimizing packet losses during handoffs, e.g., (Wei *et al.*, 2006). Although adjusting the MS-AP associations can balance the traffic loads among the APs, such as in (Velayos *et al.*, 2004), (Elshenawy *et al.*, 2009), (Zhao, 2007), providing fair services to individual MSs is not considered in these works. Some works, such as (Ahmed and Yanikomeroglu, 2009), (Alasmary and Zhuang, 2012), (Borst *et al.*, 2009), (Yang *et al.*, 2007), study the vehicular communication networks by considering the effect of different factors (such as user mobility and link

adaptation) on the service fairness to vehicles. Not much work is available in the literature to jointly consider the effects of AP resource allocations and MS handoffs on fair and efficient resource allocations.

Concerning the application popularity and all the issues above, designing a distributed scheme that can adapt to fast changing network topology, provide fair service to users, and improve the system utilization is necessary.

1.5.2 Overview of the Thesis

In this thesis we study AP resource allocations in a WMN in order to provide fair throughput to the MSs as well as efficiently utilize the channel time. In the first part of the thesis, we assume that the channel time allocations at the AP level are given, and jointly consider the handoff management of the MSs and the channel time allocations at the MS level. An optimization problem is formulated based on long-term proportional fairness, so that all the MSs in the entire WMN can receive fair average throughput, while the total throughput of the MSs is maximized. The optimum solution is based on the assumption of having global information about the channel conditions and mobility information of all the MSs, and cannot be easily implemented in a practical system. A heuristic scheme is then proposed, which allows each AP to allocate its channel time based on local information only, and the handoff decisions of the MSs are based on information exchanged between neighboring APs. Numerical results show that performance of the proposed heuristic scheme is very close to the optimum in terms of both fairness and throughput.

In the second part of thesis, we consider that different APs can share the same frequency channel and should coordinate for their transmission time, so that neighboring

APs sharing the same frequency channel do not transmit at the same time in order to avoid strong co-channel interference. We jointly study the channel time allocations at the AP level and the MS level together with the MS handoff management. An optimization problem is first formulated and solved, and the objective is to provide fair average throughput for the MSs while efficiently utilizing the channel resources. Two distributed schemes are proposed by decoupling the handoff management and time allocations. The HO-CA scheme performs heuristic handoff decisions for the MSs based on their link gains to nearby APs, and then optimizes the channel time allocations. The CA-HO scheme allocates the channel time to individual APs based on the neighboring relationship of the APs, and then allows the MSs to make handoff decisions based on possible utilities from nearby APs. In both schemes, individual APs solve a local optimization problem to allocate channel time for their associated MSs. Numerical results indicate that both the distributed schemes can achieve close-to-optimum fairness; and with higher computational complexity, the HO-CA scheme achieves higher throughput.

The remainder of the thesis is organized as follows. In Chapter 2, joint handoff management and channel time allocations at the MS level is studied. In chapter 3, joint handoff management and channel time allocations at both the AP and MS levels is studied. In Chapter 4, we conclude the thesis and discuss a few future research topics.

Chapter 2

Handoff Management and Channel Time Allocations at MS Level

In this chapter we study MS handoff jointly with channel time allocations at the MS level in order to provide fair throughput to the MSs over a long term. First, an optimization problem is formulated and solved based on long-term PF. The objective is to maximize the total throughput for all the MSs in the network subject to maintaining long-term fair throughput among MSs. A heuristic scheme is then designed, so that each AP can make decisions of its resource allocations based on local information only, and each MS can choose its associated AP based on information from nearby APs. The rest of the chapter is organized as follows. In Section 2.1, we describe the system that this work is based on. The optimum solution is formulated and solved in Section 2.2, and the heuristic scheme is proposed in Section 2.3. Numerical results are demonstrated in Section 2.4, followed by a summary in Section 2.5.

2.1 System Model

We first consider a general network, where M links share the transmission channel. Channel time is divided into equal length frames. The throughput of link m at frame t is denoted as $\eta_m(t)$, and $\overline{\eta}_m = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \eta_m(t)$ is the long-term average throughput of link m. At each frame t, $\overline{X}_m(t)$ is used to represent a weighted average rate for link m up to time t, and is given by

$$\overline{X}_m(t) = (1-\gamma)\overline{X}_m(t-1) + \gamma\eta_m(t-1), \tag{2.1}$$

where $0 \le \gamma \le 1$ is a parameter that balances the weights of the throughput in the past and the most recent frame. When γ is close to 1, more weight is given to the most recent throughput. For a special case, $\overline{X}_m(0) = 0$. A general scheduling problem can be formulated to maximize a certain objective function $U(\overline{\mathbf{X}})$, where $\overline{\mathbf{X}} = [\overline{X}_1(t), \overline{X}_2(t), \ldots, \overline{X}_M(t)]$. When $U(\overline{\mathbf{X}}) = \sum_{m=1}^{M} \log \overline{\eta}_m$, PF is achieved for the average throughput among the links. PF is proved to be the best tradeoff between effective resource utilization and users' satisfaction. It is proved in (Stolyar, 2005) that this objective can be achieved iteratively through frame-by-frame scheduling, and the objective at frame t is to find the throughput vector $\boldsymbol{\eta}(t) = [\eta_1(t), \eta_2(t), \ldots, \eta_M(t)]$ in order to maximize $\sum_{m=1}^{M} \frac{\eta_m(t)}{\overline{X}_m(t)}$. Intuitively, in order to achieve this objective, the MSs with either higher $\eta_m(t)$ (good link condition to the AP at the current frame) or lower $\overline{X}_m(t)$ (low average throughput in the past) should be given a higher priority to transmit at frame t.

In this work, we consider a WMN with multiple APs. In a typical file downloading scenario, the downlink traffic dominates the overall traffic load in the network. That

is, APs transmit to the MSs for almost all the channel time. For this reason, we focus

on the downlink transmissions from the APs to the MSs and study the throughput

provided to individual MSs. MSs move in random speeds and directions within the

service area of the network. The link gains are considered to be constant during each

frame, which can be one or several superframes. All the APs periodically transmit

the beacon signals with fixed power, which is known to the MSs and used by the

MSs to measure their link gains to the nearby APs. The MSs report the link gains to

their currently associated APs at the beginning of each frame, and this information

is used by the APs for allocating resources in this frame. The link gain information

to nearby APs can also help the MS make handoff decisions. We define a set of

binary variables $A_{im}(t)$'s. When MS m is associated to AP i at frame t, $A_{im}(t) = 1$;

otherwise, $A_{im}(t) = 0$. Then we have

$$\sum_{i \in \mathcal{I}} A_{im}(t) \leq 1, \tag{2.2}$$

for all $m \in \mathcal{M}$, where $\mathcal{I}$ is a set of all the APs, and $\mathcal{M}$ is a set of all the MSs

in the network. We consider that transmissions of different APs are orthogonal to

each other, either because they transmit at different frequency channels, or at the

same frequency channel but in different time intervals. MSs communicate directly

with their associated APs. Handoff of the MSs between different APs can only be

performed at the beginning of a frame, and the association relations between the MSs

and the APs are kept unchanged during each frame. Let T_i be the total amount of

time that AP i can transmit during one frame. The value of T_i can be the entire

duration of one frame, if AP i does not share the frequency channel with other APs.

When multiple APs share the same frequency channel, they may not all transmit at

the same time in order to avoid strong co-channel interference. The amount of the time that each AP can transmit during one frame may depend on the number of the MSs associated to each AP, interference level, and other factors. We assume that the values of T_i's are all known for now, and will study how to find the optimum T_i's in our future work.

Let $P_{im}(t)$ and $T_{im}(t)$, respectively, represent the amount of power and time that AP i transmits to MS m during frame t, $R_{im}(t)$ represent the instantaneous transmission rate of AP i when it is transmitting to MS m, and $\mathcal{M}_i(t)$ be a set of the MSs associated to AP i at frame t. The product of $T_{im}(t)$ and $R_{im}(t)$ gives the throughput (total amount of transmitted data) of MS m during frame t, or

$$T_{im}(t)R_{im}(t) = \eta_m(t), \tag{2.3}$$

for all $m \in \mathcal{M}_i(t)$. The instantaneous transmission rate to MS m is determined by the transmission power of the AP to the MS, $P_{im}(t)$, and the link gain between them, $g_{im}(t)$, i.e.,

$$R_{im}(t) = W \log_2 \left(1 + \frac{P_{im}(t)g_{im}(t)}{P_{noise}} \right), \tag{2.4}$$

where W is the transmission bandwidth, and P_{noise} is the power of the background additive white Gaussian noise.

2.2   Optimum Solution

Based on the description in the previous section, the following optimization problem can be formulated to achieve long-term proportional fair throughput among the MSs, where $\boldsymbol{\eta}(t) = [\eta_m(t), m \in \mathcal{M}]$, $\mathbf{P}(t) = [P_{im}(t), i \in \mathcal{I}, m \in \mathcal{M}]$, $\mathbf{T}(t) = [T_{im}(t), i \in$

$\mathcal{I}, m \in \mathcal{M}]$, and $\mathbf{A}(t) = [A_{im}(t), i \in \mathcal{I}, m \in \mathcal{M}]$.

$$\text{P1.} \quad \max_{\boldsymbol{\eta}(t),\mathbf{P}(t),\mathbf{T}(t),\mathbf{A}(t)} \sum_{m \in \mathcal{M}} \frac{\eta_m(t)}{X_m(t)} \tag{2.5}$$

$$\text{s.t.} \quad \eta_m(t) = \sum_{i \in \mathcal{I}} A_{im}(t) R_{im}(t) T_{im}(t), \; m \in \mathcal{M}, \tag{2.6}$$

$$\sum_{i \in \mathcal{I}} A_{im}(t) \leq 1, \; m \in \mathcal{M}, \tag{2.7}$$

$$\sum_{m \in \mathcal{M}} T_{im}(t) \leq T_i, \; i \in \mathcal{I}, \tag{2.8}$$

$$0 \leq P_{im}(t) \leq P_{\text{AP}}, \; m \in \mathcal{M}, \; i \in \mathcal{I}, \tag{2.9}$$

$$A_{im} \in \{0,1\}, \; m \in \mathcal{M}, \; i \in \mathcal{I}. \tag{2.10}$$

The problem is not convex because constraint (2.6) is not convex. Below we convert the problem into a linear optimization problem.

First, the objective function is a monotonically increasing function of $\eta_m(t)$, which is also a monotonically increasing function of $P_{im}(t)$ when $A_{im}(t) = 1$. Therefore, with the optimum solution, if $A_{im}(t) = 1$, then $P_{im}(t) = P_{\text{AP}}$, where P_{AP} is defined as the maximum transmission power of the AP. Based on this, we can rewrite (2.6) as

$$\eta_m(t) = \sum_{i \in \mathcal{I}} A_{im} T_{im}(t) \tilde{R}_{im}(t). \tag{2.11}$$

where

$$\tilde{R}_{im}(t) = W \log_2 \left[1 + \frac{P_{\text{AP}} g_{im}(t)}{P_{noise}} \right]. \tag{2.12}$$

In order to make (2.11) convex, we further define a new set of variables $Y_{im}(t)$'s, so that

$$T_{im}(t) - 10^8(1 - A_{im}(t)) \leq Y_{im}(t) \leq 10^8 A_{im}(t), \tag{2.13}$$

where 10^8 can be replaced with any number that is larger than T_i. In (2.13) we can

see that when $A_{im}(t) = 0$, the left-hand side is negative, the right-hand side is zero, and $Y_{im}(t) = 0$; when $A_{im}(t) = 1$, the left-hand side is $T_{im}(t)$, the right-hand side is a large number, and $Y_{im}(t) = T_{im}(t)$. With this, $Y_{im}(t) = A_{im}(t)T_{im}(t)$, and then (2.11) becomes

$$\eta_m(t) = \sum_{i \in \mathcal{I}} Y_{im}(t)\tilde{R}_{im}(t). \tag{2.14}$$

The original optimization problem can then be rewritten as

$$\text{P2.} \quad \max_{\boldsymbol{\eta}(t),\mathbf{T}(t),\mathbf{A}(t)} \sum_{m \in \mathcal{M}} \frac{\eta_m(t)}{\overline{X}_m(t)} \tag{2.15}$$

$$\text{s.t.} \quad \eta_m(t) = \sum_{i \in \mathcal{I}} Y_{im}(t)\tilde{R}_{im}(t), \ m \in \mathcal{M}, \tag{2.16}$$

$$\sum_{m \in \mathcal{M}} T_{im}(t) \leq T_i, \ i \in \mathcal{I}, \tag{2.17}$$

$$T_{im}(t) - 10^8(1 - A_{im}(t)) \leq Y_{im}(t), m \in \mathcal{M}, i \in \mathcal{I}, \tag{2.18}$$

$$Y_{im}(t) \leq 10^8 A_{im}(t), \ m \in \mathcal{M}, \ i \in \mathcal{I}, \tag{2.19}$$

$$0 \leq Y_{im}(t) \leq T_{im}(t), \ m \in \mathcal{M}, \ i \in \mathcal{I}, \tag{2.20}$$

$$\sum_{i \in \mathcal{I}} A_{im}(t) \leq 1, \ m \in \mathcal{M}, \tag{2.21}$$

$$A_{im}(t) \in \{0, 1\}, \ m \in \mathcal{M}, \ i \in \mathcal{I}. \tag{2.22}$$

The new problem is a mixed integer linear programming problem, which can be solved using commercial software such as AMPL. Finding a global optimum solution in a practical system is unlikely because there is no central station available to collect all required information and run the global optimization problem. Nevertheless, this provides a performance benchmark for designing practical schemes.

2.3 AP Resource Allocations and MS Handoff

2.3.1 Channel Time Allocations at MS Level

The basic idea of the proposed scheme is to decouple the handoff decisions and the AP resource allocations. Given the current associated MSs, each AP can allocate its resources to the MSs so that PF is achieved among them. For AP i, an optimization problem can be formulated as follows,

$$\text{P3.}\quad \max_{\boldsymbol{\eta}_i(t), \mathbf{T}_i(t)} \sum_{m \in \mathcal{M}_i(t)} \frac{\eta_m(t)}{\overline{X}_m(t)} \tag{2.23}$$

$$\text{s.t.}\quad \eta_m(t) = \tilde{R}_{im}(t) T_{im}(t), \ m \in \mathcal{M}_i(t), \tag{2.24}$$

$$\sum_{m \in \mathcal{M}_i(t)} T_{im}(t) \leq T_i, \tag{2.25}$$

where $\mathbf{T}_i(t) = [T_{im}(t), m \in \mathcal{M}_i(t)]$, and $\boldsymbol{\eta}_i(t) = [\eta_m(t), m \in \mathcal{M}_i(t)]$. Since the AP has the link gain information between itself and all the associated MSs, the above optimization problem can be solved locally at each AP. If this scheduling is performed iteratively for the same group of MSs, PF is achieved for the average throughput among these MSs associated to the same AP. However, the MSs may move between the coverage areas of different APs, and there is no guarantee that such fairness can be achieved among all the MSs over a long term.

Note that in problem P3, the optimum solution at frame t is for the AP to allocate all the time to the MS with the maximum $\frac{\eta_m(t)}{\overline{X}_m(t)}$. That is,

$$T_{im}(t) = \begin{cases} T_i, & \text{if } m = \arg\max_{m' \in \mathcal{M}_i(t)} \frac{\eta_{m'}(t)}{\overline{X}_{m'}(t)}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.26}$$

2.3.2   Utility-based Handoff

In this subsection, we design a scheme for the MSs to find their associated APs. This requires some coordination between the APs and the MSs. For AP i, define the average utility of the MSs currently associated to it as

$$\overline{U}_{AP,i}(t) = \frac{1}{|\mathcal{M}_i(t)|} \sum_{m \in \mathcal{M}_i(t)} \frac{\eta_m(t)}{\overline{X}_m(t)}, \tag{2.27}$$

where $|\mathcal{M}_i(t)|$ is the total number of MSs associated to AP i.

Each AP calculates its own $\overline{U}_{AP,i}(t)$ value at the end of each frame and exchanges it with its neighboring APs. With the information from its neighbors, AP i can calculate the average per MS utility in the neighboring APs as

$$\overline{U}_{av,i}(t) = \frac{\sum_{i' \in \mathcal{N}_i} \overline{U}_{AP,i'}(t)}{|\mathcal{N}_i|} \tag{2.28}$$

where $\mathcal{N}_i$ is the set of the neighboring APs of AP i, and $|\mathcal{N}_i|$ is the number of APs of AP i.

AP i then compares $\overline{U}_{AP,i}(t)$ with $\overline{U}_{av,i}(t)$. If

$$\overline{U}_{AP,i}(t) < \overline{U}_{av,i}(t) \tag{2.29}$$

then it broadcasts a "no handoff" message to notify the MSs nearby and neighboring APs that it does not accept any handoff requests from the MSs currently associated to other APs.

Each MS measures the link gain to its nearby APs at the beginning of each frame

and calculates

$$\tilde{U}_{i'm}(t) = \frac{R_{i'm}(t)T_{i'}}{\overline{X}_m(t)}, \tag{2.30}$$

which is the utility that can be achieved if AP i' transmits to MS m for T_i in the frame t. The MS then keeps a set $\mathcal{H}_m$, which includes all the nearby APs that can provide $\tilde{U}_{i'm}(t)$ larger than its currently associated AP. That is,

$$i' \in \mathcal{H}_m, \text{ if } \tilde{U}_{i'm}(t) > \tilde{U}_{im}(t), \tag{2.31}$$

where AP i is the AP that MS m is currently associated to. If $\mathcal{H}_m$ is empty, the MS keeps associated to its current AP. Otherwise, the MS sorts all the APs in $\mathcal{H}_m$ based on the calculated $\tilde{U}_{i'm}(t)$ values. The first AP in the set has the highest $\tilde{U}_{i'm}(t)$ value, and the last AP in the set has the lowest $\tilde{U}_{i'm}(t)$ value (but still larger than $\tilde{U}_{im}(t)$). The MS listens to the "no-handoff" notifications from APs in $\mathcal{H}_m$. If it does not hear the notification from the first AP in $\mathcal{H}_m$, the MS then attempts to handoff to this AP. Otherwise, it attempts to handoff to the second AP, and so on. If none of the APs in $\mathcal{H}_m$ can accept the MS, the MS stays in the current AP.

Note that an AP may receive handoff requests from multiple MSs that are currently associated to other APs, but it only accepts one such request. Let $\mathcal{M}_{h,i}$ denote a set of the MSs that are requesting to handoff to AP i. Upon receiving these handoff requests, AP i accepts the request of MS m^*, where $m^* = \arg\max_{m \in \mathcal{M}_{h,i}} \tilde{U}_{im}$. That is, MS m^* is the one that AP i can provide with the highest utility among all the requesting MSs.

For comparison, we consider two other straightforward handoff and resource allocation schemes, one based on the number of MSs associated to each AP, and the

other based on the total throughput that can be provided to all the MSs in each frame. The former is referred to as number-based handoff, and the latter is referred to as max-throughput handoff.

2.3.3 Number-based Handoff

Instead of making the handoff decisions based on the utility function, the second handoff criterion simply makes use of the number of MSs associated to each AP. Define

$$\overline{|\mathcal{M}|}_i(t) = \frac{\sum_{j \in \mathcal{N}_i} |\mathcal{M}_j(t)|}{|\mathcal{N}_i(t)|} \tag{2.32}$$

as the average number of the MSs that are associated to the APs near AP i. At the end of each frame, neighboring APs exchange the values of $|\mathcal{M}_i(t)|$. Upon receiving $|\mathcal{M}_j(t)|$'s from all neighboring APs, AP i calculates $\overline{|\mathcal{M}|}_i(t)$. If

$$|\mathcal{M}_i(t)| - \overline{|\mathcal{M}|}_i(t) > 1, \tag{2.33}$$

one MS currently associated to AP i can be handed off to the neighboring APs. Otherwise, no MS should be handed off from this AP, and the AP can accept handoff requests from MSs currently associated to other APs. If (2.33) is true, the MS that should try to hand off from AP i is given by $m^* = \arg\min_{m \in \mathcal{M}_i} g_{im}(t)$, and the destination AP that this MS is handed off to is given by $j^* = \arg\max_{j \in \mathcal{N}_i} g_{jm^*}(t)$. That is, the MS with the worst link gain to AP i tries to handoff to another AP, and the destination AP is the one that has the best link gain to it. After the handoff decisions are made for the MSs, the APs allocate their resources in the same way as in the utility-based scheme by solving the the optimization problem P3 in subsection 2.3.1.

2.3.4 Max-throughput Handoff

This scheme works very similar to the utility-based handoff, except that when choosing the destination AP that MS m should be handed off to, the criterion is

$$i_m^* = \arg\max_{j\in\mathcal{I}} \tilde{R}_{jm}(t). \tag{2.34}$$

After the handoff decisions are made for the MSs, each AP allocates its resources by solving the following optimization problem

$$\text{P4.} \quad \max_{\boldsymbol{\eta}_i(t),\mathbf{T}_i(t)} \sum_{m\in\mathcal{M}_i(t)} \eta_m(t) \tag{2.35}$$

$$\text{s.t.} \quad \eta_m(t) = \tilde{R}_{im}(t)T_{im}(t), \ m\in\mathcal{M}_i(t), \tag{2.36}$$

$$\sum_{m\in\mathcal{M}_i(t)} T_{im}(t) \leq T_i. \tag{2.37}$$

This problem is similar to P3 in subsection 2.3.1 but the objective is to maximize the total throughput of all the MSs associated to it. Overall, the objective of this scheme is to maximize the total throughput of all the MSs at each frame.

2.4 Numerical Results

In this section, we generate numerical results to demonstrate the performance of our proposed scheme. We consider a 1km×1km network service area. There are six APs located at the vertexes of a hexagon. The center of the hexagon is the center of the network service area, the distance between each AP and the center is 300m, and the distance between any two closest APs is 300m. For each AP, only the two APs that are closest to it are its neighbors and can interfere its transmissions. In specific, the

two neighbors of AP i is APs $i-1$ and $i+1$ for $i=2$, 3, 4, and 5, the two neighbors of AP 1 are APs 2 and 6, and that of AP 6 are APs 1 and 5. All APs share the same frequency channel, and neighboring APs cannot transmit at the same time. APs 1, 3, and 5 can transmit at the same time, while APs 2, 4, and 6 are not transmitting; and when APs 2, 4, and 6 transmit, APs 1, 3, and 5 cannot transmit. We specify that during each frame, APs 1, 3, and 5 transmit for half the frame, and APs 2, 4, and 6 transmit for the other half.

Table 2.1: Default simulation parameters

Parameter	Value
Bandwidth W	10 MHz
Noise power P_{noise}	-70 dBm
frame duration	150 ms
T_i	75 ms
P_{AP}	0.5 W
Number of MSs	28
MS maximum moving speed $V_{\max}$	30 m/s

We consider unbalanced MS distributions. A fraction (represented as β, $0 < \beta < 1$) of the MSs move within the circle with radius of 200m and centered at AP 1. The other MSs all move in the entire network service area. All MSs move randomly in their moving area by following the random waypoint model. At the beginning of each frame, the moving speed of an MS is uniformly distributed between 0 and $V_{\max}$, and the moving direction is uniformly distributed between 0 and 2π. The speed and direction are fixed during one frame. The moving speed and direction of different MSs are selected independently. Default parameters are listed in Table 2.1. We collect the throughput and fairness performance in the network. The throughput is the total throughput of all the MSs averaged over the time, and then divided by the total number of APs, and the throughput fairness is evaluated based on Jain's fairness

index (Jain *et al.*, 1984).

We first fix β to 90%. Fig. 2.1 shows the fairness index of different schemes. From the figure we can see that the optimum solution achieves the best fairness, and the fairness performance of the proposed utility-based scheme is close to the optimum. This indicates that the proposed scheme by separating the handoff decisions and the AP resource allocations works well. The max-throughput scheme achieves the lowest fairness. For the max-throughput scheme, each MS is associated to the AP that can provide it with the highest throughput, and each AP allocates the resource to maximize the total throughput of all the MSs associated to it. Therefore, fairness is not considered in both the handoff management and the resource allocations. Due to the uneven geographical distributions of the MSs, MSs located in the low density area (fewer MSs per unit area) can receive much higher throughput than other MSs. Furthermore, when multiple MSs associated to the same AP, the AP always transmits to only the MS with good link conditions. This results in significant unfairness. The number-based scheme attempts to balance the number of MSs associated to different APs. This helps balance the traffic loads among the APs. On the other hand, having an equal number of MSs in two APs does not mean that the APs provide the same total throughput for their associated MSs, because the actual throughput of the MSs also depends on channel conditions. Because of this reason, the fairness of the number-based scheme is between the utility-based and the max-throughput schemes.

In addition, Fig. 2.1 demonstrates the fairness varying with frame of different schemes. In the beginning, all schemes including global optimum solution start from low levels of fairness. As the frame increases, the fairness index of the global optimum solution jumps up to almost 1 within a short period of time while the rest schemes
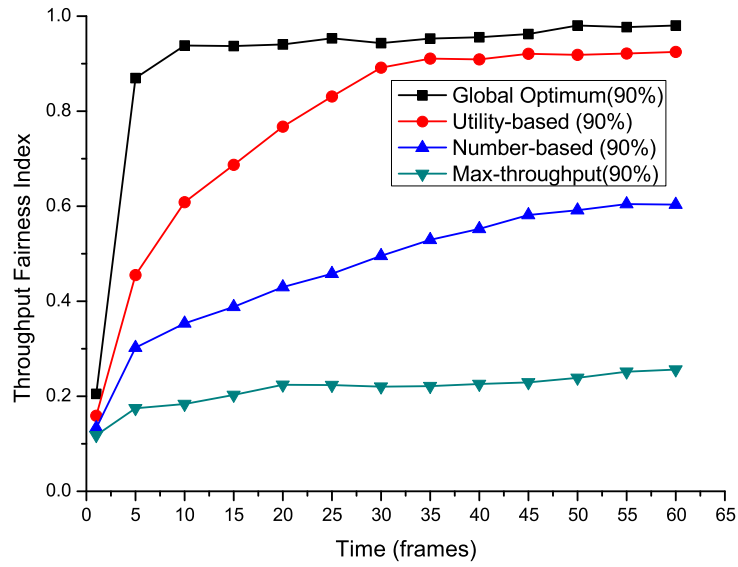
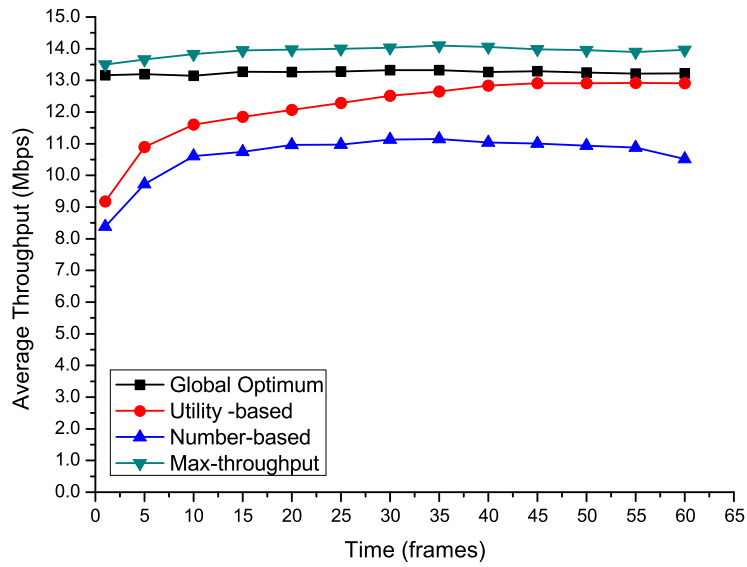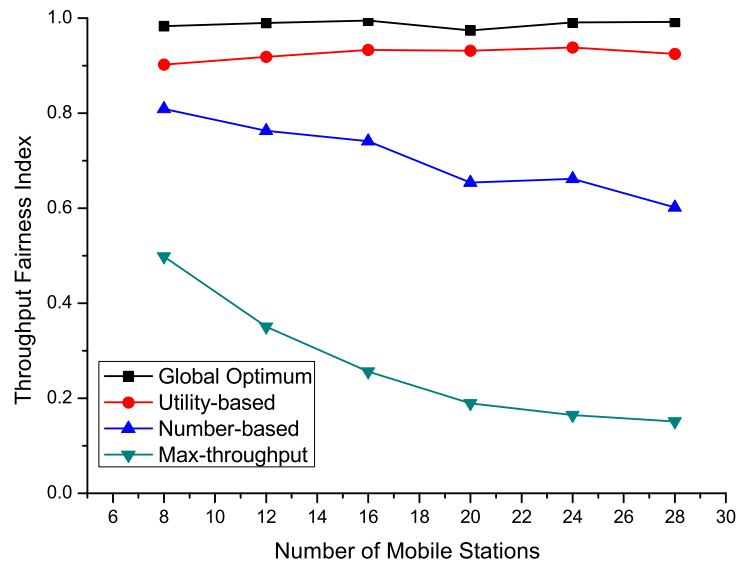Figure 2.1: Fairness index versus number of frames

Figure 2.2: Throughput versus number of frames

gradually rise to their own fairness levels. After long-term iterations, all of them converge and keep steady. Among all heuristic schemes, our proposed utility-based scheme has the shortest convergence time.

Fig. 2.2 illustrates the throughput performance versus frame. It can be seen that there's no significant change of the throughput over the time. The network level average throughput won't change much with the time since each AP selects its best candidate according to certain criterion defined in each scheme.

Fig. 2.3 shows that the fairness performance varies with the number of the MSs in the network. With more MSs, the fairness performance becomes worse. For the max-throughput scheme, having more MSs in the network means better chances for each AP to select the MS to which it can provide higher instantaneous throughput, and this results in worse fairness among the MSs. For the number-based scheme, having more MSs that are unevenly distributed in the network service area results in more unbalanced traffic loads among the APs. The utility-based scheme not only tries to balance the traffic loads among the APs, but also schedules transmissions of the MSs so that they each have their chances to be served, and therefore always achieves close-to-optimum fairness index.

Fig. 2.4 shows the throughput performance. It is seen that the network level throughput is not much affected by the number of MSs. One of the reasons is that the APs always transmit at the same full power, and the total transmission time of each AP during each frame is fixed. Comparing the throughput of different schemes, we find that the optimum solution achieves approximately the same throughput as the max-throughput scheme, while the utility-based heuristic scheme achieves slightly lower throughput. The number-based scheme achieves the lowest throughput among
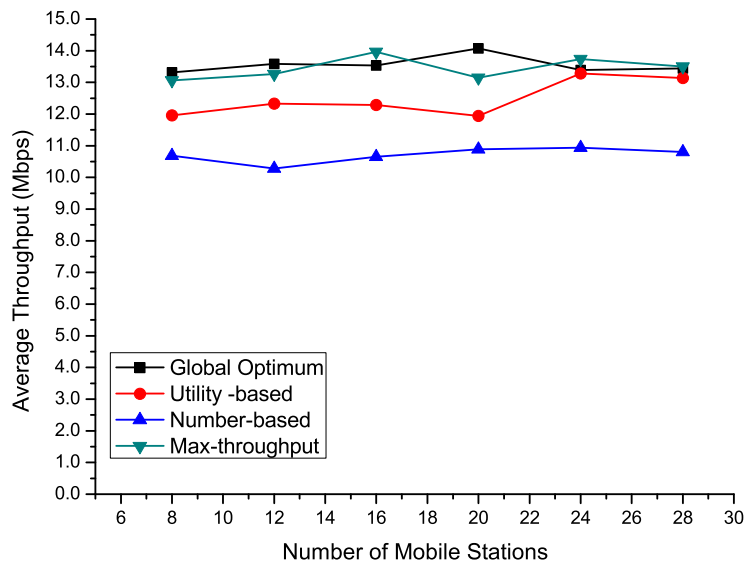
Figure 2.3: Fairness index versus number of MSs

Figure 2.4: Throughput versus number of MSs
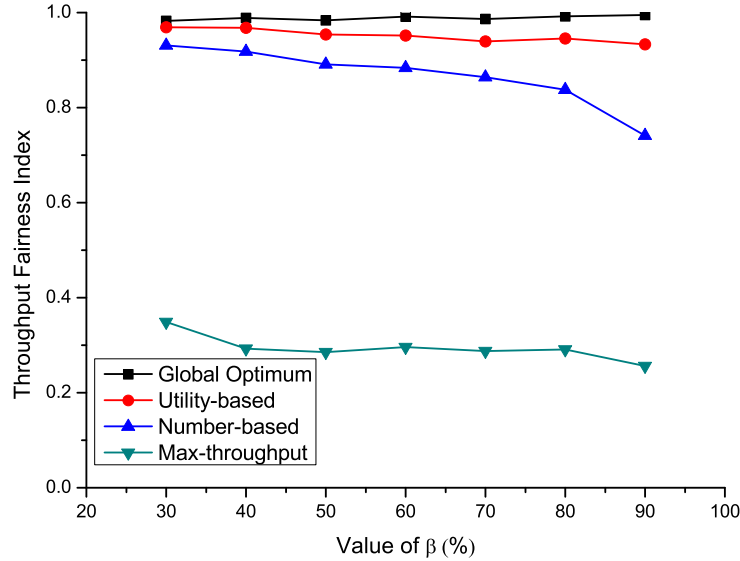
all the solutions.

Figure 2.5: Fairness index versus β

Next, we change the value of β, and compare the fairness and throughput performance of the above schemes. As β increases, more MSs are restricted to move in the close vicinity of AP 1, and the geographical distribution of the MSs becomes less even. Fig. 2.5 shows that the fairness index of the utility-based heuristic scheme is close to the optimum, the fairness index of the number-based scheme is slightly lower and decreases with β, and the max-throughput scheme results in very poor fairness. These results indicate that the proposed utility-based scheme can indeed achieve fair throughput among the MSs with different mobility patterns and geographical distributions. Fig. 2.6 shows the throughput performance when β changes, from which we can see that the utility-based heuristic scheme achieves almost the same average throughput as the optimum scheme, and slightly lower than that of the max-throughput
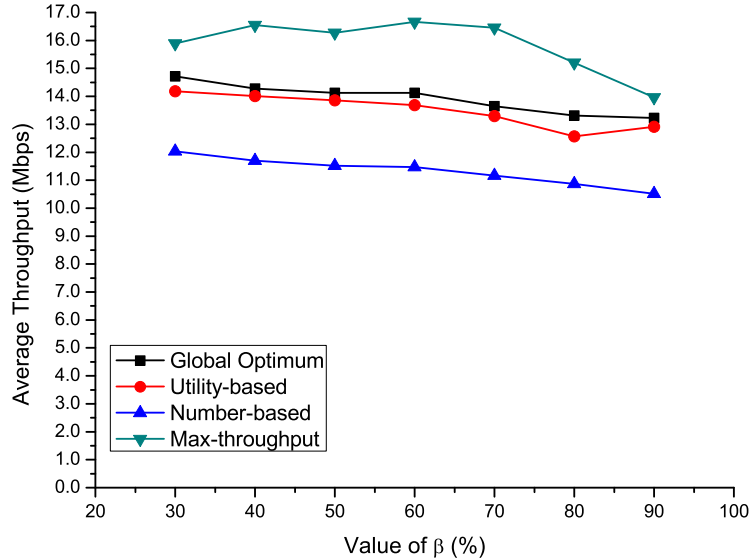
Figure 2.6: Throughput versus β

scheme, while the number-based scheme achieves the lowest throughput.

In summary, the proposed utility-based scheme works well in the sense that its performance is close to the optimum performance in terms of both throughput and fairness. Compared to the max-throughput scheme, it sacrifices the total throughput by about 10% but improves the fairness significantly. Compared to the number-based scheme, it achieves both better fairness and throughput performance.

2.5   Summary

We have studied joint handoff and resource management in wireless mesh networks in order to achieve long-term fair throughput for the MSs. A global optimization problem has been formulated and solved, and a heuristic and distributed scheme has

been proposed. The heuristic scheme is easy to implement, and achieves close-to-optimum performance. In next chapter, we will extend work by coordinating the transmission time of the APs sharing the same frequency channel, depending on the channel and traffic conditions, so that the system level performance is optimized.

Chapter 3

Handoff Management and Channel Time Allocations at both AP and MS Levels

In this chapter we study MS handoff management jointly with AP resource allocations in order to provide fair average throughput to the MSs. We consider a WMN where neighboring APs may share the same frequency channel, and MSs can handoff between different APs as they move around. In particular, different MSs may follow different mobility patterns. We consider both MS handoffs and channel time allocations, where the channel allocations are further performed at two levels, one among the APs and another among the MSs. At the AP level, the problem of channel allocations is to decide the amount of time that each AP can use to serve its associated MSs; and at the MS level, the problem is to decide the amount of time that each of their associated MSs can be served. We first formulate an optimization problem based on long-term PF. The objective is to achieve PF for the long-term throughput among all the MSs in

the network, while achieving high throughput for the system. This optimum solution requires a central station that has the global information about the channel conditions and mobility information of all MSs, and is difficult to achieve in a practical system. We then propose two distributed schemes, HO-CA and CA-HO, by decoupling the MS handoffs and the channel time allocations. The HO-CA scheme makes handoff decisions for the MSs first and then allocates the channel time to the APs. Heuristic handoff decisions are made for the MSs based on their link gains to nearby APs. Given the handoff decisions, an iterative and distributed method is used to find the optimum channel time allocations. The CA-HO scheme allocates the channel time to the APs first and then makes handoff decisions for the MSs. The channel time is first allocated to individual APs based on the neighboring relationship of the APs. After this, each MS estimates possible utilities if it is served by the nearby APs, and then makes handoff decisions. In both schemes, each AP solves a local optimization problem to allocate time for its own associated MSs. Both schemes require neighboring APs to exchange limited information, and can be implemented in practical systems.

The rest of the chapter is organized as follows. In Section 3.1, we describe the system that this work is based on. In Section 3.2, the global optimization problem is formulated, followed by a local optimization problem for each AP to optimize its time allocations to associated MSs. The HO-CA scheme is presented in Section 3.3, and the CA-HO scheme is presented in Section 3.4. Numerical results are demonstrated in Section 3.5, followed by a summary in Section 3.6.

3.1   System Model

We consider the same WMN as described in Section 2.1, except that in this chapter the transmission time of different APs is correlated and should be jointly optimized. In a typical WMN scenario, the locations of the APs may be decided based on convenience or other reasons, and not optimized from the perspective of QoS provisioning and resource utilization. Therefore, there is no guarantee that neighboring APs are assigned to different channels or APs assigned to the same frequency channel do not interfere with each other. To an extreme case, all the APs in the network may share the same frequency channel. Therefore, the APs should coordinate for their transmission time, so that neighboring APs sharing the same frequency channel do not transmit at the same time in order to avoid strong co-channel interference.

Consider that channel time is divided into equal size frames, each of which has a duration T_{frame}. The amount of the time that each AP can transmit during one frame depends on the total number of APs sharing the same channel, the number of the MSs associated to each AP, and other factors. We define *Group of NeighBoring* APs, or GNB. Each GNB is a set of APs that satisfy the following conditions: i) all APs in the group are neighbors of one another, ii) all APs in the group share the same frequency channel, and iii) any AP not in the group either is not a neighbors of at least one AP in the group, or does not use the same frequency channel. Let N be the total number of GNBs in the considered WMN, $\mathcal{G}_n$ be a set of the APs in the nth GNB, and $T_i(t)$ be the total amount of time that AP i can transmit during frame t, then we have

$$\sum_{i \in \mathcal{G}_n} T_i(t) \leq T_{\mathrm{frame}}, \tag{3.1}$$

54

for $n = 1, 2, \ldots, N$.

3.2   Optimum Solutions

This section includes two subsections. In Subsection 3.2.1, the global optimum solution that jointly considers handoff decisions and channel time allocations at both the AP and MS levels is found, and the objective is to achieve PF of the long-term throughput among the MSs. This result will be used as the performance benchmark for comparing with the distributed schemes proposed in Sections 3.3 and 3.4. In Subsection 3.2.2, a local optimization problem is formulated for each AP to allocate its own channel time to associated MSs, assuming the channel time allocations at the AP level are given.

3.2.1   Global Optimum Solution

Based on the description in the previous section, the following optimization problem can be formulated to achieve long-term PF of the throughput among the MSs, where $\boldsymbol{\eta}(t) = [\eta_m(t), m \in \mathcal{M}]$, $\mathbf{T}(t) = [T_{im}(t), i \in \mathcal{I}, m \in \mathcal{M}]$, and $\mathbf{A}(t) = [A_{im}(t), i \in \mathcal{I}, m \in \mathcal{M}]$.

$$\text{P1.} \quad \max_{\boldsymbol{\eta}(t), \mathbf{T}(t), \mathbf{A}(t), \mathbf{Y}(t)} \sum_{m \in \mathcal{M}} \frac{\eta_m(t)}{\overline{X}_m(t)} \tag{3.2}$$

$$\text{s.t.} \quad \eta_m(t) = \sum_{i \in \mathcal{I}} Y_{im}(t) R_{im}(t), \; m \in \mathcal{M}, \tag{3.3}$$

$$R_{im}(t) = W \log_2 \left[1 + \frac{P_{\text{AP}} g_{im}(t)}{P_{noise}} \right], \tag{3.4}$$

$$\sum_{i \in \mathcal{G}_n} T_i(t) \leq T_{\text{frame}}, \; n = 1, 2, \ldots, N, \tag{3.5}$$

$$\sum_{m \in \mathcal{M}_i} T_{im}(t) \leq T_i(t), \; i \in \mathcal{I}, \tag{3.6}$$

$$T_{im}(t) - 10^8(1 - A_{im}(t)) \leq Y_{im}(t), m \in \mathcal{M}, i \in \mathcal{I}, \tag{3.7}$$

$$Y_{im}(t) \leq 10^8 A_{im}(t), \ m \in \mathcal{M}, \ i \in \mathcal{I}, \tag{3.8}$$

$$0 \leq Y_{im}(t) \leq T_{im}(t), \ m \in \mathcal{M}, \ i \in \mathcal{I}, \tag{3.9}$$

$$\sum_{i \in \mathcal{I}} A_{im}(t) \leq 1, \ m \in \mathcal{M}, \tag{3.10}$$

$$A_{im}(t) \in \{0, 1\}, \ m \in \mathcal{M}, \ i \in \mathcal{I}. \tag{3.11}$$

Compared to the optimization problem defined in Section 2.2, constraint 3.5 is extra because $T_i(t)$'s become unknon variables in the network considered in this chapter. The problem is a mixed integer linear programming problem, and can be solved using commercial software such as AMPL. Finding the global optimum solution in a practical system is difficult because it requires a central station to collect global information, including link gains from all the APs to all the MSs, to calculate $R_{im}(t)$'s for all i and m and to inform the APs and MSs of the results regarding channel time allocations and MS handoff decisions. Nevertheless, the solution provides a performance benchmark for designing practical schemes.

3.2.2   Optimum Time Allocations at MS Level

If we consider the objective functions in (3.2) as the total utility of the system, each term in the summation is the utility achieved by one MS. Then $\sum_{m \in \mathcal{M}_i(t)} \frac{\eta_m(t)}{\overline{X}_m(t)}$ is the utility achieved for all the MSs associated to AP i at frame t, or we can simply refer it to as the total utility of AP i at frame t. For AP i, given its own available transmission time $T_i(t)$ in a frame and the link gain information between itself and all the associated MSs, the following optimization problem can be formulated and

solved locally in order to find the allocated time for individual MSs,

$$\text{P2.} \quad \max_{\boldsymbol{\eta}_i(t), \mathbf{T}_i(t)} \sum_{m \in \mathcal{M}_i(t)} \frac{\eta_m(t)}{\overline{X}_m(t)} \tag{3.12}$$

$$\text{s.t.} \quad \eta_m(t) = R_{im}(t)T_{im}(t), \ m \in \mathcal{M}_i(t), \tag{3.13}$$

$$\sum_{m \in \mathcal{M}_i(t)} T_{im}(t) \leq T_i(t), \tag{3.14}$$

where $\mathbf{T}_i(t) = [T_{im}(t), m \in \mathcal{M}_i(t)]$, and $\boldsymbol{\eta}_i(t) = [\eta_m(t), m \in \mathcal{M}_i(t)]$. The optimum solution to problem P2 is

$$T_{im}(t) = \begin{cases} T_i(t), & \text{if } m = \arg\max_{m' \in \mathcal{M}_i(t)} \frac{\eta_{m'}(t)}{\overline{X}_{m'}(t)}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.15}$$

That is, at each frame the AP simply allocates all the time to one MS that can achieve the maximum utility. This local optimum solution can be implemented within each AP if $T_i(t)$ and $A_{im}(t)$'s for all $m \in \mathcal{M}_i(t)$ are known to the AP, i.e., channel allocations at the AP level are completed and the MS-AP associations are given.

In the next two sections, distributed schemes will be designed to solve the problems of MS handoffs and channel time allocations. In Section 3.3, we propose a scheme that considers the handoff decisions first and then allocates channel time to the APs. This is referred to as HO-CA. In the CA-HO scheme proposed in Section 3.4, static channel time is first allocated to the APs, based on which the MSs make their handoff decisions. In both schemes, individual APs allocate the channel time to their own associated MSs using the local optimum solution in (3.15).

3.3   HO-CA Scheme

Since the overall objective of the system is to maximize the total utility achieved to all the MSs, each MS should be associated to the AP that can possibly provide it with the maximum utility. In order to decouple the handoff management and channel time allocations, a simple approach to the handoff management is for each MS to associate to the AP with the best link gain. This, although may not be the optimum solution from the system point of view, does provide a good chance for the AP to serve the MS with high transmission rate, which is important towards high utility for the MS. Given the MS-AP associations, the remaining part of this section is to find the optimum channel time allocations.

From the solution to problem P2 in Subsection 3.2.2 we know that the utility for AP i is a nondecreasing function of $T_i(t)$, and we can denote it as $f_i(T_i(t))$. Given $A_{im}(t)$'s, the optimum channel time allocations can be found from the following optimization problem

$$\text{P3.} \quad \min_{\mathbf{T}(t)} \ -\sum_{i \in \mathcal{I}} f_i(T_i(t)) \tag{3.16}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{G}_n} T_i(t) \le T_{\text{frame}}, \ \text{for } n = 1, 2, \ldots, N. \tag{3.17}$$

Note that the function $f_i(T_i(t))$ incorporates the details for optimum channel time allocations at the MS level, given that AP i is allocated to $T_i(t)$. Therefore, solving problem P3 directly still requires a central station that collects all the link gain information between each MS and its associated AP. In the remaining part of this section, we use a decomposition method to solve problem P3 distributively. The basic method is first decomposing the original problem into multiple subproblems, which

are coupled in the constraints. The problem is solved in iterations. In each iteration, the subproblems are solved in parallel with the information passed from the coupling variables until the solutions converge.

When decomposing the original problem into subproblems, there is a tradeoff between the complexity for solving individual subproblems and the number of iterations for the solutions to converge. In general, if the number of subproblems is large, less complexity is required for solving each subproblem, but more iterations may be required for the solutions to converge. The decomposition for problem P3 is based on GNBs. The original problem is decomposed into N subproblems, each of which is formulated for one GNB. Within each GNB, the APs are all neighbors of each other, and can exchange the link gain information and find the optimum solution for the subproblem of the GNB. Meanwhile, the solutions among the GNBs are coupled because some APs belong to multiple GNBs, and these APs can help find out whether the solutions have converged.

We now describe the subproblem formulation. In order to have concise expressions, we drop the t variables. This should not cause any confusion because all the descriptions are for a given frame. Since the APs belonging to multiple GNBs play important roles in this distributed solution, we define K_i as the total number of GNBs that AP i belongs to. We use K_i fake APs (FAPs) to replace AP i, so that each of the FAPs belongs to only one of the GNBs that AP i belongs to. We define a number of FAP sets as $\mathcal{F}_1 = [1, K_1]$, and $\mathcal{F}_i = [\sum_{i'=1}^{i-1} K_{i'} + 1, \sum_{i'=1}^{i} K_{i'}]$ for $i \in \mathcal{I}$ and $i > 1$, and then use j as the index for the FAPs. Then all FAPs with $j \in \mathcal{F}_i$ are used to replace AP i. Define the allocated time for FAP j as Z_j, then $Z_j = T_i$ if $j \in \mathcal{F}_i$. The total utility for AP i is equal to a sum of the utilities of all its FAPs. Define $h_j(Z_j)$

as the total utility of FAP j. We have

$$f_i(T_i) = \sum_{j \in \mathcal{F}_i} h_j(Z_j). \tag{3.18}$$

Exactly how the FAPs in $\mathcal{F}_i$ should share the utility (or MSs) of AP i is not important to the proposed method. One example is that all the FAPs have an equal share of the total utility of the AP. That is,

$$h_j(Z_j) = \frac{1}{K_i} f_i(T_i) \tag{3.19}$$

for $j \in \mathcal{F}_i$. Other relationship between $f_i(T_i)$ and $h_j(Z_j)$ also works, as long as (3.18) is satisfied. The optimization problem P3 can then be rewritten as

$$\text{P4.} \quad \min_{\mathbf{Z},\mathbf{T}} -\sum_j h_j(Z_j) \tag{3.20}$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{G}_n} Z_j \leq T_{\text{frame}}, \ n = 1, 2, \ldots, N, \tag{3.21}$$

$$Z_j = T_i, \ i \in \mathcal{I} \text{ and } j \in \mathcal{F}_i. \tag{3.22}$$

At this point, we distinguish the APs with $K_i = 1$ from those with $K_i > 1$. The APs with $K_i = 1$ are referred to as local APs, because they each belong to only one GNB; while the APs with $K_i > 1$ each belong to more than one GNB and are referred to as net-APs. Without loss of generality, we assume that the first M^{loc} APs are net-APs, and the remaining APs are local APs. Define a column vector $\mathbf{T}^{net} = \{T_i, \ i \in [1, M^{loc}]\}$. The FAPs replacing the local and net-APs are referred to as local and net-FAPs, respectively. We define a set $\mathcal{J}^{net}$ to include all the net-FAPs, and set $\mathcal{J}^{loc}$ to include all the local FAPs. Furthermore, define sets $\mathcal{J}_n^{loc}$ and $\mathcal{J}_n^{net}$,

respectively, to include the local and net-FAPs in the nth GNB.

Now we define several new variables so that problem P4 can be written in a concise matrix form. We define column vectors $\mathbf{Z}_n^{loc} = \{Z_j, \ j \in \mathcal{J}_n^{loc}\}$, and $\mathbf{Z}_n^{net} = \{Z_j, \ j \in \mathcal{J}_n^{net}\}$. Define p_n as the total number of net-FAPs in the nth GNB, then $\mathbf{Z}_n^{net}$ has p_n elements. We further define a column vector $\mathbf{Z}^{net} = [\mathbf{Z}_1^{net}; \ldots; \mathbf{Z}_N^{net}]$. The total number of elements in $\mathbf{Z}^{net}$ is given by $p = \sum_{n=1}^{N} p_n$. A $p_n \times M^{loc}$ matrix $\mathbf{E}_n$ is defined for the nth GNB as follows

$$
E_{n,ji} =
\begin{cases}
1, & \text{if } Z_j = T_i, \ j \in \mathcal{J}_n^{net}, \ i \in [1, M^{loc}] \\
0, & \text{otherwise},
\end{cases}
\tag{3.23}
$$

then $\mathbf{E} = [\mathbf{E}_1; \mathbf{E}_2; \ldots; \mathbf{E}_N]$ is a $p \times M^{loc}$ matrix. The constraints defined by (3.22) can be rewritten in a concise format as

$$
\mathbf{Z}^{net} = \mathbf{E}\mathbf{T}^{net},
\tag{3.24}
$$

which can be partitioned into N sub-conditions, each for one GNB, and the condition for the nth one is given by

$$
\mathbf{Z}_n^{net} = \mathbf{E}_n \mathbf{T}^{net}.
\tag{3.25}
$$

Note that the N sub-conditions defined by (3.25) are correlated because they all share the same vector $\mathbf{T}^{net}$. Physically, this is because that FAPs belonging to different GNBs may correspond to the same AP.

Define $H(\mathbf{Z}^{loc}, \mathbf{Z}^{net}) = -\sum_j h_j(Z_j)$, and $G(\mathbf{Z}_n^{loc}, \mathbf{Z}_n^{net}) = \sum_{j \in \mathcal{G}_n} Z_j$, problem P4

can be written in a matrix form as

$$\text{P4.1} \quad \min_{\mathbf{Z}^{loc}, \mathbf{Z}^{net}, \mathbf{T}} \quad H(\mathbf{Z}^{loc}, \mathbf{Z}^{net}) \tag{3.26}$$

$$\text{s.t.} \quad G(\mathbf{Z}_n^{loc}, \mathbf{Z}_n^{net}) \leq T_{\text{frame}}, \ n = 1, 2, \ldots, N. \tag{3.27}$$

$$\mathbf{Z}^{net} = \mathbf{E}\mathbf{T}^{net}. \tag{3.28}$$

The partial Lagrangian of the above problem can be written as

$$\text{P5.} \quad \min_{\mathbf{Z}^{loc}, \mathbf{Z}^{net}, \mathbf{T}} \quad H(\mathbf{Z}^{loc}, \mathbf{Z}^{net}) + \boldsymbol{\lambda}^T(\mathbf{Z}^{net} - \mathbf{E}\mathbf{T}^{net}) \tag{3.29}$$

$$\text{s.t.} \quad G(\mathbf{Z}_n^{loc}, \mathbf{Z}_n^{net}) \leq T_{\text{frame}}, \ n = 1, 2, \ldots, N, \tag{3.30}$$

where $\boldsymbol{\lambda}$ is a column vector with p elements, and (3.29) can be further rewritten as

$$\min_{\mathbf{Z}^{loc}, \mathbf{Z}^{net}, \mathbf{T}} \sum_{n=1}^{N} \left[H(\mathbf{Z}_n^{loc}, \mathbf{Z}_n^{net}) + \boldsymbol{\lambda}_n^T \mathbf{Z}^{net} \right] - \boldsymbol{\lambda}^T \mathbf{E}\mathbf{T}^{net}, \tag{3.31}$$

where $\boldsymbol{\lambda}_n$ is a column vector including the p_n elements corresponding to the nth GNB in $\boldsymbol{\lambda}$.

Next, we use the dual decomposition method in (S. Boyd and Mattingley, 2008) to solve the problem. To find dual function, we should first minimize over $\mathbf{T}^{net}$, which results in $\boldsymbol{\lambda}^T \mathbf{E} = 0$. With this condition, the problem can be decomposed into N subproblems, each for one GNB. For the n-th GNB, the subproblem is given as

$$\text{P6.} \quad \min_{\mathbf{Z}_n^{loc}, \mathbf{Z}_n^{net}} \quad H(\mathbf{Z}_n^{loc}, \mathbf{Z}_n^{net}) + \boldsymbol{\lambda}_n^T \mathbf{Z}^{net} \tag{3.32}$$

$$\text{s.t.} \quad G(\mathbf{Z}_n^{loc}, \mathbf{Z}_n^{net}) \leq T_{\text{frame}}. \tag{3.33}$$

Note that the N subproblems defined by P6 (for $n = 1, 2, \ldots, N$) are not independent. Constraint (3.28), or equivalently, constraint (3.25) for $n = 1, 2, \ldots, N$, couples the subproblems, and specifies that all the FAPs replacing the same AP should have the same Z_j value. Therefore, in the subproblems defined by P6, each element in $\mathbf{Z}_n^{loc}$ is a local variable to only one subproblem, and each element in $\mathbf{Z}_n^{net}$ is a public variable that connects different subproblems.

Based on subsection 4.4 in (S. Boyd and Mattingley, 2008), the following algorithm can be used to solve the dual problem of the original problem with dual gap of zero.

0. Choose initial vector $\boldsymbol{\lambda}$ that satisfies $\mathbf{E}^T \boldsymbol{\lambda} = 0$ and repeat the following steps until the process converges.

1. Solve problem P6 within individual GNBs and obtain $\mathbf{Z}_n^{loc}$ and $\mathbf{Z}_n^{net}$.

2. Find $\hat{\mathbf{T}}^{net} = \left(\mathbf{E}_n^T \mathbf{E}_n\right)^{-1} \mathbf{E}_n^T \mathbf{Z}_n^{net}$ for each GNB, where $\hat{\mathbf{T}}^{net}$ is an estimate of $\mathbf{T}^{net}$.

3. Update $\boldsymbol{\lambda}_n = \boldsymbol{\lambda}_n + \alpha_k (\mathbf{Z}_n^{net} - \mathbf{E}_n \hat{\mathbf{T}}^{net})$, where α_k is an appropriate step that can be chosen in any of the standard ways (S. Boyd and Mattingley, 2008).

When implementing the above algorithm in a real system, we assume that matrix $\mathbf{E}$ is known to all the APs. Initially, individual FAPs can calculate their own λ_j values (λ_j is the jth element in $\boldsymbol{\lambda}$). FAPs within the same GNB exchange necessary information in order to solve the sub-problem P6 for their own GNBs, and the optimization problems are solved in parallel in different GNBs. Updating $\hat{\mathbf{T}}^{net}$ in Step 2 and $\boldsymbol{\lambda}_n$ in Step 3 is also done within each GNB. After this, new solution is obtained within each GNB. At the end of each iteration, net-AP i obtains the solutions from all the GNBs that it belongs, and compares the values of Z_j's of its FAPs. If they are sufficiently close, the solution converges. Note that in each iteration, in order to solve

P6, FAP j should solve problem P2 by assuming that $T_i = Z_j$ in the current frame and find $f_i(T_i)$, and then find $h_j(Z_j)$ based on the relationship between $h_j(Z_j)$ and $f_i(T_i)$ in (3.19). From this process we can see that this algorithm does not require interactions among APs belonging to different GNBs, and the scheme does not require a central controller. When the solutions converge, the time allocations are completed at both the AP and the MS levels.

3.4   CA-HO Scheme

In the previous scheme, most of the efforts are used for finding the channel time allocations at the AP level. The distributed and iterative method can be time consuming, especially when the number of APs is large. In this section, we propose a relatively static method for channel time allocations at the AP level, based on which MSs make handoff decisions in order to maximize the total utility at each frame.

First, the condition in (3.1) should be satisfied when considering the channel time allocations at the AP level. This constraint basically indicates that the neighboring relationship is important to the time allocated to different APs. For example, allocating time to the APs having more neighboring APs is more critical than to those having fewer or no neighboring APs. We consider to allocate the channel time at the AP level based on the relatively fixed neighboring relationship of them. Within each frame, an AP having more neighboring APs is assigned to less time so that its neighboring APs can be assigned to more time. In this way, the cumulative available channel transmission time for all the APs can be larger, and the overall system throughput can be higher. With this motivation, we choose to solve the following

optimization problem in order to find T_i's.

$$\text{P7.} \quad \max_{\mathbf{T}} \sum_{i \in \mathcal{I}} \log T_i \tag{3.34}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{G}_n} T_i \leq T_{\text{frame}}, \ \ n = 1, 2, \ldots, N, \tag{3.35}$$

where the logarithm objective function together with the constraint allows high cumulative transmission time of the APs, while preventing zero time allocated to some APs, which may be resulted from some other objective functions such as $\max_{\mathbf{T}} \sum_{i \in \mathcal{I}} T_i$. Assume the neighboring relationship is known to all the APs, the above problem can be solved by each AP offline, which then broadcasts its T_i value to nearby MSs.

Given the available time for each AP, the utility-based handoff scheme presented in subsection 2.3.2 can be used for MSs to make handoff decisions.

3.5 Numerical Results

In this section, we generate numerical results to demonstrate the performance of the proposed schemes. We consider that all the APs in the network service area share the same frequency channel. Two APs are neighbors if their distance is less than a certain distance D. We consider uneven MS distributions. A fraction (represented as β, $0 < \beta < 1$) of the MSs move within a circle that is centered at a given AP and has a radius of 200m, and the other MSs move in the entire network service area. All MSs move randomly in their movement areas by following the random waypoint model. At the beginning of each frame, the moving speed of an MS is uniformly distributed between 0 and V_{max}, and the moving direction is uniformly distributed between 0 and 2π. The speed and direction are fixed during one frame. The moving

speed and direction of different MSs are selected independently. Below we consider different scenarios of AP locations and MS movement patterns, and collect the average throughput and fairness performance in each scenario. The average throughput is the total throughput of all the MSs divided by the total number of APs, and the throughput fairness is evaluated based on average throughput of individual MSs using the Jain's fairness index (Jain *et al.*, 1984). For comparison, we also plot the results for a "max-throughput" scheme, which always maximizes the total throughput of all the MSs. The results for this scheme are obtained by solving the optimization problem P1 in Section 3.2 but changing the objective function to maximize $\sum_m \eta_m(t)$ in each frame t. Default parameters are listed in Table 3.1.

Table 3.1: Default simulation parameters

Parameter	Value
Bandwidth W	10 MHz
Noise power P_{noise}	-70 dBm
frame duration	150 ms
P_{AP}	0.5 W
Number of MSs	28
MS maximum moving speed V_{max}	30 m/s
Maximum distance between neighboring APs D	300m

Scenario 1: We consider a network, where six APs are equally spaced in a line as shown in the upper part of Fig. 3.1. The distance between any two neighboring APs is 200m. At any time, at most three APs can transmit simultaneously. We first fix $\beta = 30\%$, force a fraction of β MSs to move around AP 2, and collect the results of fairness index and average throughput by varying number of MSs. Fig. 3.2 shows the fairness index of different schemes, from which we can see that the optimum PF solution achieves perfect fairness, the fairness index of both the proposed distributed schemes is very close to 1, while the HO-CA scheme achieves slightly better fairness
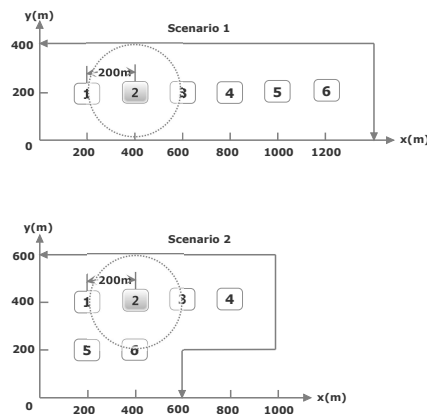
Figure 3.1: AP positions in scenarios 1 and 2

than the CA-HO scheme. These results indicate that the proposed schemes that decouple the MS handoff and channel time allocations work well in terms of achieving fair average throughput among the MSs. In the HO-CA scheme, the link gain-based handoff decisions may result in different number of MSs associated to different APs, but the optimum channel time allocations can compensate this effect, and allocate more time to the APs associated with more MSs. Similarly, in the CA-HO scheme, the static channel time allocations at the AP level do not adapt to traffic changes, but the handoff decisions made by the MSs based on the available channel time of individual APs can mitigate some of this effect and achieve good fairness for the average throughput among the MSs. In addition, Fig. 3.2 shows that the fairness index of the max-throughput scheme is much below 1, and can be worse when the number of the MSs is large.

The throughput performance of different schemes is shown in Fig. 3.3. The max-throughput scheme achieves the highest throughput at a price of poor fairness. The throughput performance of the optimum PF solution is slightly lower. This is due to the tradeoff between throughput and fairness among the MSs. The figure shows

some gap between the throughput performance of the distributed schemes and the optimum PF solution, and this is due to the effect of decoupling the channel time allocations and handoff decisions in the proposed schemes.

Overall, we see that the HO-CA scheme achieves both better fairness and higher throughput than the CA-HO scheme. In the HO-CA scheme, although the handoff decisions are made heuristically, the channel time is allocated in an optimum way once the handoff decisions are made; on the other hand, in the CA-HO scheme the channel time at the AP level is allocated statically without considering the MS locations, and following that the handoff decisions are made again heuristically. Therefore, overall the HO-CA scheme can better deal with the uneven traffic load distributions than the CA-HO scheme, and decoupling the channel time allocations and handoff decisions results in more performance degradation to the CA-HO scheme than to the HO-CA scheme. This affects both the fairness and throughput of the schemes.

We then change β to 90% to create a scenario of more uneven MS distributions in the network, and the fairness and throughput performance are shown in Figs. 3.4 and 3.5, respectively. It is seen that the fairness index of the proposed distributed schemes is still close to the optimum, and the throughput performance in this case is not much different from that when $\beta = 30\%$. This demonstrates that the proposed distributed schemes work well for highly unevenly distributed MSs.

Next, we evaluate the fairness and throughput performance of the schemes by varying β. As β increases, more MSs are restricted to move in the close vicinity of one AP, and the geographical distribution of the MSs becomes more uneven. Fig. 3.6 shows that the fairness index of the proposed schemes is still very close to 1, although slightly decreases with β. Meanwhile, Fig. 3.7 shows that the throughput performance
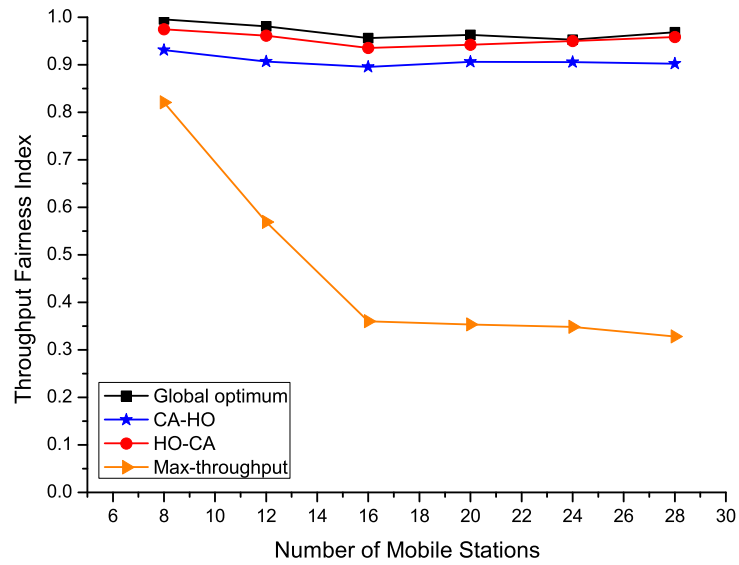
Figure 3.2: Scenario 1, Fairness index versus number of MSs ($\beta = 30\%$)
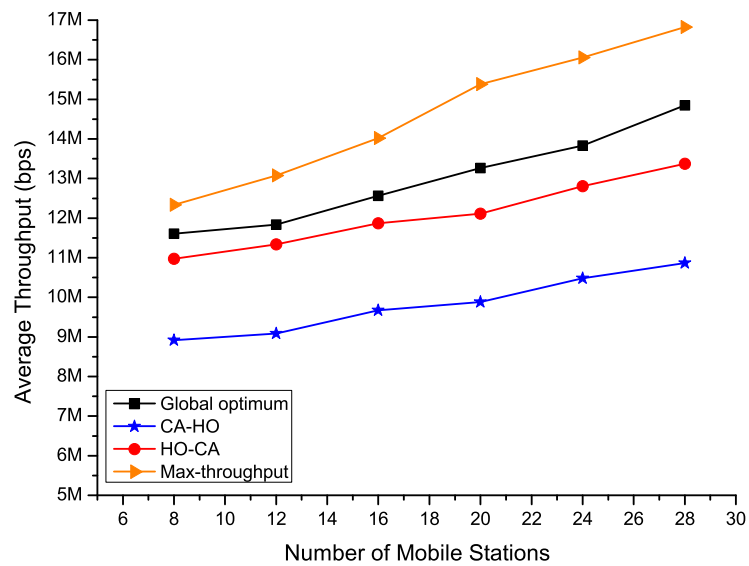
Figure 3.3: Scenario 1, Throughput versus number of MSs ($\beta = 30\%$)
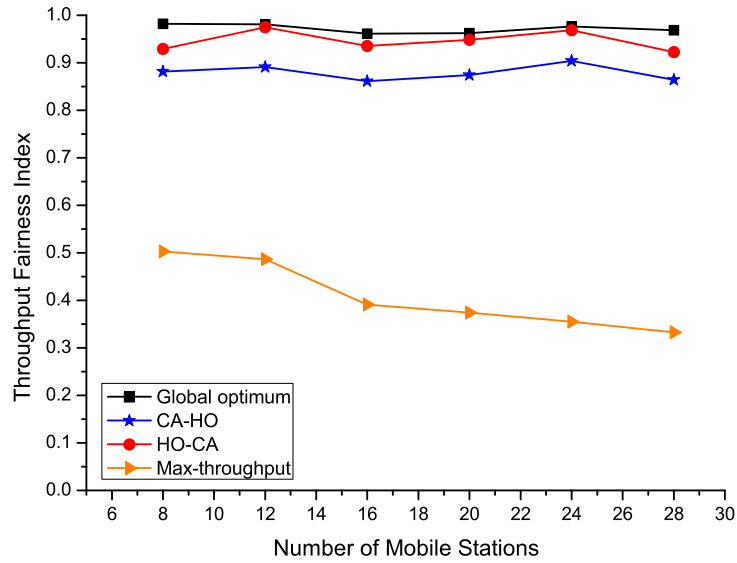
Figure 3.4: Scenario 1, Fairness index versus number of MSs ($\beta = 90\%$)
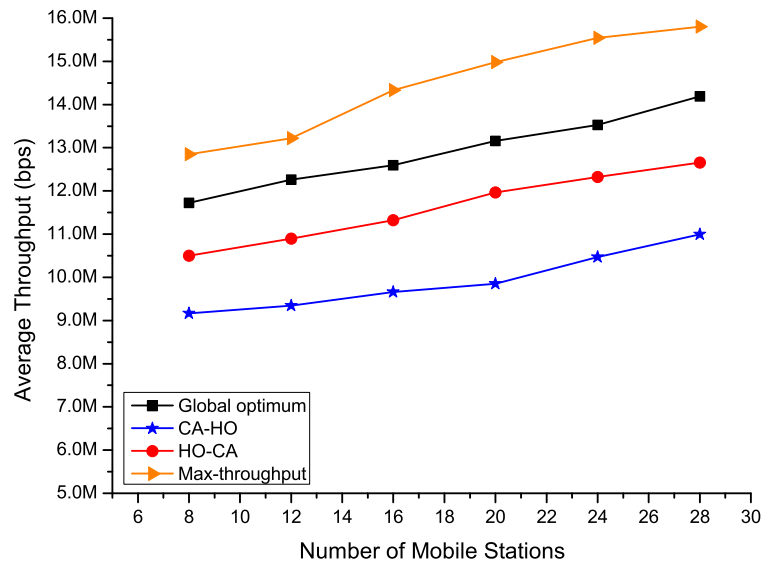
Figure 3.5: Scenario 1, Throughput versus number of MSs ($\beta = 90\%$)

of all the proposed schemes is only slightly affected by β. These results indicate that the proposed distributed schemes work well for highly unevenly distributed MSs. The schemes not only provide almost equally good fairness for different levels of unevenly distributed MSs, but also keep approximately the same level of throughput regardless of the MS location distributions.
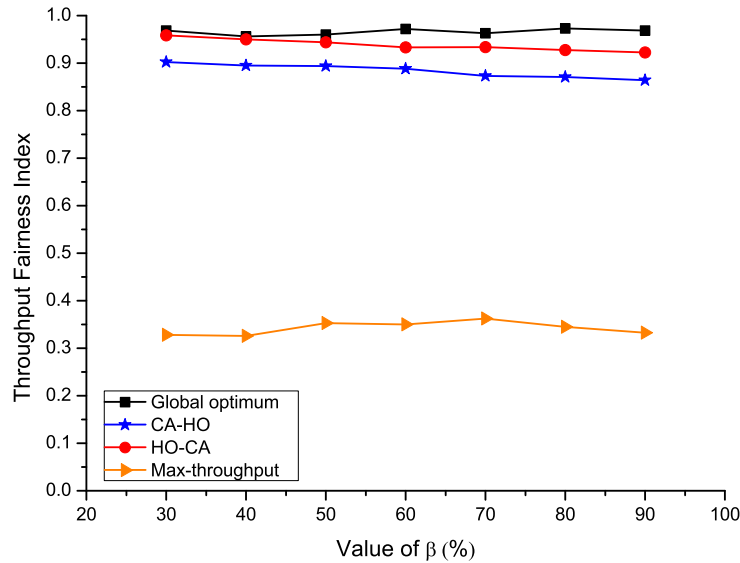
Figure 3.6: Scenario 1, Fairness index versus β

Scenario 2: In the second scenario, APs 1-4 are equally spaced along one line with distance of 200m in between, and APs 5 and 6 are located in another parallel line with distance of 200m in between. The distances between the two lines is also 200m. This topology is shown in the lower part of Fig. 3.1. We first have a fraction of 30% MSs move around AP 2 and the remaining MSs move in the entire network service area. Compared to scenario 1, AP 2 in scenario 2 has more neighboring APs, and therefore, resource allocations at AP 2 have more significant effect on the

71
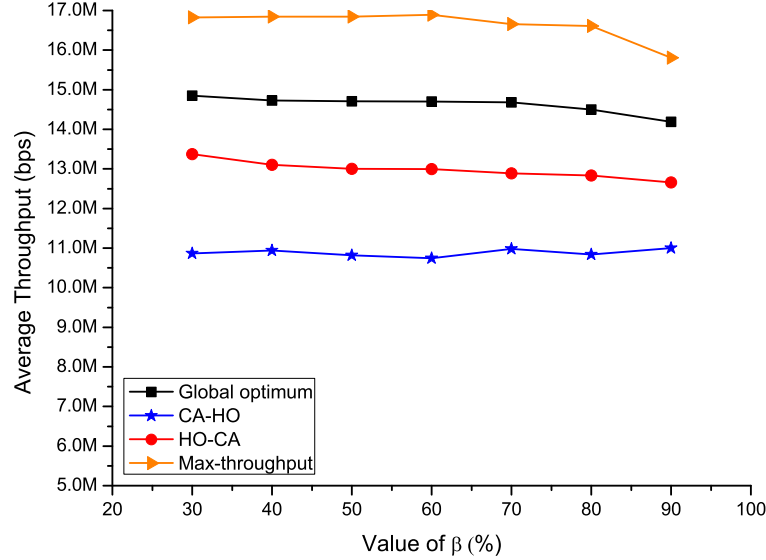
Figure 3.7: Scenario 1, Throughput versus β

overall system performance in this scenario than in scenario 1. Figs. 3.8 and 3.9 show the fairness index and throughput performance, respectively. Although the fairness index for both proposed distributed schemes in scenario 2 is still quite close to 1, by comparing Figs. 3.2 and 3.8 we can see that the fairness index of both schemes in this scenario is slightly lower than that in scenario 1, and the drop for the CA-HO scheme is slightly larger than that for the HO-CA scheme. Fig. 3.9 also shows that throughput performance of the HO-CA scheme in this scenario is close to the optimum, while the throughput of the CA-HO scheme drops quite significantly. By comparing Figs. 3.9 and 3.3, we can see that the throughput (of both optimum and heuristic) in this scenario is lower than that in scenario 1.

When a large number of MSs gather around one AP, this AP is overloaded. In order to provide fair throughput for all the MSs, the overloaded AP should be allocated

to more channel time, and as a result, less time is available for the neighboring APs. If the overloaded AP has more neighboring APs, then the cumulative transmission time for all the APs in the network becomes shorter, and the overall throughput of the network is reduced. This explains that the optimum throughput in scenario 2 is lower than in scenario 1. In terms of fairness, the irregular topology in scenario 2 makes it more difficult to provide close-to-perfect fairness, compared to scenario 1. This does not affect the HO-CA scheme very significantly, because the optimum channel time allocations at both the AP and MS levels can reduce the effect of non-optimum handoff decisions. On the other hand, in the CA-HO scheme, since the overloaded AP has more neighboring APs, it is always allocated to relatively shorter channel time, which affects both the throughput and fairness of the scheme negatively.
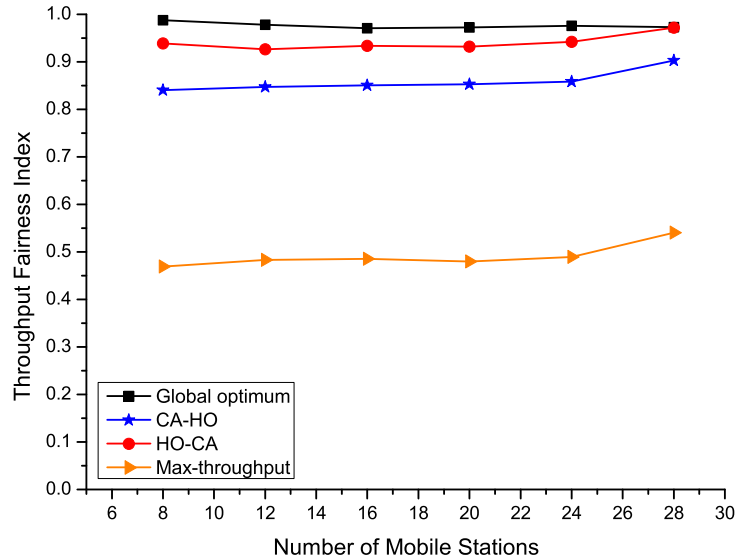
Figure 3.8: Scenario 2, Fairness index versus number of MSs ($\beta = 30\%$)

We then change β to 90% and collect the fairness index and throughput results,
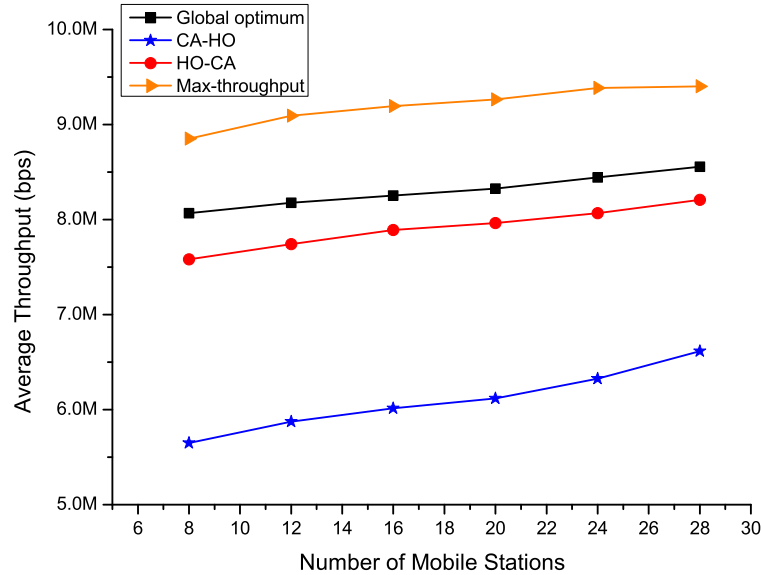
Figure 3.9: Scenario 2, Throughput versus number of MSs ($\beta = 30\%$)

which are shown in Figs. 3.10 and 3.11. Comparing with the results for $\beta = 30\%$, neither the throughput or the fairness index has significant changes. This further indicates that the proposed schemes are not much sensitive to the uneven distribution of the MSs.

Next, we change the value of β and collect the fairness and throughput performance in scenario 2. Figs. 3.12 and 3.13 show the fairness and throughput performance, respectively. The results further indicate that the performance is not very much sensitive to the uneven distribution of the MSs. The fairness index of both the schemes decreases slightly with β, but still much higher than that of the max-throughput scheme. The throughput for the HO-CA scheme is close to the optimum throughput, and almost not affected by changing β, while the throughput of the CA-HO scheme decreases slightly with β. This again demonstrates that the HO-CA
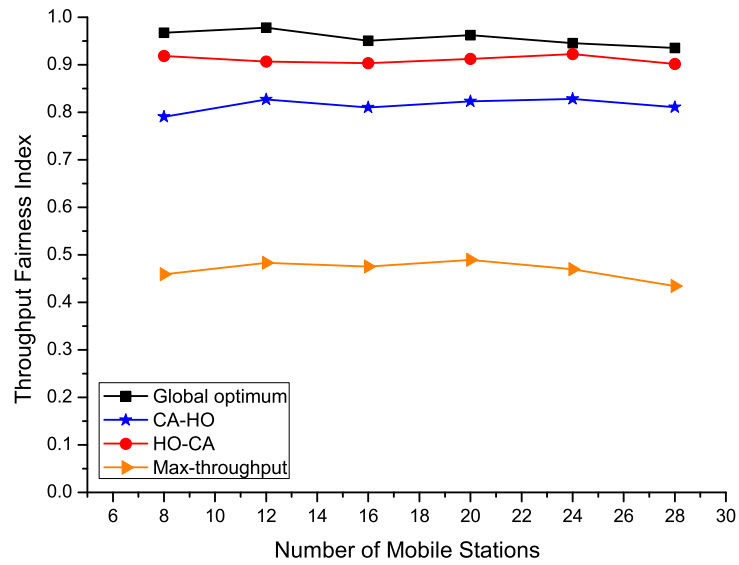
Figure 3.10: Scenario 2, Fairness index versus number of MSs ($\beta = 90\%$)
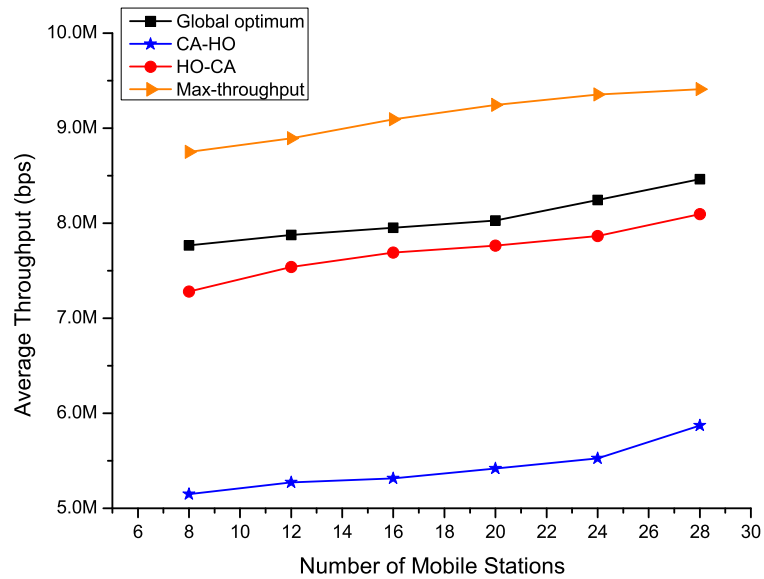
Figure 3.11: Scenario 2, Throughput versus number of MSs ($\beta = 90\%$)

scheme is more robust to the changes to MS location distributions than the CA-HO scheme.
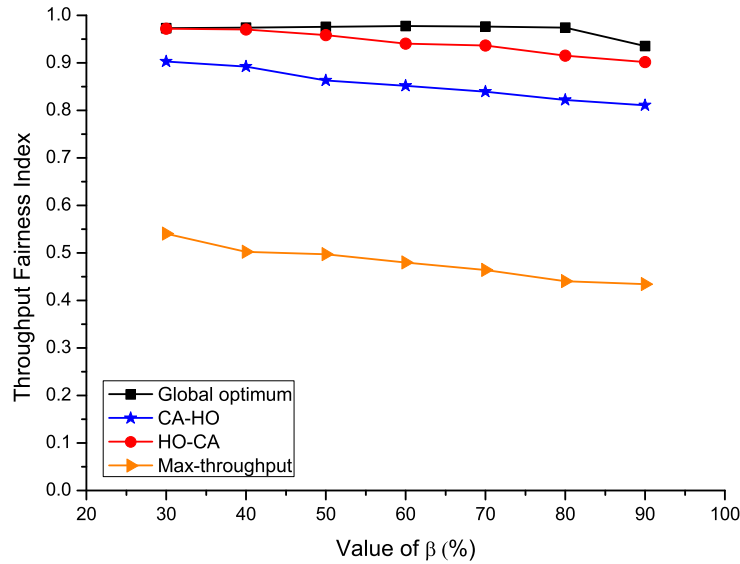
Figure 3.12: Scenario 2, Fairness index versus β

Scenario 3: In this scenario we keep the same geographical locations of the APs as in scenario 2, but have a fraction of the MSs move around AP 1 instead of AP 2. As AP 1 does not have as many neighboring APs as AP 2, the overload condition in scenario 3 may have a different effect on the network performance as in scenario 2. We first fix $\beta = 30\%$ and change the number of MSs, Figs. 3.14 and 3.15 show the fairness and throughput performance, respectively. We then fix the number of MSs and change β, Figs. 3.18 and 3.19, respectively, show the fairness and throughput performance. For comparison, performance results of the proposed schemes in scenario 2 are also shown in these figures. From Figs. 3.14 and 3.18 we can see that the fairness index of both the proposed schemes in this scenario is lower than in scenario 2, although

76
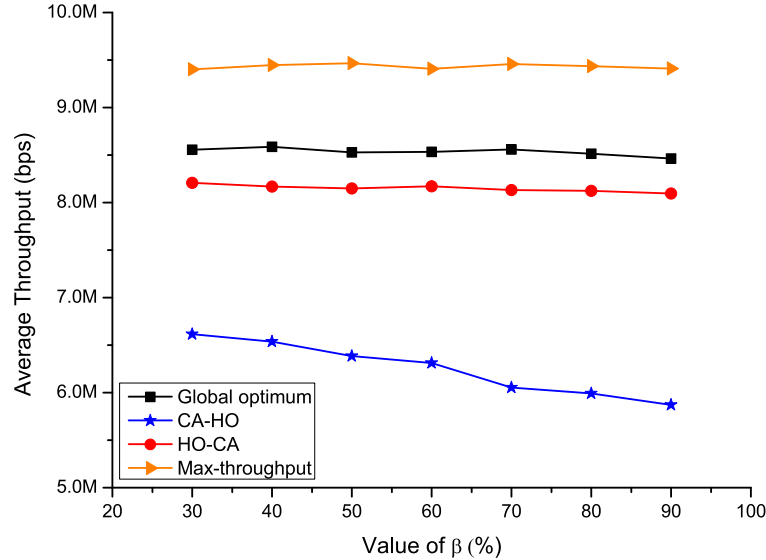
Figure 3.13: Scenario 2, Throughput versus β

still much higher than that of the max-throughput scheme. This is because that the overloaded AP (AP 1) has fewer neighbors, which limits both the channel time adjustment between APs and handoff choices of the MSs. As a result, the fairness index can be poor. As shown in Fig. 3.18, the fairness index of both the proposed schemes drop to about 0.6 when $\beta = 90\%$.

Figs. 3.15 and 3.19 show that the throughput of the HO-CA scheme in scenario 3 is almost the same as in scenario 2, while the throughput of the CA-HO scheme in scenario 3 is slightly higher than in scenario 2. The main reason for this is that, in the HO-CA scheme, the channel time is allocated adaptively based on the traffic load requests in each AP. The amount of time allocated to each AP and each MS is optimized to the current MS-AP associations. Therefore, the scheme has more flexibility to adapt to the network condition changes without significantly sacrificing

the throughput performance. In the CA-HO scheme, the channel time allocated to each AP is determined by the relative locations of the APs, and is the same in scenarios 2 and 3. In scenario 2, the overloaded AP (AP 2) has more neighboring APs, and this allows more MSs in the overloaded area to connect to other APs in order to achieve better fairness. As a result, the throughput is sacrificed because more MSs have to be served by APs that do not have the best link gains to them. While in scenario 3, MSs in the overloaded area do not have as many handoff choices as in scenario 2. Although these MSs are served with lower average throughput, other MSs in non-overloaded areas can be served with higher throughput. Therefore, the overall throughput in scenario 3 can be higher than in scenario 2.
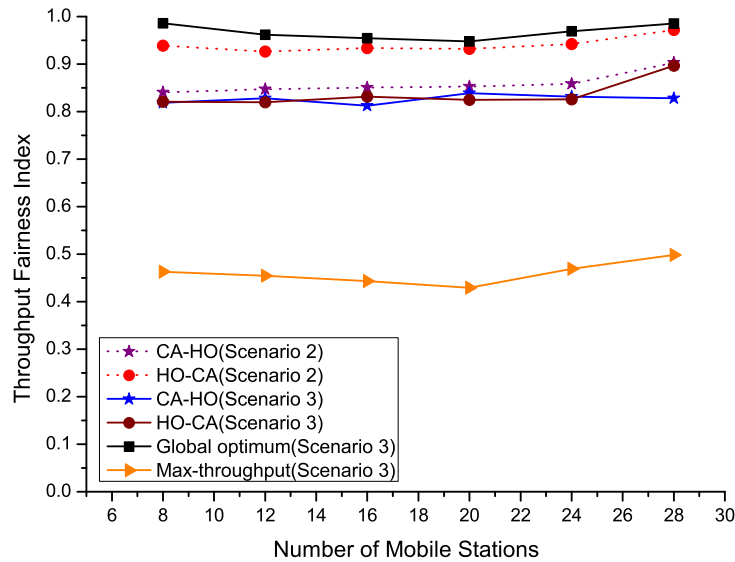
Figure 3.14: Scenario 3, Fairness index versus number of MSs ($\beta = 30\%$)

We then change β to 90% and collect the fairness index and throughput results, which are shown in Figs. 3.16 and 3.17. The observations from these figures are
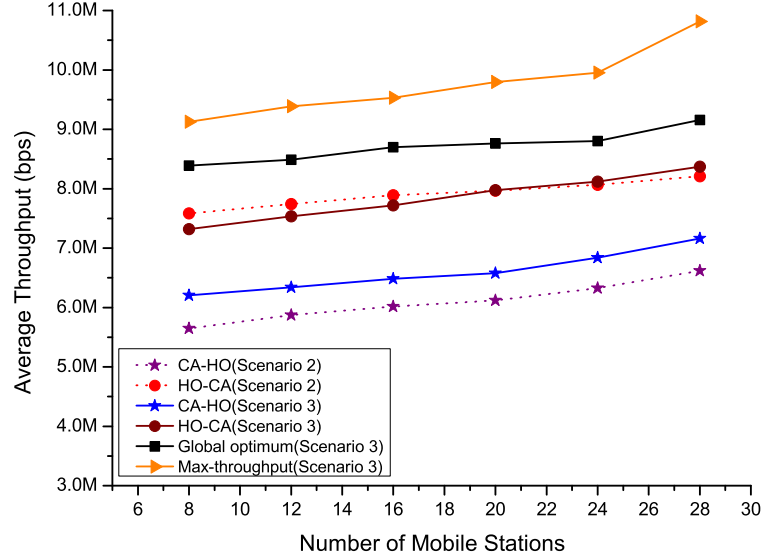
Figure 3.15: Scenario 3, Throughput versus number of MSs ($\beta = 30\%$)

consistent with that in Figs. 3.14 and 3.15.

Scenario 4: In this scenario we consider random AP locations. 6 APs are uniformly distributed in a square service area of 1km×1km. A fraction of β MSs move around one randomly picked AP, and all the other MSs move in the entire network service area. We simulate a large number of such random topologies and average the performance over these topologies. The results are shown in Figs. 3.20 and 3.21 for fairness index and throughput, respectively, when $\beta = 30\%$. In addition, Figs. 3.24 and 3.25 show the fairness and throughput performance, respectively, for different values of β when the number of MSs is fixed. The results indicate that both the proposed schemes achieve very good fairness index, which is close to the optimum and much better than that of the max-throughput scheme. Regarding the throughput performance, the HO-CA scheme achieves higher throughput than the CA-HO
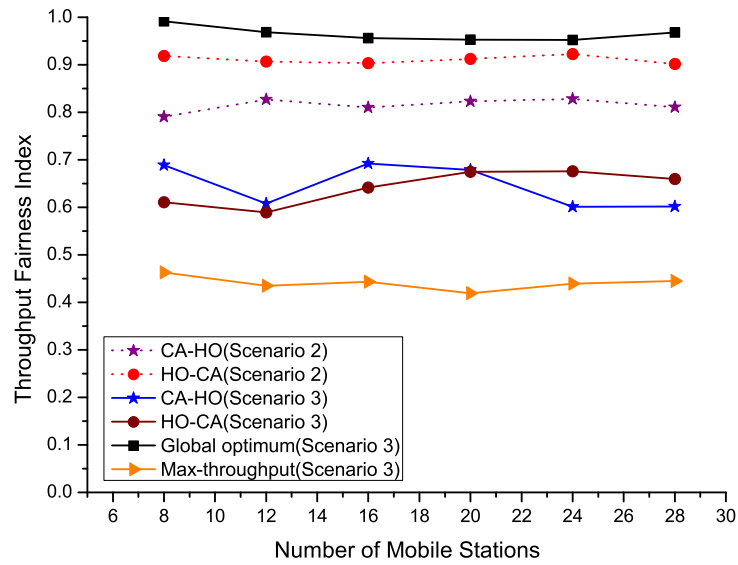
Figure 3.16: Scenario 3, Fairness index versus number of MSs ($\beta = 90\%$)
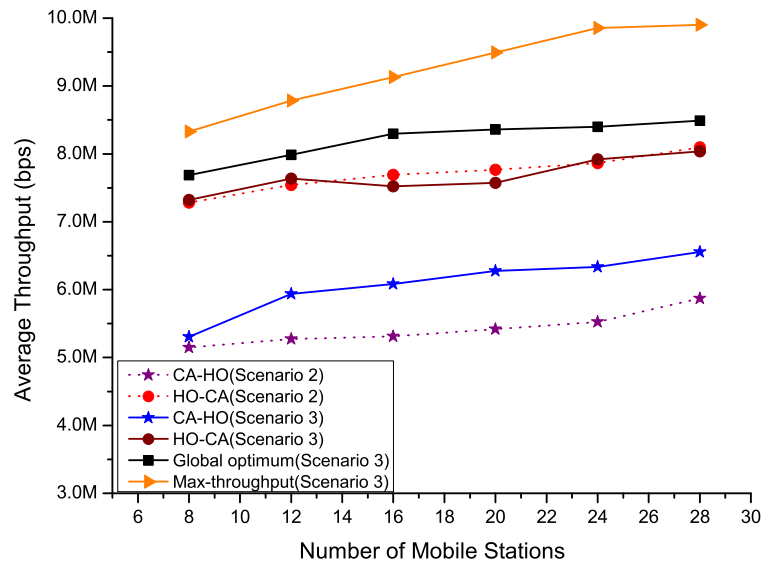
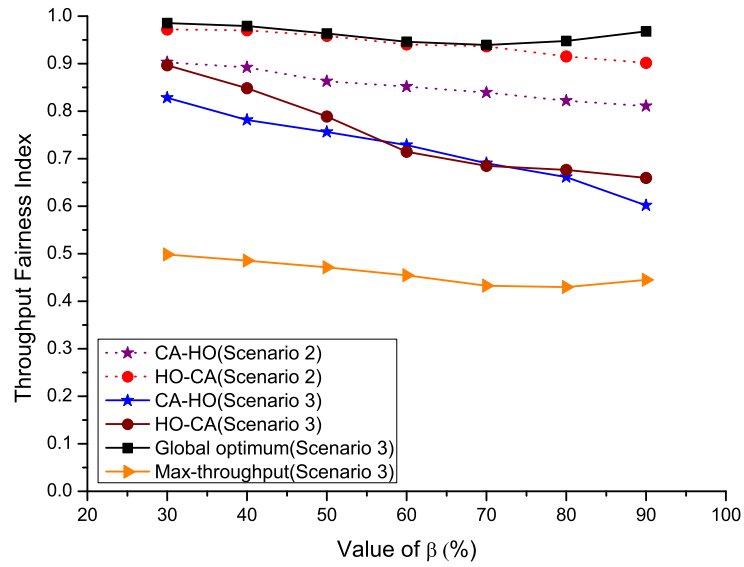Figure 3.17: Scenario 3, Throughput versus number of MSs ($\beta = 90\%$)

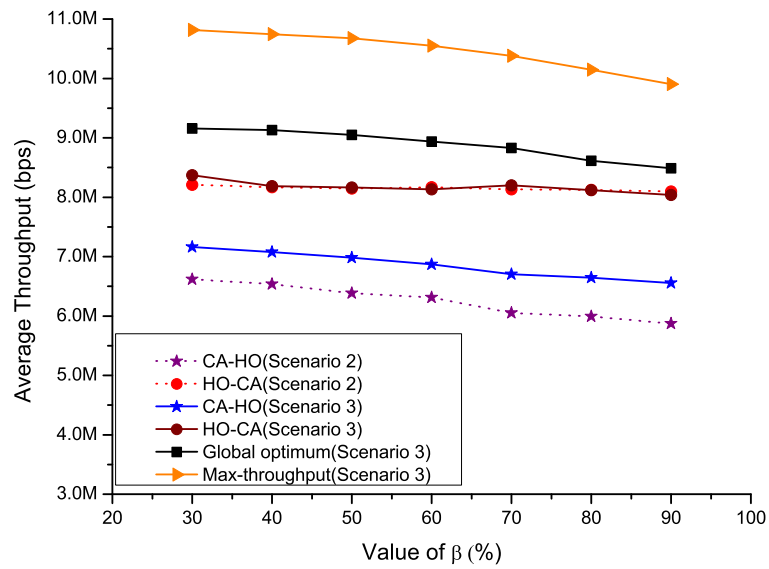Figure 3.18: Scenario 3, Fairness index versus β

Figure 3.19: Scenario 3, Throughput versus β

scheme for the whole range of β. While both the throughput and fairness index drop with β, the changes are very slight. Overall, the observations in this random topology are consistent with that in the previous scenarios.
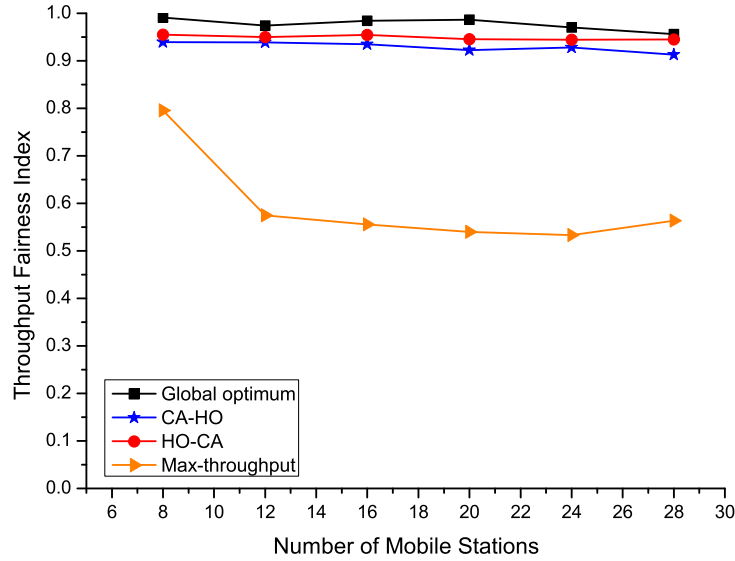
Figure 3.20: Scenario 4, Fairness index versus number of MSs ($\beta = 30\%$)

Scenario 5: This scenario is the same as scenario 4, except that 8 APs are randomly deployed in the network service area of 1km×1km. A fraction of β MSs move around one randomly picked AP, and all the other MSs move in the entire network service area. We simulate a large number of such random topologies and average the fairness index and throughput over these topologies. The results are shown in Figs. 3.26 and 3.27 for $\beta = 30\%$ and in Figs. 3.28 and 3.29 for $\beta = 90\%$. In addition, Figs. 3.30 and 3.31, respectively, show the fairness and throughput performance for different values of β when the number of MSs is fixed. Basically, both the fairness and throughput performance of the proposed schemes in this scenario are consistent
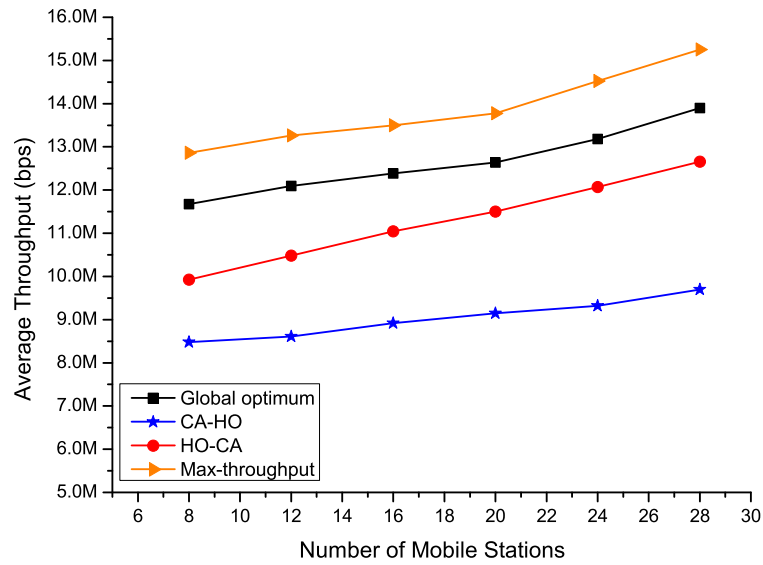
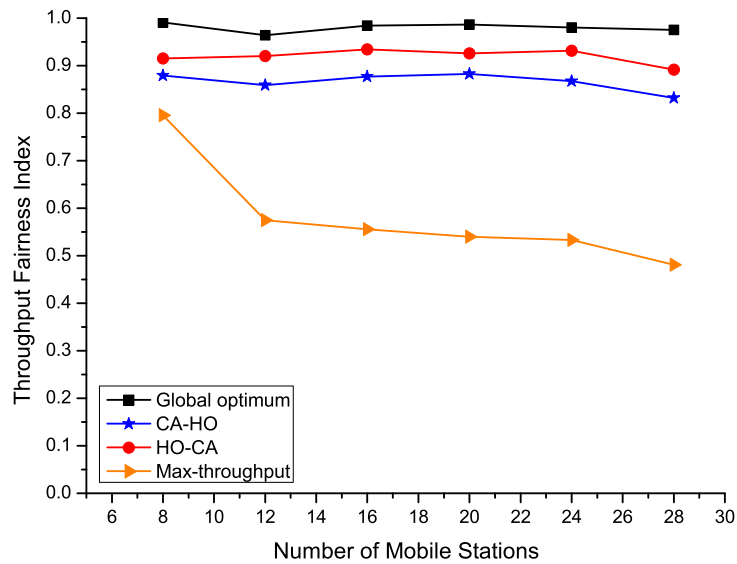Figure 3.21: Scenario 4, Throughput versus number of MSs ($\beta = 30\%$)

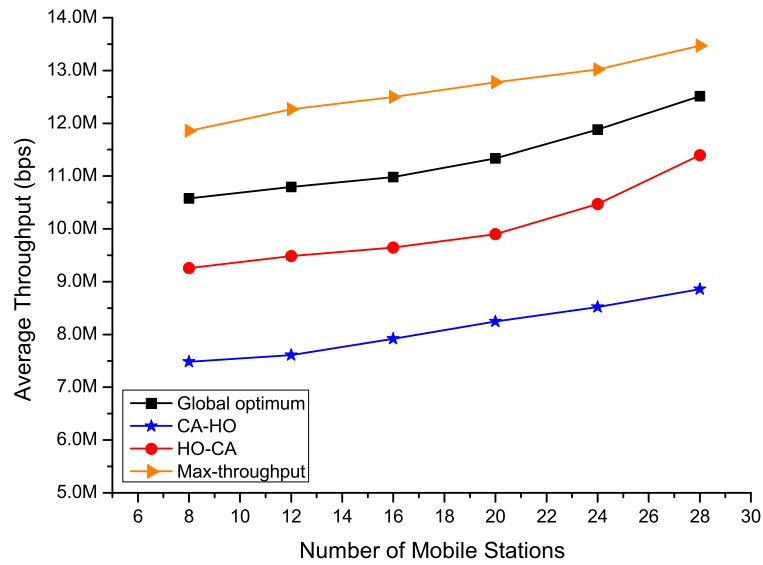Figure 3.22: Scenario 4, Fairness index versus number of MSs ($\beta = 90\%$)

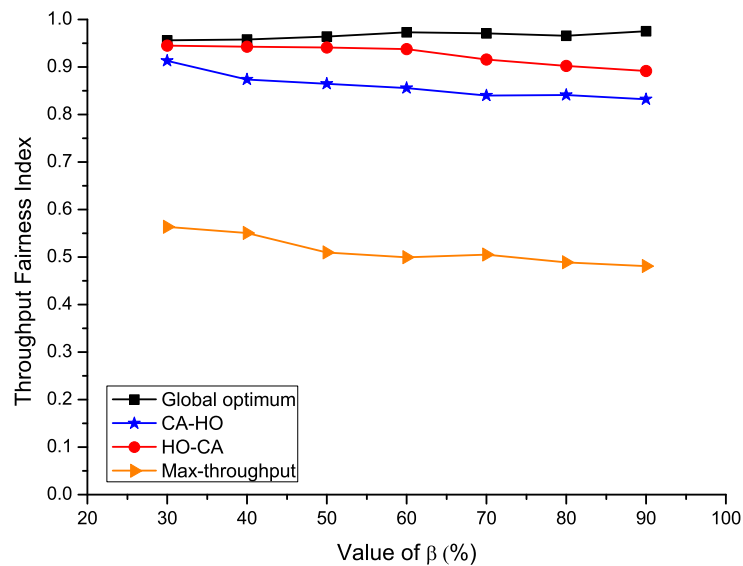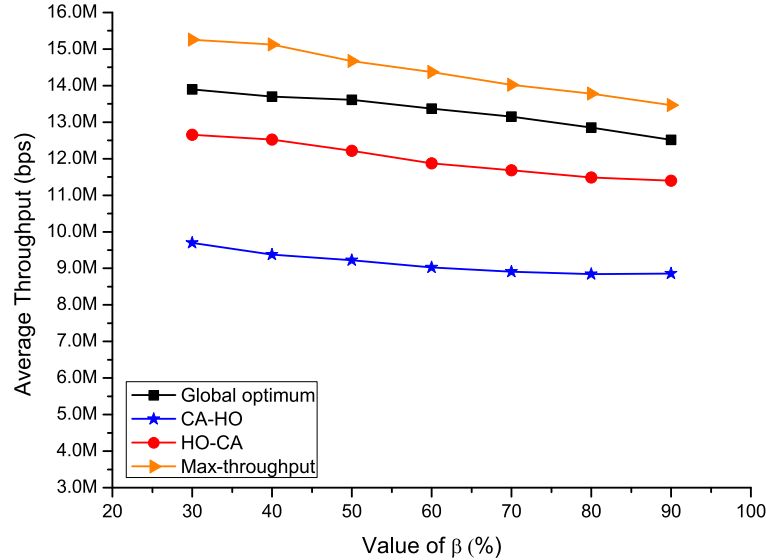Figure 3.23: Scenario 4, Throughput versus number of MSs ($\beta = 90\%$)

Figure 3.24: Scenario 4, Fairness index versus β

84

Figure 3.25: Scenario 4, Throughput versus β

with the previous scenario. Compared to the results in scenario 4, the throughput in this scenario is slightly higher due to more APs.

Scenario 6: In this scenario, we collect and analyze the results while the number of APs varies. APs are randomly deployed in the network service area of 1km×1km. A fraction of β MSs move around one randomly picked AP, and all the other MSs move in the entire network service area. For each case with different number of APs, We simulate a large number of such random topologies and average the fairness index and throughput over these topologies. The results are shown in Figs. 3.32 and 3.33 for fairness index and throughput performance, respectively. For the fairness index, the observations from Fig. 3.32 are basically consistent with the previous results. That is, the fairness index of the proposed schemes is close to optimum, and much higher than that achieved by the max-throughput scheme. In addition, the HO-CA scheme
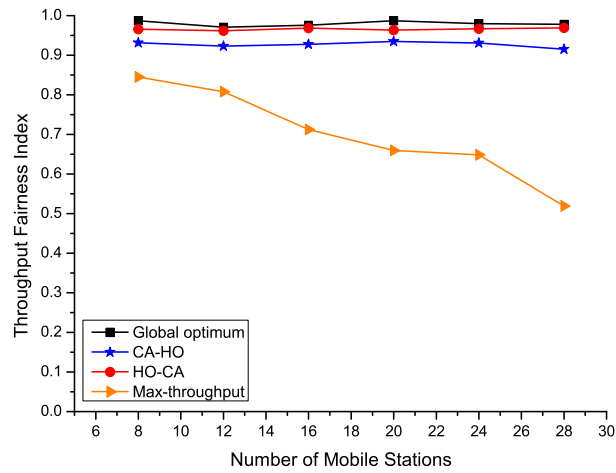
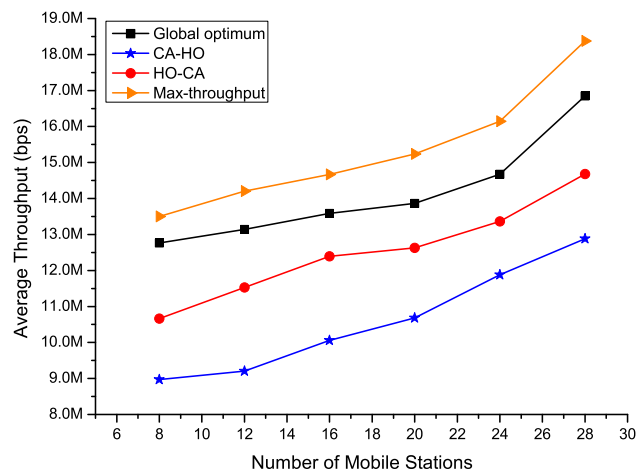Figure 3.26: Scenario 5, Fairness index versus number of MSs ($\beta = 30\%$)

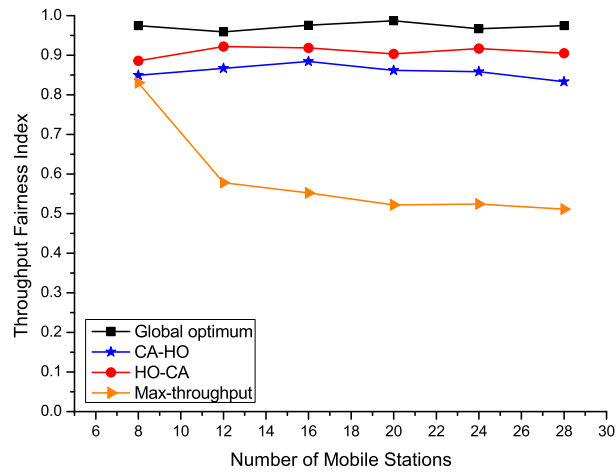Figure 3.27: Scenario 5, Throughput versus number of MSs ($\beta = 30\%$)

Figure 3.28: Scenario 5, Fairness index versus number of MSs ($\beta = 90\%$)
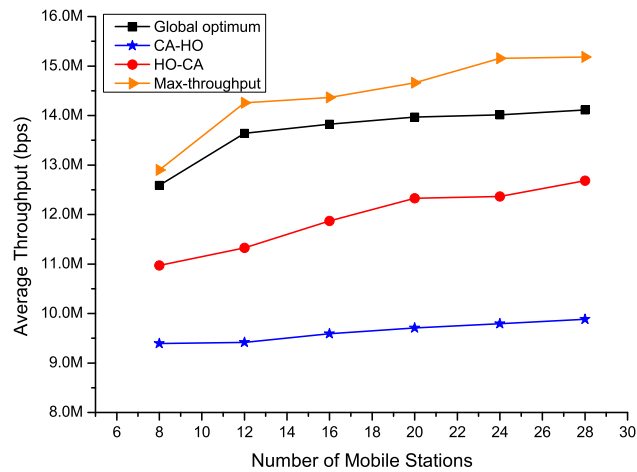

Figure 3.29: Scenario 5, Throughput versus number of MSs ($\beta = 90\%$)
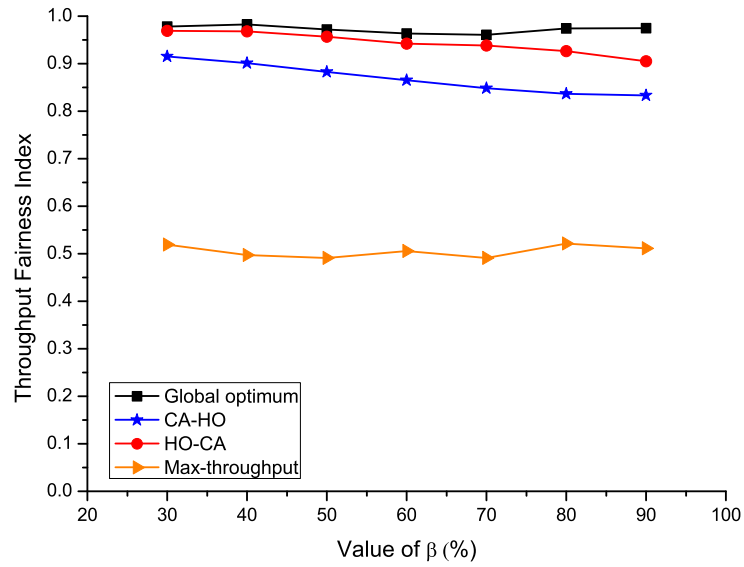
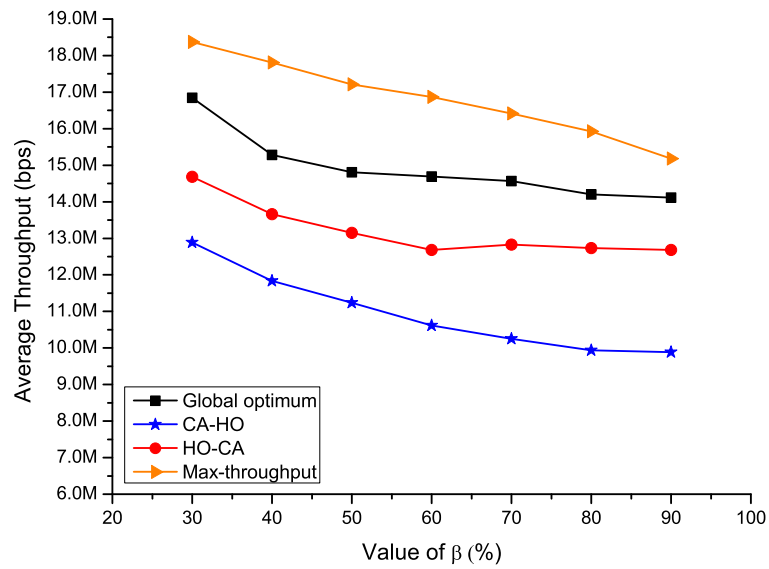Figure 3.30: Scenario 5, Fairness index versus β

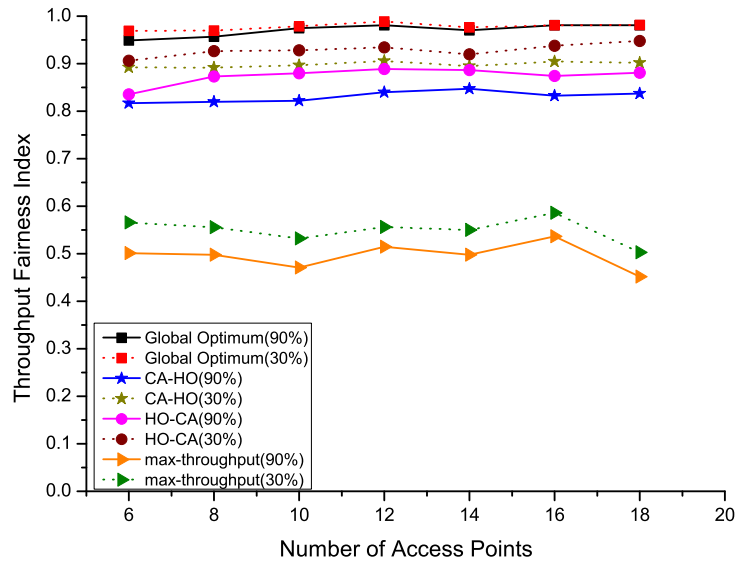Figure 3.31: Scenario 5, Throughput versus β

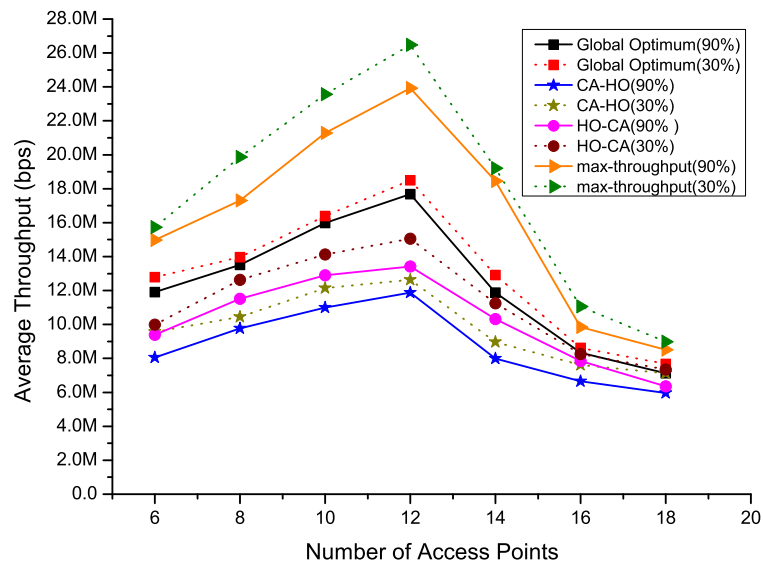Figure 3.32: Scenario 6, Fairness index versus number of APs

Figure 3.33: Scenario 6, Throughput versus number of APs

achieves slightly better fairness than the CA-HO scheme; and for each scheme, the fairness index at $\beta = 30\%$ is slightly higher than that at $\beta = 90\%$.

For the throughput performance, Fig. 3.33 shows that the HO-CA scheme achieves slightly higher throughput than the CA-HO scheme; and for each scheme, the throughput at $\beta = 30\%$ is slightly higher than that at $\beta = 90\%$. These observations are all consistent with the previous scenarios. In addition, as the number of APs increases, the throughput increases and then decreases. This is due to some contradictory effects of the number of APs on the throughput performance. On one hand, having more APs provides the MSs more chances to handoff to the APs that can seve them with high transmission rate, which increases the throughput. On the other hand, as the number of APs increases, more APs become neighbors of one another, and more APs have time conflicts, which reduces the available transmission time for each AP and decreases the overall throughput of the network. When the number of APs is relatively small, the first effect dominates, and the throughput increases with the number of APs; and as the number of APs further increases, the second effect dominates, and the throughput decreases with the number of APs.

3.6   Summary

We have proposed two distributed schemes for channel allocations and handoff management in wireless mesh networks in order to provide fair throughput for mobile stations. Our results have demonstrated that the proposed schemes can provide good fairness even when the location distributions of the MSs are highly uneven in the network service area.

Chapter 4

Conclusions and Future Work

In this thesis we have studied how to provide fair throughput for the MSs while efficiently utilizing the channel resources in a WMN, where handoff management and channel timeline allocations are jointly considered. The objective is to achieve long-term PF for the average throughput among the MSs and in the meanwhile improve the network resource utilization. We considered two different systems. In the first system, the channel time allocations to the APs are fixed, and handoff management of the MSs and channel time allocations at the MS level are jointly studied; and in the second system, handoff management of the MSs and channel time allocations at both the AP and MS levels are jointly studied. For each system, the optimum solution is obtained, and heuristic and practical schemes are designed. Our results have demonstrated that the proposed schemes can effectively balance the throughput provided to different MSs through both handoff management and channel time allocations. The proposed heuristic schemes can indeed achieve close-to-optimum fairness and very good resource utilization.

The proposed schemes are important to wireless mesh networks for supporting mobile stations with uneven geographical distributions and different mobility patterns. More importantly, the idea of designing distributed resource management schemes can be extended to design similar resource management strategies in other wireless networks that do not have a central controller. This work can bet further extended to better support different QoS of the MSs and improve the radio resource utilization.

The current work considers best effort data traffic only, which is delay insensitive and does not have a minimum throughput requirement. The work can be extended to manage the MS handoffs and allocate the channel time in order to satisfy the minimum throughput requirements of some users while providing fair throughput for the remaining users. In addition, the real-time performance of the proposed distributed scheme can be further studied. The proposed schemes can be modified to give different priorities to the MSs in order to provide services with different urgency.

In this work, the transmission rates are adaptively changed based on the current transmission conditions. This makes it flexible for MSs to handoff to different APs in order to balance the traffic loads and achieve better throughput fairness. In order to simplify the implementation, fixed transmission power is assumed for all the APs. The work can be further extended to consider both adaptive transmission power and transmission rates. By allowing APs to adjust their transmission power, co-channel interference can be better controlled, and the resource utilization can be further improved. However, the complexity for designing a close-to-optimum resource management scheme becomes more challenging.

The channel time allocations within an AP are optimized by assuming that perfect link gain formation is known to each AP for all its associated MSs. In a practical

system, this information may not be accurate. For example, there may be errors at the MSs when measuring the link gains, or the signaling transmissions can experience errors when the MSs report the measured link gains to the APs. This should be taken into consideration in making handoff decisions and resource allocations in order to make the schemes more practical.

In the current work we assume that neighboring APs do not transmit at the same time in order to avoid possible interference to the receiving MSs. Next we will extend the work by incorporating transmission scheduling into the resource allocations in order to further improve the channel reuse efficiency.

This thesis studies the downlink transmission scheduling only, where the AP makes the scheduling decisions for all the transmissions to its associated MSs. For the uplink transmissions, the problem formulation should consider the special characteristics of the access channel.

Bibliography

Adya, A., Bahl, P., Padhye, J., Wolman, A., and Zhou, L. (2004). A multi-radio unification protocol for IEEE 802.11 wireless networks. In *Proceedings of the First IEEE International Conference on Broadband Networks (BroadNets)*, pages 344–354.

Ahmed, M. and Yanikomeroglu, H. (2009). Throughput fairness and efficiency of link adaptation techniques in wireless networks. *IET Communications*, **3**(7), pages 1227–1238.

Ahmed, M. H. (2005). Call admission control in wireless networks: a comprehensive survey. *IEEE Communications Surveys & Tutorials*, **7**(1), pages 49–68.

Akyildiz, I. F., McNair, J., Ho, J. S., Uzunalioglu, H., and Wang, W. (1999). Mobility management in next-generation wireless systems. *Proceedings of the IEEE*, **87**(8), pages 1347–1384.

Akyildiz, I. F., Xie, J., and Mohanty, S. (2004). A survey of mobility management in next-generation all-IP-based wireless systems. *IEEE Wireless Communications Magazine*, **11**(4), pages 16–28.

Akyildiz, I. F., Wang, X., and Wang, W. (2005). Wireless mesh networks: a survey. *Computer Networks*, **47**(4), pages 445–487.

Alasmary, W. and Zhuang, W. (2012). Mobility impact in IEEE 802.11 p infrastructureless vehicular networks. *Ad Hoc Networks*, **10**(2), pages 222–230.

Alicherry, M., Bhatia, R., and Li, L. E. (2005). Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks. In *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking, New York, NY*, pages 58–72.

Alliance, W. (2005). http://www.wimedia.org/,wimedia alliance.

Alliance, W.-F. (1999). http://www.wi-fi.org/,wi-fi alliance.

Alliance, Z. (2003). http://www.zigbee.org/.

Amir, Y., Danilov, C., Hilsdale, M., Musăloiu-Elefteri, R., and Rivera, N. (2006). Fast handoff for seamless wireless mesh networks. In *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services, New York, NY*, pages 83–95.

Aoun, B. and Boutaba, R. (2006). Max-min fair capacity of wireless mesh networks. In *Proceedings of IEEE International Conference on Mobile Ad hoc and Sensor Systems (MASS), Vancouver, BC*, pages 21–30.

Banchs, A., Serrano, P., and Oliver, H. (2007). Proportional fair throughput allocation in multirate IEEE 802.11e wireless LANs. *Wireless Networks*, **13**(5), pages 649–662.

Bejerano, Y., Han, S.-J., and Li, L. E. (2004). Fairness and load balancing in wireless LANs using association control. In *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking, New York, NY*, pages 315–329.

Bensaou, B. and Fang, Z. (2007). A fair MAC protocol for IEEE 802.11-based ad hoc networks: design and implementation. *IEEE Transactions on Wireless Communications*, **6**(8), pages 2934–2941.

Bicket, J., Aguayo, D., Biswas, S., and Morris, R. (2005). Architecture and evaluation of an unplanned 802.11 b mesh network. In *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking, New York, NY*, pages 31–42.

Boche, H., Schubert, M., and Stanczak, S. (2007). Proportionally fair resource allocation for wireless networks. *http://www.eurasip.org/Proceedings/Ext/WSA07/papers/1569015387.pdf*.

Bononi, L., Di Felice, M., Molinaro, A., and Pizzi, S. (2009). A cross-layer architecture for efficient multi-hop communication in multi-channel multi-radio wireless mesh networks. In *Proceedings of the 6th IEEE Annual Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops, Rome*, pages 1–6.

Borst, S., Hegde, N., and Proutiere, A. (2009). Mobility-driven scheduling in wireless networks. In *Proceedings of IEEE INFOCOM 2009, Rio de Janeiro*, pages 1260–1268.

BWN-Mesh, G. I. o. T. (2005). http://www.ece.gatech.edu/research/labs/bwn/mesh/.

Calì, F., Conti, M., and Gregori, E. (2000). Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit. *IEEE/ACM Transactions on Networking*, **8**(6), pages 785–799.

Chen, W.-T. and Chen, P.-Y. (2003). Group mobility management in wireless ad hoc networks. *Proceedings of the IEEE 58th on Vehicular Technology Conference, Orlando, Florida*, **4**, pages 2202–2206.

Chen, Y.-S., Chuang, M.-C., and Chen, C.-K. (2008). Deucescan: deuce-based fast handoff scheme in IEEE 802.11 wireless networks. *IEEE Transactions on Vehicular Technology*, **57**(2), pages 1126–1141.

Chen, Y.-Y., Liu, S.-C., and Chen, C. (2006). Channel assignment and routing for multi-channel wireless mesh networks using simulated annealing. In *Proceedings of IEEE on Global Telecommunications Conference (GLOBECOM), New York, NY*, pages 1–5.

Chi, K.-H., Jiang, J.-H., and Yen, L.-H. (2006). Cost-effective caching for mobility support in IEEE 802.1 X frameworks. *IEEE Transactions on Mobile Computing*, **5**(11), pages 1547–1560.

Churchill, S. (2007). http://www.dailywireless.org/2007/04/23/belair-live-in-london/.

Cicconetti, C., Lenzini, L., and Mingozzi, E. (2008). Interference-aware distributed scheduling in TDMA wireless mesh networks. In *Proceedings of the 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems, Atlanta, GA*, pages 505–506.

Cisco (2007). http://www.cisco.com/en/us/products/ps6548/.

Dugar, A., Vaidya, N., and Bahl, P. (2001). Priority and fair scheduling in a wireless LAN. *Proceedings of the IEEE Military Communications Conference for Network-Centric Operations: Creating the Information Force*, **2**, pages 993–997.

Dutta, A., Wong, K. D., Burns, J., Jain, R., McAuley, A., Young, K., and Schulzrinne, H. (2002). Realization of integrated mobility management protocol for ad-hoc networks. *Proceedings of the IEEE MILCOM*, **1**, pages 448–454.

Elshenawy, M., El-Darieby, M., and Abdulhai, B. (2009). Vehicular mobility management schemes for balancing loads among wmn access points. In *Proceedings of the 69th IEEE Vehicular Technology Conference, New York, NY*, pages 1–5.

Fashandi, S. and Todd, T. D. (2005). Real-time handoff in solar/battery powered ESS mesh networks. *Proceedings of the IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Berlin*, **3**, pages 1489–1494.

Ferng, H.-W., Lee, C.-F., Huang, J.-J., and Chiu, G.-M. (2005). Designing a fair scheduling mechanism for IEEE 802.11 wireless LANs. *IEEE on Communications Letters*, **9**(4), pages 301–303.

Forum, W. (2001). http://www.wimaxforum.org/home,wimax forum.

Grossglauser, M. and Tse, D. N. (2002). Mobility increases the capacity of ad hoc wireless networks. *ACM Transactions on Networking*, **10**(4), pages 477–486.

Group, I. . S. (2001a). http://www.ieee802.org/16/,ieee 802.16 standard group.

Group, I. . S. (2001b). http://www.ieee802.org/20/,ieee 802.20 standard group.

Group, I. . S. (2002a). http://www.ieee802.org/11/,ieee 802.11 standard group.

Group, I. . S. (2002b). http://www.ieee802.org/15/,ieee 802.15 standard group.

Gummalla, A. C. V. and Limb, J. O. (2000). Wireless medium access control proto-
cols. *IEEE Communications Surveys & Tutorials*, **3**(2), pages 2–15.

Haas, Z. J. and Liang, B. (1999). Ad hoc mobility management with uniform quorum
systems. *IEEE/ACM Transactions on Networking (TON)*, **7**(2), pages 228–240.

Hauser, J. (2003). Draft par for IEEE 802.11 ESS mesh. *IEEE Document Number:
IEEE 802.11-03/759r2*.

Heusse, M., Rousseau, F., Guillier, R., and Duda, A. (2005). Idle sense: an optimal
access method for high throughput and fairness in rate diverse wireless LANs. *ACM
SIGCOMM Computer Communication Review*, **35**(4), pages 121–132.

Huang, R., Zhang, C., and Fang, Y. (2007). A mobility management scheme for wire-
less mesh networks. In *IEEE on Global Telecommunications Conference, Washing-
ton, DC*, pages 5092–5096.

Huang, X. L. and Bensaou, B. (2001). On max-min fairness and scheduling in wireless
ad-hoc networks: analytical framework and implementation. In *Proceedings of the
2nd ACM International Symposium on Mobile Ad Hoc Networking & Computing,
New York, NY*, pages 221–231.

Jain, R., Chiu, D.-M., and Hawe, W. R. (1984). *A quantitative measure of fairness*

and discrimination for resource allocation in shared computer system. Eastern Research Laboratory, Digital Equipment Corporation.

Jalali, A., Padovani, R., and Pankaj, R. (2000). Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. *Proceedings of the IEEE 51st Vehicular Technology Conference Proceedings, Tokyo*, **3**, pages 1854–1858.

Jiang, H., Wang, P., and Zhuang, W. (2007). A distributed channel access scheme with guaranteed priority and enhanced fairness. *IEEE Transactions on Wireless Communications*, **6**(6), pages 2114–2125.

Jiang, S., Rao, J., He, D., Ling, X., and Ko, C. C. (2002). A simple distributed PRMA for MANETs. *IEEE Transactions on Vehicular Technology*, **51**(2), pages 293–305.

Jones, B. (2005). DSRC-linking the vehicle and the road. In *VII Public Meeting, ITS America*.

Jou, T. and Eastlake, D. (2004). ESS MESH network study group meeting minutes.

Jubin, J. and Tornow, J. D. (1987). The DARPA packet radio network protocols. *Proceedings of the IEEE*, **75**(1), pages 21–32.

Jun, J. and Sichitiu, M. L. (2003). Fairness and QoS in multihop wireless networks. *Proceedings of the IEEE 58th Vehicular Technology Conference, Orlando, Florida*, **5**, pages 2936–2940.

Kelly, F. (1997). Charging and rate control for elastic traffic. *European transactions on Telecommunications*, **8**(1), pages 33–37.

Kelly, F. P., Maulloo, A. K., and Tan, D. K. (1998). Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research society*, **49**(3), pages 237–252.

Kinney, P. (2003). IEEE 802.15 general interest in mesh networking, IEEE 802.15 request for information of a mesh network study group.

Kiyon (2006). http://www.kiyon.com.

Korkmaz, G., Ekici, E., and Ozguner, F. (2006). A cross-layer multihop data delivery protocol with fairness guarantees for vehicular networks. *IEEE Transactions on Vehicular Technology*, **55**(3), pages 865–875.

Kyasanur, P. and Vaidya, N. H. (2005). Routing and interface assignment in multi-channel multi-interface wireless networks. *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, **4**, pages 2051–2056.

Leiner, B. M., Ruther, R., and Sastry, A. R. (1996). Goals and challenges of the DARPA GloMo program [global mobile information systems]. *IEEE Personal Communications Magazines*, **3**(6), pages 34–43.

Li, B. (2005). End-to-end fair bandwidth allocation in multi-hop wireless ad hoc networks. In *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems, Columbus, OH*, pages 471–480.

Li, L., Halpern, J. Y., Bahl, P., Wang, Y.-M., and Wattenhofer, R. (2005). A cone-based distributed topology-control algorithm for wireless multi-hop networks. *IEEE/ACM Transactions on Networking*, **13**(1), pages 147–159.

Liu, X., Chong, E. K. P., and Shroff, N. B. (2001). Opportunistic transmission scheduling with resource-sharing constraints in wireless networks. *IEEE Journal on Selected Areas in Communications*, **19**(10), pages 2053–2064.

Luo, H., Lu, S., and Bharghavan, V. (2000). A new model for packet scheduling in multihop wireless networks. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, New York, NY*, pages 76–86.

Luo, J., Rosenberg, C., and Girard, A. (2010). Engineering wireless mesh networks: joint scheduling, routing, power control, and rate adaptation. *IEEE/ACM Transactions on Networking*, **18**(5), pages 1387–1400.

Mahmud, S. A., Khan, S., Khan, S., and Al-Raweshidy, H. (2006). A comparison of MANETs and WMNs: commercial feasibility of community wireless networks and MANETs. In *Proceedings of the 1st International Conference on Access Networks, New York, NY*, page 18.

Meng, X., Tan, K., and Zhang, Q. (2006). Joint routing and channel assignment in multi-radio wireless mesh networks. *Proceedings of the IEEE International Conference on Communications, New York, NY*, **8**, pages 3596–3601.

Mingyi Hong, A. G. (2013). Mechanism design for base station association and resource allocation in downlink ofdma network. *http://arxiv.org/pdf/1204.6105.pdf*.

Mohsenian Rad, A. and Wong, V. W. (2006). Joint optimal channel assignment and congestion control for multi-channel wireless mesh networks. *Proceedings of the IEEE International Conference on Communications (ICC), Istanbul*, **5**, pages 1984–1989.

Mohsenian-Rad, A. H. and Wong, V. W. (2007). Joint logical topology design, interface assignment, channel allocation, and routing for multi-channel wireless mesh networks. *IEEE Transactions on Wireless Communications*, **6**(12), pages 4432–4440.

Nash Jr, J. F. (1950). The bargaining problem. *Econometrica: Journal of the Econometric Society*, pages 155–162.

Net-X, U. o. I. (2007). http://www.crhc.uiuc.edu/wireless/netx.html.

Piggin, P., Lewis, B., and Whitehead, P. (2003). Mesh networks in fixed broadband wireless access: multipoint enhancements for the 802.16 standard.

Qiao, D. and Shin, K. G. (2002). Achieving efficient channel utilization and weighted fairness for data communications in IEEE 802.11 WLAN under the DCF. In *Proceedings of the IEEE International Workshop on Quality of Service, Miami Beach*, pages 227–236.

Qiao, D. and Shin, K. G. (2003). UMAV: A simple enhancement to the IEEE 802.11 DCF. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, Big Island, HI*, pages 1–9.

Rad, A. M. and Wong, V. W. (2007). Joint channel allocation, interface assignment and mac design for multi-channel wireless mesh networks. In *Proceedings of the 26th IEEE International Conference on Computer Communications, Anchorage, AK*, pages 1469–1477.

Ramanathan, R. (2001). On the performance of ad hoc networks with beamforming

antennas. In *Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking & Computing, New York, NY*, pages 95–105.

Raniwala, A., Pradipta, D., and Sharma, S. (2007). End-to-end flow fairness over IEEE 802.11-based wireless mesh networks. In *Proceedings of the 26th IEEE International Conference on Computer Communications, Anchorage, AK*, pages 2361–2365.

S. Boyd, L. Xiao, A. M. and Mattingley, J. (2008). Notes on decomposition methods. *EE364B, Stanford University*, pages 1–36.

Salem, N. B. and Hubaux, J.-P. (2005). A fair scheduling for wireless mesh networks. In *Proceedings of the IEEE Workshop on Wireless Mesh Networks (WiMesh), Santa Clara, CA*, pages 1–8.

So, J. and Vaidya, N. H. (2004). Multi-channel mac for ad hoc networks: handling multi-channel hidden terminals using a single transceiver. In *Proceedings of the 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing, New York, NY*, pages 222–233.

Sobrinho, J. L. and Krishnakumar, A. S. (1996). Real-time traffic over the IEEE 802.11 medium access control layer. *Bell Labs Technical Journal*, **1**(2), pages 172–187.

Spyropoulos, A. and Raghavendra, C. S. (2003). Asympotic capacity bounds for ad-hoc networks revisited: the directional and smart antenna cases. *Proceedings of the IEEE Global Telecommunications Conference*, **3**, pages 1216–1220.

Staehle, B., Staehle, D., and Pries, R. (2009). Effects of link rate assignment on the max-min fair throughput of wireless mesh networks. In *Proceedings of the 21st International Teletraffic Congress, Paris*, pages 1–8.

Stolyar, A. L. (2005). On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation. *Operations research*, **53**(1), pages 12–25.

Sucec, J. and Marsic, I. (2002). Location management for hierarchically organized mobile ad hoc networks. *Proceedings of the IEEE Wireless Communications and Networking Conference, Orlando, FL*, **2**, pages 603–607.

Sundaresan, K., Sivakumar, R., Ingram, M. A., and Chang, T.-Y. (2004). A fair medium access control protocol for ad-hoc networks with MIMO links. *Proceedings of the 23rd IEEE Annual Joint Conference of the IEEE on Computer and Communications Societies, Hong Kong*, **4**, pages 2559–2570.

Tang, J., Xue, G., and Zhang, W. (2005). Interference-aware topology control and QoS routing in multi-channel wireless mesh networks. In *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing, New York, NY*, pages 68–77.

Tang, J., Xue, G., and Zhang, W. (2006). Maximum throughput and fair bandwidth allocation in multi-channel wireless mesh networks. In *Proceedings of IEEE INFOCOM, Barcelona, Catalunya, Spain*, pages 1–10.

Tassiulas, L. and Sarkar, S. (2002). Maxmin fair scheduling in wireless networks.

Proceedings of the 21st IEEE Annual Joint Conference on Computer and Communications Societies, **2**, pages 763–772.

Tropos (2006). http://www.tropos.com/solutions/.

Trung, T. M. and Mo, J. (2010). A multichannel TDMA MAC protocol to reduce end-to-end delay in wireless mesh networks. *ETRI journal*, **32**(5), pages 819–822.

Vaidya, N., Dugar, A., Gupta, S., and Bahl, P. (2005). Distributed fair scheduling in a wireless LAN. *IEEE Transactions on Mobile Computing*, **4**(6), pages 616–629.

Velayos, H., Aleo, V., and Karlsson, G. (2004). Load balancing in overlapping wireless LAN cells. *Proceedings of the IEEE International Conference on Communications (ICC),*, **7**, pages 3833–3836.

Viswanathan, H. and Mukherjee, S. (2005). Performance of cellular networks with relays and centralized scheduling. *IEEE Transactions on Wireless Communications*, **4**(5), pages 2318–2328.

Wei, H.-Y., Kim, S., Ganguly, S., and Izmailov, R. (2006). Seamless handoff support in wireless mesh networks. In *Proceedings of the 1st Workshop on Operator-Assisted (Wireless Mesh) Community Networks, Berlin*, pages 1–8.

Whitehead, P., Piggin, P., Lewis, B., and Lynch, S. (2003). Mesh extensions to IEEE 802.16 and 16a, IEEE 802.16 proposal.

Wu, K. and Harms, J. (2006). Performance study of proactive flow handoff for mobile ad hoc networks. *Wireless Networks*, **12**(1), pages 119–135.

Xiang, W., Pratt, T., and Wang, X. (2004). A software radio testbed for two-transmitter two-receiver space-time coding OFDM wireless LAN. *IEEE Communications Magazine*, **42**(6), pages S20–S28.

Xie, B., Yu, Y., Kumar, A., and Agrawal, D. P. (2006). Load-balancing and inter-domain mobility for wireless mesh networks. In *Proceedings of the IEEE Global Telecommunications Conference, San Francisco, CA*, pages 1–6.

Xie, J. and Wang, X. (2008). A survey of mobility management in hybrid wireless mesh networks. *IEEE Network*, **22**(6), pages 34–40.

Yang, K., Ou, S., Chen, H.-H., and He, J. (2007). A multihop peer-communication protocol with fairness guarantee for IEEE 802.16-based vehicular networks. *IEEE Transactions on Vehicular Technology*, **56**(6), pages 3358–3370.

Yang, Y. and Zhao, D. (2011). Joint handoff and energy management for a wireless mesh network. In *Proceedings of the IEEE International Conference on Communications (ICC), Kyoto*, pages 1–5.

Zhang, J., Wu, H., Zhang, Q., and Li, B. (2005). Joint routing and scheduling in multi-radio multi-channel multi-hop wireless networks. In *Proceedings of the 2nd International Conference on Broadband Networks*, pages 631–640.

Zhang, Y. and Qu, S. (2006). Joint power control and channel assignment algorithms for wireless ad hoc networks. In *Proceedings of the 4th International Conference on Ubiquitous Intelligence and Computing, Berlin Heidelberg*, pages 270–279. Springer.

Zhao, D. (2007). Throughput fairness in infrastructure-based IEEE 802.11 mesh networks. *IEEE Transactions on Vehicular Technology*, **56**(5), pages 3210–3219.

Zhao, W. and Xie, J. (2010). A novel Xcast-based caching architecture for inter-gateway handoffs in infrastructure wireless mesh networks. In *Proceedings of the IEEE on INFOCOM, San Diego, CA*, pages 1–9.

Zhou, L. (1997). The Nash bargaining theory with non-convex problems. *Econometrica: Journal of the Econometric Society*, **65**(3), pages 681–685.