

**MIMO-Based Wireless Mesh Networks:
Resource Allocation and Performance Evaluation**

by

An Thai Nguyen

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Master of Science

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada

Copyright © 2008 An Thai Nguyen

**THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION**

**MIMO-Based Wireless Mesh Networks:
Resource Allocation and Performance Evaluation**

BY

An Thai Nguyen

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of
Manitoba in partial fulfillment of the requirement of the degree**

Of

Master of Science

An Thai Nguyen © 2008

Permission has been granted to the University of Manitoba Libraries to lend a copy of this thesis/practicum, to Library and Archives Canada (LAC) to lend a copy of this thesis/practicum, and to LAC's agent (UMI/ProQuest) to microfilm, sell copies and to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

Supervisor: Prof. E. Hossain

ABSTRACT

Multiple input multiple output (MIMO) is a promising physical layer approach to boost the performance of multihop wireless networks such as the wireless mesh networks (WMNs). WMNs are dynamically self-organized and self-configured networks which are built using a wireless mesh backbone consisting of mesh routers. The wireless mesh backbone, which provides multihop wireless connectivity to mesh clients, requires high throughput and reliability in order to provide quality of service (QoS) for real-time as well as non-real-time applications. To achieve high channel utilization and to guarantee the QoS requirements of different types of traffic, the radio resources such as the antennas and transmission power in a mesh router need to be allocated optimally among the different flows. Resource allocation and end-to-end performance analysis for MIMO-based wireless mesh backbone networks pose significant research challenges. In order to realize the full potential of MIMO technology, higher layer protocols must be designed to be cognizant of the MIMO link capability. In particular, channel state information (CSI) from the physical layer should be exploited for optimal resource allocation at the medium access control (MAC) layer. An antenna assignment and allocation (ASA) scheme is presented which uses only a subset of total available antennas for each type of service and provides differentiated QoS among different flows in a mesh router. Also, to provide efficient channel utilization, the ASA technique considers adaptive modulation and coding (AMC) to exploit CSI. This scheme is developed based on a Markov Decision Process (MDP) formulation of the antenna assignment problem. The MDP formulation exploits a queueing analytical model for the data queues at a mesh router. The performance of the proposed scheme is compared with the traditional weighted round-robin type of scheme for antenna scheduling. Numerical results demonstrate the efficacy of the proposed scheme. To this end, for performance analysis of a wireless mesh backbone in an end-to-end MIMO transmission scenario, we propose a tandem queueing model. This model considers the implementation of AMC in the physical layer and ARQ-error recovery in the link layer. Both exact and approximate decomposition approaches are developed to solve the tandem queueing problem. The performance analysis results obtained

from the analytical model show the significant performance gain achieved through using MIMO links compared to that for using single input single output (SISO) links.

Examiners:

Prof. E. Hossain, Supervisor, Dept. of Electrical & Computer Engineering

Dr. J. Diamond, *TRLabs* and Dept. of Electrical & Computer Engineering

Prof. P. Thulasiraman, External Examiner, Dept. of Computer Science

Table of Contents

Abstract	ii
Table of Contents	v
List of Figures	viii
Acknowledgement	x
1 Introduction	1
1.1 Wireless Mesh Networks and Its Applications	1
1.2 Directional Antenna	3
1.3 Multiple Input Multiple Output (MIMO) Technology	5
1.3.1 MIMO Channel Model	6
1.3.2 Spatial Diversity Gain on MIMO system	7
1.3.3 Spatial Multiplexing MIMO Systems	9
1.4 Objective, Motivation, and Scope of The Thesis	11
1.5 Organization of The Thesis	14
2 Resource Allocation in MIMO-Based Wireless Networks	16
2.1 Medium Access Control Protocol in MIMO Wireless Networks	17
2.1.1 Centralized Medium Access Control in MIMO Wireless Networks	17
2.1.2 Distributed Medium Access Control in MIMO Wireless Networks	18
2.1.2.1 Medium Access Control Exploiting MIMO Spatial Diversity	20
2.1.2.2 Medium Access Control Exploiting Spatial Multiplex- ing MIMO	21
2.2 Power and Rate Control in MIMO Wireless Networks	22
2.3 Admission Control in MIMO Wireless Networks	23

2.4	Antenna Selection and Assignment in MIMO Wireless Mesh Networks	24
2.5	Cross-Layer Design in MIMO Systems: Perspectives and Challenges	25
2.5.1	Complexity in Neighbor Discovery	25
2.5.2	Difficulty in Obtaining the Channel State Information	26
2.5.3	Different Performance Objectives and Different MAC Design	27
2.5.4	Difficulty in Optimizing Resource Allocation in a High Mobility Environment	27
3	Quality of Service in MIMO Wireless Mesh Networks: An Antenna Selection Approach	29
3.1	Introduction	29
3.2	Related Work	30
3.3	System Model, Assumptions, Problem Definition and Solution Methodology	32
3.3.1	Physical and Link Layer Model	33
3.3.1.1	Signal Model	33
3.3.1.2	Channel Model	36
3.3.2	Problem Definition and Solution Methodology	39
3.4	Formulation of the Markov Decision Process Model	42
3.4.1	Packet Arrival Process	42
3.4.2	Packet Departure Process	43
3.4.3	State Space and Transition Probability Matrix	44
3.4.4	Queueing Performance for Each Type of Service	47
3.4.4.1	Steady State Probability	47
3.4.4.2	Average Queue Length	47
3.4.4.3	Packet Dropping Probability	48
3.4.4.4	Queue Throughput	49
3.4.4.5	Average Delay for one Packet	49
3.4.5	Markov Decision Process Model	49
3.5	Performance Evaluation	54
3.5.1	Complexity of Policy Evaluation	54
3.5.2	Parameter Setting	55
3.5.3	Queueing Performance	56

3.5.4	Simulation Methodology	57
4	A Tandem Queue Model for Performance Analysis in MIMO Wire-	
	less Mesh Networks	64
4.1	Introduction	64
4.2	System Model and Assumptions	65
4.2.1	The physical and link layer models	66
4.3	An Exact Tandem Queue Model	67
4.3.1	End to End Packet Dropping Probability	70
4.3.2	Average Delay	71
4.4	General Case ($L > 2$)	71
4.5	Solution of the Tandem Queueing Model: Decomposition Approach .	72
4.5.1	Technical Approach	72
4.5.2	Tandem Queue with Exogenous Traffic	75
4.6	Validation of Decomposition Approach and Typical Numerical Results	75
4.7	Summary	78
5	Conclusion	81
5.1	Summary	81
5.2	Future Research Works	82
	Bibliography	85

List of Figures

Figure 1.1	A simple wireless mesh network.	2
Figure 1.2	Transmission using directional and omidirectional antenna. . .	5
Figure 1.3	SISO, SIMO, MISO, and MIMO transmission.	8
Figure 1.4	Spatial Diversity transmits the same data stream across all antennas, and makes use of multipath to increase the chance of correctly decoding the received signal, i.e. lower bit error rate.	9
Figure 1.5	Spatial Multiplexing transmits independent data streams in the same time slot and frequency band simultaneously.	10
Figure 1.6	VBLAST system diagram.	11
Figure 1.7	Capacity of an i.i.d. Rayleigh channel with different SNR. . .	12
Figure 1.8	Capacities of $n \times n$ MISO channel, $1 \times n$ SIMO channel and $n \times n$ MIMO channel (for SNR = 0 dB).	13
Figure 2.1	Joint design of PHY and MAC layers for MIMO links in ad hoc networks.	20
Figure 2.2	Complexity in neighbor discovery.	26
Figure 3.1	Converting a MIMO channel into a parallel eigenchannel by using Singular Decomposition Value technique.	35
Figure 3.2	Transmission model between two mesh routers.	39
Figure 3.3	The procedure to find the optimal solution for ASA.	41
Figure 3.4	Algorithm for the Antenna Assignment Controller.	54
Figure 3.5	Computational complexity when the system has 2 separate subchannels and K is the number of transmission modes.	55
Figure 3.6	Computational complexity when system using 5 transmission modes and m is number of separate subchannels.	56

Figure 3.7	Average packet delay under different packet arrival rate with $\zeta_1 = \zeta_2$ and subchannel assigned to best-effort queue has $\gamma_2 = 10dB$ (from analysis).	57
Figure 3.8	Average packet blocking probability under different packet arrival rate with $\zeta_2 = 0.8$ and subchannel assigned to best-effort queue has $\gamma_2 = 10dB$ (from analysis).	58
Figure 3.9	Relation between states and the deterministic policy, where $\zeta_1 = 0.8, \zeta_2 = 0.5, \gamma_1 = 15dB$ and $\gamma_2 = 5dB$.	59
Figure 3.10	Packet throughput under different packet arrival rate (for $\zeta_2 = 0.8$, and $\gamma_2 = 5$ dB).	60
Figure 3.11	Packet blocking probability under different SNR (for $\zeta_1 = \zeta_2 = 0.5$, and $\gamma_2 = 5$ dB).	61
Figure 3.12	Impact of connection arrival rate on average delay for QoS-sensitive queue (for $\zeta_2 = 0.5$ and $\gamma_2 = 5$ dB).	62
Figure 3.13	Average delays for best-effort and QoS-sensitive traffic (for $\zeta_2 = 0.7$ and $\gamma_1 = 9$ dB $\gamma_2 = 5$ dB).	63
Figure 4.1	A multihop wireless network with multiple ongoing connections.	66
Figure 4.2	A tandem system of queues.	66
Figure 4.3	A tandem system with two queues.	68
Figure 4.4	Impact of packet arrival rate on average delay (for SISO: $\gamma_0 = 15$ dB, for 2×3 MIMO: $\gamma_0 = 5$ dB, for 3×3 MIMO: $\gamma_0 = 3$ dB, $Q_1 = Q_2 = 50, L = 2$).	76
Figure 4.5	Impact of packet arrival rate on end - end packet dropping probability (for SISO: $\gamma_0 = 15$ dB, for 2×3 MIMO: $\gamma_0 = 5$ dB, for 3×3 MIMO: $\gamma_0 = 3$ dB, $Q_1 = Q_2 = 50, L = 2$).	77
Figure 4.6	Impact of packet arrival rate on average delay in 2×2 MIMO with $\gamma_0 = 5$ dB, $Q_1 = Q_2 = 50, L = 2$, with different exogenous traffic.	78
Figure 4.7	Impact of packet arrival rate on average delay in 2×2 MIMO with $\gamma_0 = 5$ dB, $Q_k = 50$, with different number of hops, i.e., $H = L - 1$.	79
Figure 4.8	Impact of end - end packet dropping probability on average delay in 2×2 MIMO with $\gamma_0 = 5$ dB, $Q_k = 50$, with different number of hops, i.e., $H = L - 1$.	80

Acknowledgement

I would like to express my gratitude to Dr. Ekram Hossain for his abundant help, valuable assistance, support and guidance. I am also very grateful for his encouragement, patience with time for teaching me technical writing and introducing me to several exciting research topics.

I am indebted to Dr. Jeff Diamond for his excellent comments and suggestions during my thesis revision period. I feel providential to have learned from his remarkable combination of technical knowledge and research enthusiasm. I also appreciate Dr. P. Thulasiraman for her time and effort to serve in my thesis committee.

This research has been supported by *TRLabs* grant, for which I am grateful.

My warmest thanks are to my colleagues in the University of Manitoba for their sharing and caring, the friendship and the cooperation. Especial thanks to Dr. Long Bao Le and Dr. Dusit Niyato for their numerous timely advices and support. I would also like to thank my other friends, Tien Nguyen, Hai Pham, Quy Nguyen and Trung Nguyen, for the time we had together and the encouragement they reserved for me.

My heartfelt thanks are to the Friesen family, with whom I have shared many of my dreams, I am forever thankful for their sharing and caring, the support and the happy smiles. Their hospitality is wonderful and it is very helpful for completing my thesis.

My deepest love is, as always, with my parents and my brother family. All my life, they have always been there to love, to take care, to listen and to support. I am proud of my mom and my dad for teaching me to be a good person, and I thank my parents for letting me know that their faith in me would never fade.

Chapter 1

Introduction

1.1 Wireless Mesh Networks and Its Applications

Wireless Mesh Networks (WMNs) are dynamically self-organized and self-configured wireless networks [6]. Nodes in these networks automatically establish an ad hoc network and maintain connectivity in the network. This feature helps to reduce the cost of establishing the backbone network, decreases the complexity of network maintenance, and also improves the reliability of network coverage. In fact, there are two types of nodes in wireless mesh networks, namely, static nodes (mesh routers) and mobile nodes (mesh clients). The mesh routers have a minimal mobility and they form the backbone of the network. The client nodes in WMNs can freely and dynamically self organize into arbitrary and temporary ad hoc network topologies, allowing wireless devices to seamlessly inter-network in coverage areas of a WMN backbone. The network backbone provides multihop connectivity among the mesh routers and Internet gateways. That is, when a mesh client wants to connect to the Internet, it connects to the mesh router and then the mesh router will establish a connection (using wireless multihop connectivity) with other mesh routers until it finds the Internet gateway. Mesh networks are expected to play a key role to provide last mile connections and to backhaul traffic in wireless broadband networks [6]. Note that relay-based implementation to extend the coverage and enhance the system capacity was also proposed for centralized wireless networks (i.e., cellular and wireless LAN networks) [77], [63].

The unique topology of WMNs leads to many strong points that are hard to find in other access network technologies. WMNs provide a new way to access the Internet to potential users who are currently not serviced due to geographical, financial or

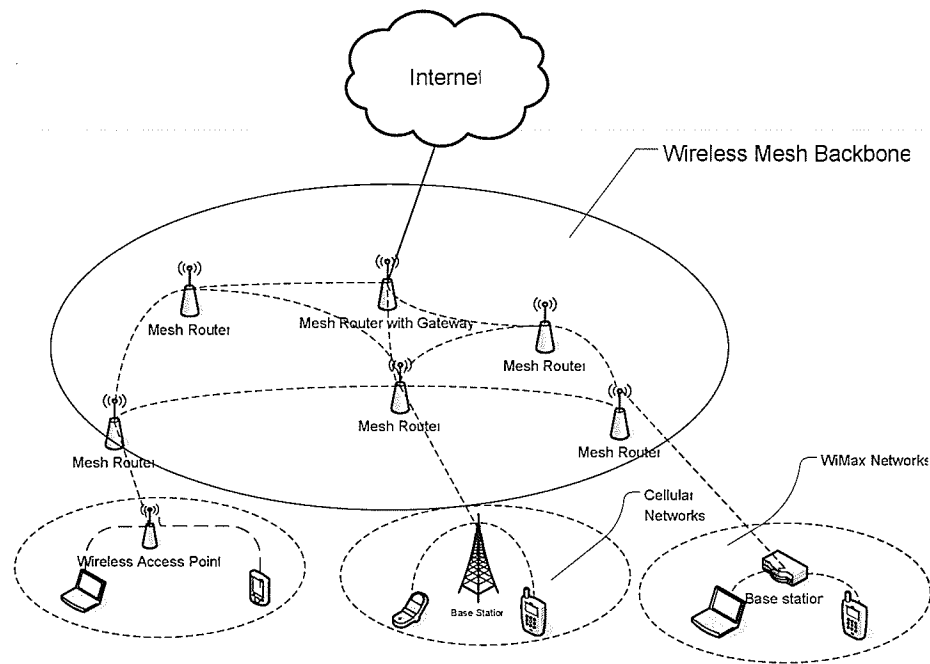


Figure 1.1. A simple wireless mesh network.

technological restrictions [40]. Users and Internet service providers are expected to benefit from WMNs in many ways as follows:

- **Scalability:** The network can grow as more and more customers are added. If the capacity becomes an issue, additional gateways can be added.
- **Multiple types of network access:** In WMNs, both backhaul access Internet and peer to peer communications are supported [41].
- **Low up-front investment:** Incremental expansion is possible in WMNs, which eliminates the challenge of a large initial investment required in deploying wired Internet access services, e.g., DSL or cable-based networks.
- **Reliability:** The mesh structure assures the availability of multiple paths for each node in the network. If a node fails, others will take over its traffic enabling uninterrupted services.
- **QoS:** Quality of Service provisioning for multimedia traffic is crucial [71]. With a careful design and enough gateways in appropriate locations, WMNs can support QoS for multimedia services. This feature of WMNs will be highlighted

in this thesis.

- **Flexibility:** Different configurations can be adapted to meet the requirements of a particular situation.
- **Mobile user support:** Mobile users can connect to the network as long as they stay within the range of any other user node (mobile or not).
- **Roaming service:** As customers travel, roaming contracts between companies can be agreed upon for added value to the service.

Due to wireless channel impairments and complicated interactions among protocols in different layers, engineering multihop wireless networks with quality of service (QoS) assurance is a very challenging task.

WMNs are expected to support real-time applications such as video streaming and gaming etc. These applications require low and stable packet delay, and constant throughput. These requirements may not be easily achieved in interference-limited mesh networks using conventional antenna technologies. In this context, the use of smart antennas and multiple input multiple output (MIMO) is a promising PHY approach to boost the network throughput.

To achieve high-speed wireless communications in a WMN backbone, we propose to use MIMO links in between the mesh routers. In the physical layer, we consider adaptive modulation and coding (AMC) and we capture the effects of correlated channel fading by using a finite state Markov chain (FSMC) model. Automatic repeat request (ARQ) is used for error control so that the reliability of the transmission can be ensured. The details of the MIMO wireless mesh backbone network where each node in the network is equipped with multiple antennas and adaptive modulation is used in transmission links in single-hop and multihop wireless scenario will be described in Chapter 3 and 4, respectively.

1.2 Directional Antenna

The development of RF technology and circuit design have been contributing towards the development of efficient antenna technology for high-speed wireless communications. Traditionally, there are two types of smart antennas: directional antenna and multiple input and multiple output (MIMO) [34]. A directional antenna generates

multiple predefined narrow beams directed to each particular direction in order to enhance the received signal strength (RSS) and/or signal-to-noise ratio (SNR) and applies one at a time towards the direction of interest. There are two major types of directional antennas [46]:

- **Switched-beam antenna:** This type of directional antenna system measures RF power from a set of predefined beams and then selects one from a set of predefined beams/antennas.
- **Digitally Adaptive Beamformers (DAB):** These systems use adaptive techniques to enhance the radio link. In such a system, each separate antenna is down converted, digitized, and processed .

However, when we use directional antennas to design a WMN backbone, the existing standard for MAC protocols (e.g., IEEE 802.11) may not work efficiently or take full advantage of using a directional antenna because those standards were designed for omni-directional antenna technology. The difference between the transmission scenarios using directional antenna and omi-directional antenna is shown in Figure 1.2. Using directional antennas may aggravate the hidden node problem in traditional IEEE 802.11-based networks.

A solution was investigated in [19] to reduce the number of hidden nodes in such networks when using directional antennas. But such a solution which requires fast steerable directional antenna increases cost and system complexity.

The results in [57] showed that by using antenna arrays, significant improvements in channel capacity (or spectrum efficiency) can be achieved. Also, antenna arrays can help reduce multipath fading and thereby increase the data rate. Such an antenna array is known as a smart antenna. Smart antenna technology can be classified into three types [34]: adaptive antenna array technology, MIMO technology, and space division multiple access technology. Despite its technical merits, however, recently, beamforming has not found commercial adoption due to its requirement for rich channel knowledge at the transmitter. We will focus on the MIMO technology in the rest of the thesis.

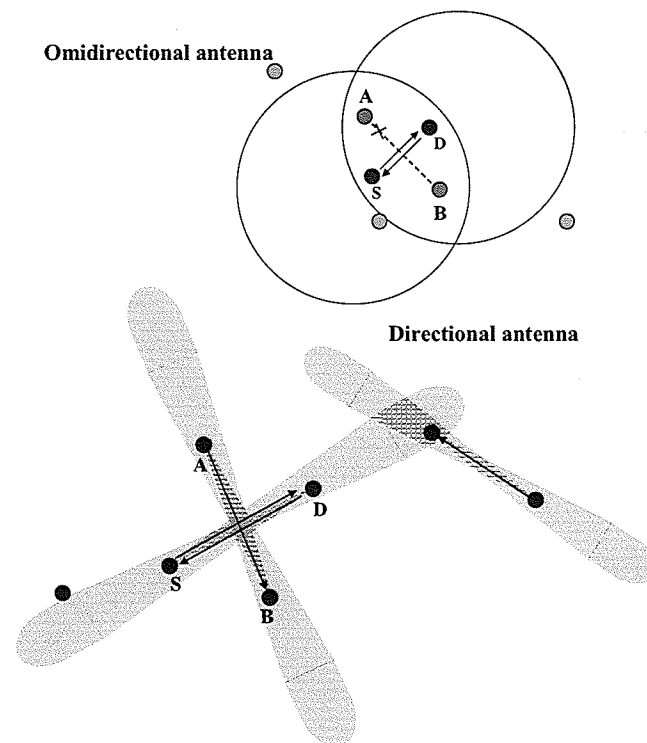


Figure 1.2. *Transmission using directional and omidirectional antenna.*

1.3 Multiple Input Multiple Output (MIMO) Technology

Although there are a lot of benefits of using directional antennas, in order to reduce the deployment cost and being closer to the users, the wireless mesh routers are mostly deployed at low to moderate heights in order of 3-10 m where direct LOS (main component to create beamforming in smart antenna techniques) is difficult to be guaranteed [34]. Depending on the position of wireless routers we can have three different communications scenarios that affect the channel propagation statistics [2]. The results in [2] also showed that, in the large scattering angles, MIMO performs better than directional antennas.

MIMO wireless systems are those that have multiple antenna elements at both sides, the transmitter and the receiver. They were first investigated by computer simulations in the 1980s [66] and were studied analytically later. Recent research on

MIMO, related to both channel capacity and the design of communication schemes, demonstrates a great improvement of performance in a MIMO-based wireless system. Indeed, MIMO antennas use sophisticated physical layer techniques to provide significant benefits over conventional antenna technology. Multiple independent data streams can be sent over the MIMO antenna elements. MIMO links can also suppress interference from neighboring links as long as the total useful streams and interfering streams are not greater than the number of receiving antenna elements. For these reasons, MIMO technology is increasingly being considered for use in interference-limited WMNs and have been adopted in third generation (3G) cellular systems (e.g., WCDMA) as well as in WLAN and WIMAX standards. However, the benefits of the MIMO technology in improving network performance are limited unless the higher layer protocols are able to exploit these capabilities [11].

Multiple antennas in a MIMO system can be used in two different ways. One is increasing the amount of *diversity* to combat channel fading and the other is the transmission of several parallel data streams to increase the system capacity (this scheme called *spatial multiplexing*). Given a MIMO channel both gains can in fact be simultaneously obtained. However, the study in [73] showed that there is a *fundamental tradeoff* between how much of each type of gain any coding scheme can extract: higher *spatial multiplexing gain* comes at the price of sacrificing *diversity*.

1.3.1 MIMO Channel Model

We consider here a single user Gaussian channel, similar to [53], with multiple transmitting and/or receiving antennas. We denote the number of transmitting antennas by N_t and the number of receiving antennas by N_r . The varying channel impulse response between the j -th ($j = 1, 2, \dots, N_t$) transmit antenna and the i -th ($i = 1, 2, \dots, N_r$) receive antenna is denoted by $h_{i,j}(\tau, t)$, then the composite MIMO channel matrix is given by $N_r \times N_t$ matrix $\mathbf{H}(\tau, t)$, where

$$\mathbf{H}(\tau, t) = \begin{bmatrix} h_{1,1}(\tau, t) & h_{1,2}(\tau, t) & \cdots & h_{1,N_t}(\tau, t) \\ h_{2,1}(\tau, t) & h_{2,2}(\tau, t) & \vdots & h_{2,N_t}(\tau, t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_r,1}(\tau, t) & h_{N_r,2}(\tau, t) & \cdots & h_{N_r,N_t}(\tau, t) \end{bmatrix} \quad (1.1)$$

In this chapter, as well as the rest of the thesis, we consider the case where elements of \mathbf{H} follow an i.i.d. Gaussian distribution with zero-mean and have independent real and imaginary parts each with variance $1/2$. Equivalently, each entry of \mathbf{H} has a uniform phase and Rayleigh magnitude. This choice models a Rayleigh fading environment with enough separation between the receiving antennas and the transmitting antennas such that the fades for each transmitting-receiving antenna pair are independent. In all cases, we will assume that the realization of \mathbf{H} is known as the receiver, or, equivalently, the channel output consists of the pair $(y; \mathbf{H})$, and the distribution of \mathbf{H} is known at the transmitter.

The vector $[h_{1,j}(\tau, t), h_{2,j}(\tau, t), \dots, h_{N_r,j}(\tau, t)]^T$ is referred to as the spatio-temporal signature induced by the j -th transmit antenna across the receive antenna array. Furthermore, given that the signal $s_j(t)$ is launched from the j -th transmit antenna, the signal received at the i -th receive antenna is given by

$$y_i(t) = \sum_{j=1}^{N_t} h_{i,j}(\tau, t) \times s_j(t) + n_i(t), \quad i = 1, 2, \dots, N_r. \quad (1.2)$$

where $n_i(t)$ is additive white Gaussian noise (AWGN) in the receiver.

1.3.2 Spatial Diversity Gain on MIMO system

Spatial diversity can be obtained by placing multiple antennas at the transmitter and/or the receiver (Figure 4.2). If the antennas are placed sufficiently far apart, the channel gains (or the varying channel impulse) between different antenna pairs fade more or less independently and independent signal paths are created [62]. The distance between two adjacent antennas determines interference between this pair and typical antenna separation of half to one carrier wavelength is sufficient. In fact, there are three types of antenna diversity: receive diversity (Single Input Multiple Output or SIMO channel), transmit diversity (Multiple Input Single Output or MISO) and diversity in both side (Multiple Input Multiple Output or MIMO channel). However, in this work, we focus on diversity in both sides. The details of SIMO and MISO can be found in [62].

Again, multiple antenna channel systems can provide *spatial diversity* which can be used to improve the reliability of wireless communication links. The diversity here

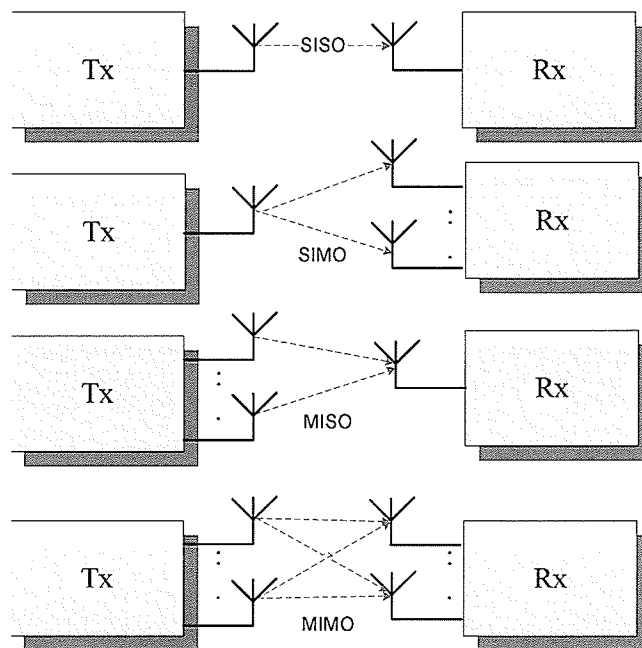


Figure 1.3. *SISO, SIMO, MISO, and MIMO transmission.*

can be seen as the number of independent duplicates of the same signal at the receiver side. The resulting signal at the receiver side can be demodulated and decoded in the usual way. Then the probability that all the signal components fade simultaneously is reduced. The results in [73] showed that at high SNR, the probability of error (averaged over the fading gain \mathbf{H} as well as additive noise) is much smaller than in the situation where only one antenna is used at the receiver side. This main achievement at high SNR, in terms of probability of error, is called *diversity gain*.

For a system with N_t transmit and N_r receive antennas, the *maximal diversity gain* provided by the channel is $N_t \times N_r$ assuming that perfect channel information is available at both sides. This information can be obtained by using pilot symbols at the transmitter and feedback from the receiver. In fact, extracting spatial diversity gain in the absence of channel knowledge at the transmitter is possible using suitably designed transmit signals. The corresponding technique is known as a space-time coding (STC) scheme [59]. The simplest STC scheme is the Alamouti scheme [7] where there are two antennas at both, transmitter and receiver, sides without any feedback from the receiver. This scheme has been proposed in several third generation cellular

standards for transmit diversity. As an extension of the Alamouti scheme, space time block coding (STBC) was introduced in [58], which is able to use an arbitrary number of transmit antennas and is able to achieve the full diversity promised by the transmit and receive antennas. Also, at the receiver side, maximum likelihood (ML) decoding is used to decouple the signals transmitted from different antennas and perfect CSI at the receiver side is assumed.

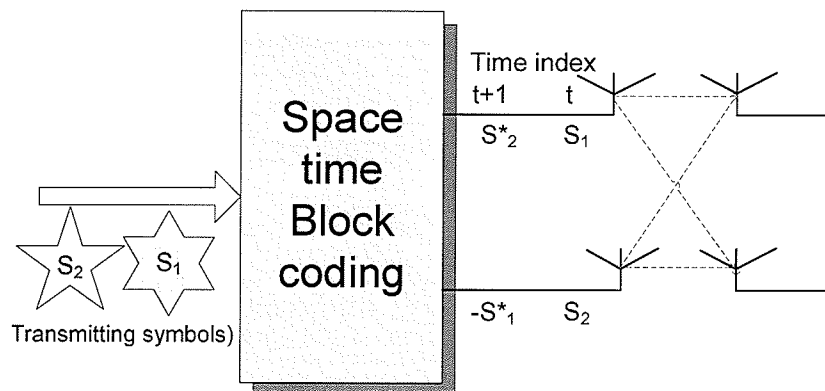


Figure 1.4. *Spatial Diversity transmits the same data stream across all antennas, and makes use of multipath to increase the chance of correctly decoding the received signal, i.e. lower bit error rate.*

1.3.3 Spatial Multiplexing MIMO Systems

Besides providing *diversity* to improve the reliability of wireless links, an alternative way of exploiting the multiple antenna elements is the so-called “*spatial multiplexing*” or BLAST approach [27]. The most well known model is V-BLAST [67], which is the first spatial multiplexing technique implemented in laboratory scenarios, and the principle of this model is shown in Figure 1.6. In this case, multiple antennas can support higher data rate than single antenna channels, or in other words, different data streams are transmitted (in parallel) from the different transmit antennas (Figure 1.7).

Another popular technique, with coding over signals transmitted on different antennas, is D-BLAST. In D-BLAST, the input data stream is transmitted on different antennas using time slots in a diagonal fashion. The advantage of this method is that

each individual sub-stream passes through all the sub-channels; hence if there is an error in one sub-channel, protected by code, it does not cause the loss of the stream [73]. However, the disadvantage of this method is the complexity in implementation. The receiver must demultiplex the signal in order to reconstruct the transmitted symbols. Multiple receive antenna elements are used for separating the different data streams at the receiver side. The results in [73] showed that the multiple antenna channel can be viewed as $\min(N_t, N_r)$ parallel spatial channels, so that the total number of *degrees of freedom* is $\min(N_t, N_r)$. Since one can transmit independent signals in parallel via multiple spatial channels, this method is called spatial multiplexing.

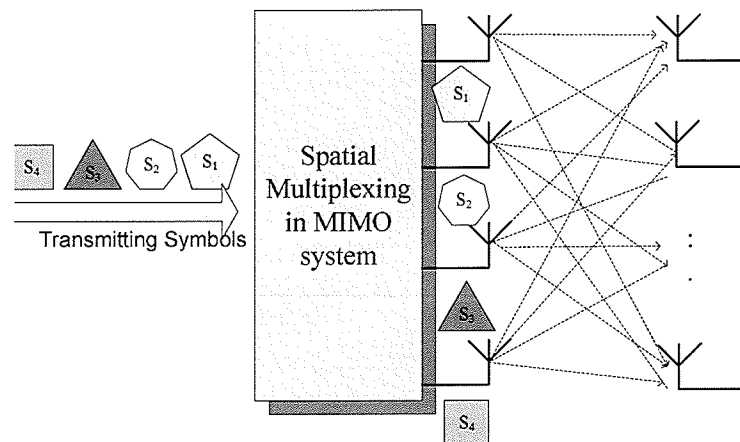


Figure 1.5. *Spatial Multiplexing transmits independent data streams in the same time slot and frequency band simultaneously.*

The advantage of this method is that the data rate can be increased by a factor of $\min(N_t, N_r)$ (i.e., either N_t or N_r) without requiring additional frequency spectrum. There are some practical schemes such as layered space time receiver structure [26] and space time codes [59] which allow us to achieve the highest performance in a MIMO wireless system.

Due to the fact that wireless technology is now moving towards the era where very high data rates are needed for multimedia services, the achievement of spatial multiplexing in data rate is significant to the wireless system designers. A practical MIMO system requires an algorithm to adapt the coding and transmission parameters to variations of the environment. This is called dynamic link adaptation in

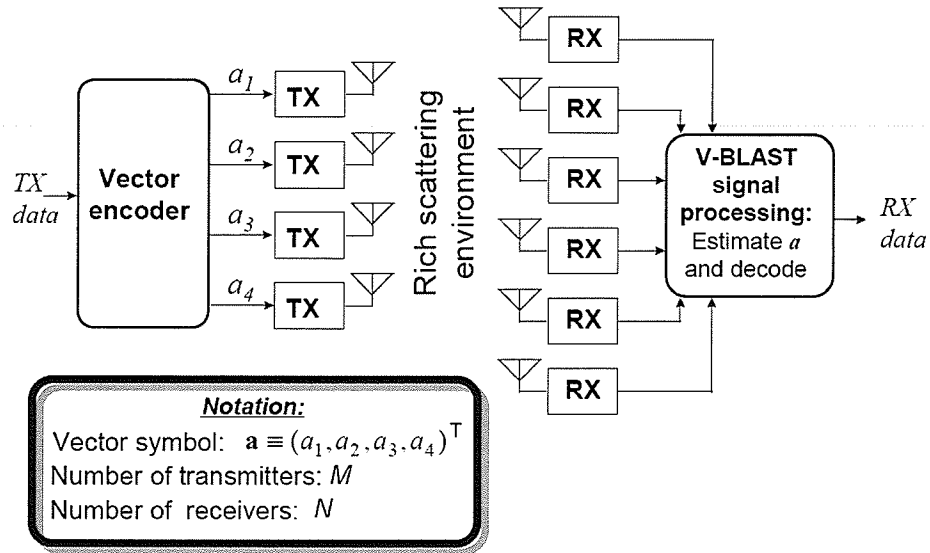


Figure 1.6. VBLAST system diagram.

which adaptive modulation is one of the most important solutions. Combining adaptive modulation with multiple antenna systems will require some feedback to change the modulation level at the transmitter. When the channel information state (CSI) is available at the receiver side (e.g., the VBLAST model), to achieve the highest possible modulation level, the post processing SNR of each stream is fed back to the transmitter [28], [42]. In this work, we focus on the case where CSI is available at both sides. In such a case, the MIMO channel can be decomposed into multiple separate channels by using the singular value decomposition (SVD) method. The details of this method and the design and analysis of adaptive modulation in a MIMO system in this particular case will be investigated later in this thesis.

1.4 Objective, Motivation, and Scope of The Thesis

A wireless mesh backbone network will require high throughput and reliability. Also, the QoS requirements for different users need to be satisfied. The use of multiple antennas at the transmitter and the receiver (i.e., MIMO wireless links) is an emerging cost-effective technology that offers substantial transmission rate in making high-

speed wireless connectivity in a wireless mesh backbone a reality. To recognize the significant improvement of using MIMO over the other schemes such as SISO, SIMO, MISO, capacity comparison of the these schemes was presented in [62] as shown in Figures 1.7 and 1.8. Indeed, we can see clearly that, the capacity and rate of a MIMO system is very large and there is a huge improvement compared to SISO, i.e., at moderate and high signal-to-noise ratio (SNR), the capacity of an $n \times n$ MIMO system is about n times the capacity of a SISO system.

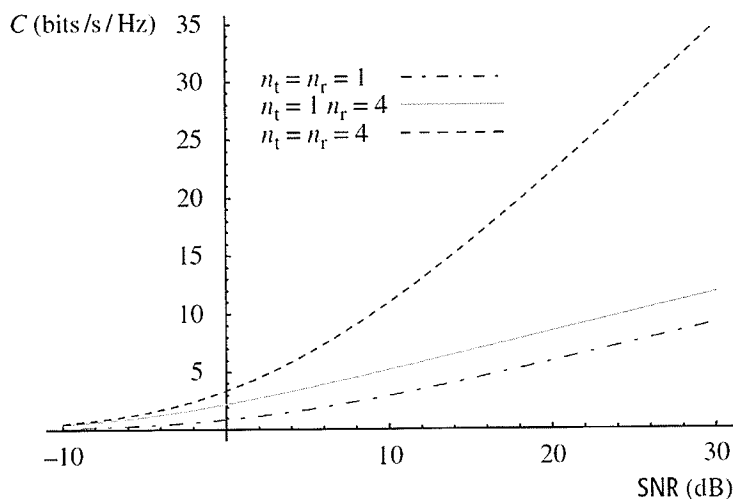


Figure 1.7. Capacity of an *i.i.d.* Rayleigh channel with different SNR.

In a WMN, the wireless mesh backbone, which provides multihop wireless connectivity to mesh clients, requires high throughput and reliability in order to provide quality of service (QoS) for real-time as well as non-real-time applications. To achieve high channel utilization and to guarantee the QoS requirements of different types of traffic, the radio resources such as the antennas and transmission power in a mesh router need to be allocated optimally among the different flows. Resource allocation and end-to-end performance analysis for MIMO-based wireless mesh backbone networks pose significant research challenges. In order to realize the full potential of MIMO technology, higher layer protocols must be designed to be cognizant of the MIMO link capability. In particular, channel state information (CSI) from the physical layer should be exploited for optimal resource allocation at the medium access

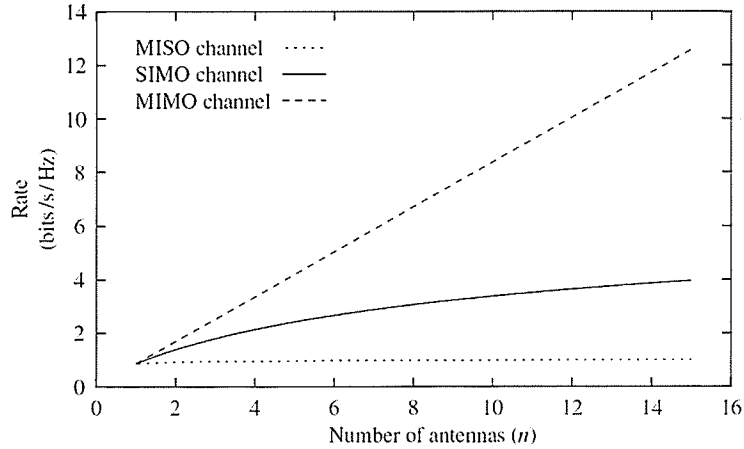


Figure 1.8. Capacities of $n \times n$ MISO channel, $1 \times n$ SIMO channel and $n \times n$ MIMO channel (for SNR = 0 dB).

control (MAC) layer.

The thesis focuses on the application of MIMO technology in WMNs. The problem of radio resource allocation at a mesh router is addressed. In particular, the problem of antenna selection and assignment (ASA) among different flows in a mesh router is solved. The solution can achieve a differentiated QoS among different flows in a mesh router. Also, to provide efficient channel utilization, the ASA scheme considers adaptive modulation and coding (AMC) to exploit CSI. The ASA solution works at the link layer to satisfy the QoS requirements for each type of users while taking full advantage of high-speed transmissions achievable through the MIMO technology.

The system model assumed here considers the following physical and link layer aspects:

- **Physical layer:** Singular Value Decomposition is used to decompose a MIMO channel into m independent channels. AMC is used to enhance the transmission rate according to CSI, and we use a finite state Markov channel (FSMC) model for different transmission modes of AMC.
- **Link layer:** An Automatic Repeat Request (ARQ) protocol to retransmit erroneous packets over a multi-rate wireless link is considered.

The ASA scheme, which uses only a subset of total available antennas for each type of services, is promising in order to 1) achieve fairness, 2) maximize the through-

put, and 3) guarantee the QoS requirements for different types of traffic. This scheme is developed based on a Markov Decision Process (MDP) formulation of the antenna assignment problem. The MDP formulation exploits a queueing analytical model for the data queues at the transmitter node. The performance of the proposed scheme is compared with the traditional weighted round-robin type of schemes for antenna scheduling. Numerical results are presented to demonstrate the efficacy of the proposed scheme.

To this end, for performance analysis of a wireless mesh backbone in an end-to-end MIMO transmission scenario, a tandem queue model is developed. Both exact and approximate decomposition approaches are proposed to solve the tandem queueing problem. The analytical model enables us to quantify the differences in end to end performance between SISO and MIMO scenarios in a wireless mesh backbone.

Note that, most of the works on MIMO in the existing literature have focused on the physical layer aspects. Recent research on MIMO communication systems has mainly focused on either increasing capacity by employing spatial multiplexing (SM) or mitigating fading by employing spatial diversity of a single wireless link. Understanding the impacts of MIMO technology on higher layers in a multihop wireless network, and in particular, the development of link adaptation, scheduling, and re-transmission algorithms that make explicit the use of MIMO nature of the system is of significant interest.

1.5 Organization of The Thesis

The organization of this thesis is as follows:

- **Chapter 2:** provides a background and discusses several key issues and techniques for resource allocation and management in a MIMO-based wireless network in the literature. The issues and design approaches for MAC schemes in MIMO wireless networks are provided.
- **Chapter 3:** presents a service differentiation model for the singler user scenario where wireless clients considering two types of traffic, namely, QoS-sensitive and best-effort traffic. AMC is used at the physical layer to increase the transmission rate by exploiting the dynamic channel variations. An optimization problem is

formulated and a solution based MDP is obtained.

- **Chapter 4:** presents a tandem queueing model for multi-hop communications in MIMO-based wireless mesh backbone networks. Both exact and approximate decomposition approaches are presented to solve the tandem system of queues.
- **Chapter 5:** summarizes the contribution of the thesis and outlines several directions for future research.

Chapter 2

Resource Allocation in MIMO-Based Wireless Networks

Radio resource management is one of most important components in wireless network design. Traditionally, in telecommunication contexts, resource sharing models are called trunk reservation models due to the most often used policy for admitting new customers based on thresholds depending only on the currently available resources [31]. In a wireless network, the main objective of radio resource allocation is to maximize the number of concurrent transmissions based on currently available resources, so that the throughput can be maximized. Note that, packet-level performances in a wireless network depend not only on the resource sharing mechanism among multiple users, but also on the radio link level error control mechanisms.

Radio frequency spectrum is the most scarce resource for wireless communications. In a multiple access wireless communication environment, many users may have to share a limited amount of bandwidth. At the same time, to meet the rapidly growing demand for different wireless communications services (e.g., video on demand, online gaming), significant efforts are being made towards efficient radio resource management to improve the wireless spectrum efficiency.

Although power consumption and node mobility, which are two major issues in an ad hoc network, have been removed in a wireless mesh backbone network, efficient radio resource allocation in a wireless mesh backbone network still remains as a major issue. To maintain wireless connectivity in the network, wireless mesh routers in a wireless mesh backbone should not only send their own packets, but also forward packets of other nodes. To guarantee the end-to-end QoS, resource allocation controllers should decide how much of the resources (e.g., number of antennas,

transmission power) must be allocated to each traffic path within the wireless mesh backbone. Since a wireless mesh backbone does not have a centralized “base-station” which coordinates with the mesh routers, the wireless mesh routers are expected to have the characteristics of self-organization and auto-configuration. These characteristics offer many benefits such as low upfront investment, increased reliability and scalability.

In the following, we discuss several key issues and techniques for resource allocation and management in a MIMO-based wireless network.

2.1 Medium Access Control Protocol in MIMO Wireless Networks

2.1.1 Centralized Medium Access Control in MIMO Wireless Networks

A multiple access control (or medium access control) technique allows different users to share the transmission media and it has significant impact on the higher layer protocol performance. In centralized scenarios, there are three fundamental multiple access methods: i.e., time-division multiple access (TDMA), frequency-division multiple access (FDMA), and space-division multiple access (SDMA). In a MIMO transmission scenario, TDMA can be implemented by assigning the entire time slot to only one user during one scheduling period as in [17]. In this work, a multiuser MIMO system was considered for downlink transmissions where each user is given an individual probability of outage constraint, defined as the probability that the short-term signal-to-noise ratio (SNR) at the receiver is smaller than a given threshold. An optimal solution for power allocation and time sharing among users was presented with the aim to minimizing the overall transmit power while meeting the user’s outage probability constraints by jointly optimizing the user’s power allocation and time-sharing (i.e., the number of time slots). By solving the MPE (minimum power equation) and the convex version of the original problem, an algorithm which can obtain a joint solution for both power and time allocation for the users has been proposed. Results have shown that the proposed method is nearly optimal. How-

ever, the MAC layer queueing dynamics and the interactions between the MAC and physical layers were not considered.

An FDMA-based resource allocation scheme for MIMO networks was considered in [38]. A spatially greedy scheduling scheme was used considering downlink transmissions in a MIMO wireless cellular packet data system. The scheduler at the base station decides to schedule transmissions to one or more users based on their current channel states with an aim to minimize average delays while maximizing the sum of allocated rates to the users.

SDMA plays an important role in MIMO transmissions. SDMA divides a geographical space, where the users are located, into smaller spaces. The key element of the design is a one-to-one map between the space divisions and the bandwidth divisions (in time slots, frequency divisions, etc). Therefore, it is important to consider how the users in the network will be grouped together. In particular, since the different spatial channels are nonorthogonal, it is critical that only spatially compatible users be chosen to be time- or frequency-coincident. However, SDMA requires channel state information (CSI) at both the transmitters and receivers. Since the receiving nodes cannot cooperate, transmitters have to ensure that data destined for one node do not interfere with the data to other nodes it is attempting to communicate with [56]. This is only possible if the transmitter can separate the users spatially, which in turn is only possible if their channels are known. Also, in ad hoc networks, all nodes can be transmitter or receiver at any time. Therefore, in order to implement SDMA, nodes which want to communicate in a given area have to ensure that the set of nodes in this area are in the receive mode.

2.1.2 Distributed Medium Access Control in MIMO Wireless Networks

In contrast to centralized scenarios, distributed scenarios requires little coordination and adopt contention-based multiple access schemes, such as ALOHA and carrier-sense multiple access (CSMA) protocols. In an ALOHA system, each user transmits when it has data to send, and then waits for an acknowledgement. If a collision occurs, the user backs off for a random period and retransmits the message. A decentralized random access (RA) strategy in MIMO systems, a simple variation of slotted ALOHA,

combined with OFDMA and SDMA was presented in [50]. In this strategy, users randomly access a subcarrier with an access probability that is designed to provide a desired number of multiple access so that by exploiting spatial processing with receive beamforming, the signals of the transmitting users can be separated and detected.

The operation of a wireless mesh backbone network is similar to a “static” wireless ad hoc network. When MIMO links are implemented in point to point communications in an ad hoc network, the number of successful transmissions at the receiver side depends on the factors such as the number of available antennas, channel/or sub-channel quality, post processing SNR etc. While a large amount of work in the literature focused on the PHY layer aspects of a MIMO communication system, only a few works addressed the MAC issues in a MIMO wireless network.

Research on wireless ad hoc networks based on the traditional IEEE 802.11 standard typically assumes the use of omnidirectional antennas at all nodes. A popular model that is often adopted at the MAC layer is the use of a four-way handshake (i.e. RTS, CTS, DATA, ACK) over single hop scenarios. When MIMO links are used in an ad hoc network, the MAC design becomes more complex due to the issues such as obtaining channel state information, maintaining channel state, trade-offs among rate, range, and reliability, and access methods [76].

For efficient MAC operations in a MIMO-based wireless network, a flexible physical layer is required which provides various modes of operations. This helps the MAC protocol to suitably choose from multiple modes. Also, a cross-layer approach to the MAC design can take full advantage of the lower layer, i.e., the physical layer. Furthermore, in order to match the requirement of unattended and decentralized architecture of an ad hoc network, this MAC mechanism has to be distributed in nature rather than to be centralized. The importance of cross layer design in MIMO links based ad hoc networks is shown in Figure 2.1, and more details can be found in [4].

In this section, we will review some of the works on joint design of physical layer and MAC layer protocols for MIMO ad hoc networks using both spatial multiplexing and diversity multiplexing modes.

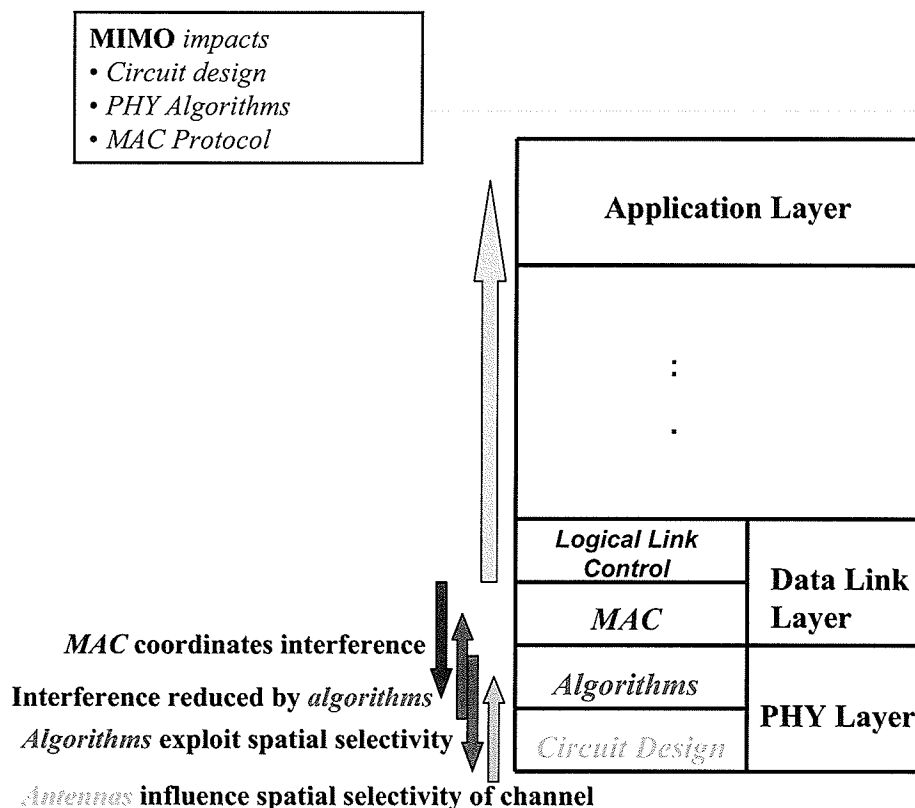


Figure 2.1. Joint design of PHY and MAC layers for MIMO links in ad hoc networks.

2.1.2.1 Medium Access Control Exploiting MIMO Spatial Diversity

Spatial diversity (SD) MIMO can improve the robustness of transmission and increase the coverage range of MIMO communications. SD is particularly useful for ad hoc networks due to the fact that mobility of users in an ad hoc network may deteriorate channel quality and the reliability of communications.

In [37], an SD-MIMO MAC based on the RTS/CTS mechanism of IEEE 802.11 distributed coordination function (DCF) was proposed for an ad hoc network. Here each node is equipped with M element antennas, thus there are M^2 degrees of freedom for communication. Assuming that space time codes are used for four-way handshaking, the proposed scheme can achieve the full order spatial diversity. This SD-MIMO

MAC is mostly similar to the CSMA/CA algorithm (RTS/CTS/DATA/ACK). However, in the multiple antenna scenarios, the threshold to decide on the status of the channel (idle or busy) is based on the average interference power across receive antennas. Also, it takes into account the impact of spatial diversity (i.e. encoding and decoding by using space time code) on overhearing and also multi-rate transmission for DATA packets according to channel conditions. The MAC layer scheduling and QoS issues were, however, not considered.

2.1.2.2 Medium Access Control Exploiting Spatial Multiplexing MIMO

As mentioned before, spatial multiplexing (SM) MIMO increases the transmission capacity by transmitting independent data streams in the same time slot and frequency band simultaneously from each transmit antenna. In this case, multiple data streams are differentiated at the receiver using channel information about each propagation path.

The fundamental objective of a MAC protocol is to avoid collision. In a traditional network using omnidirectional antennas, a source node broadcasts the RTS message. If the destination node agrees on this data transmission by sending back a CTS message, then all nodes in the range of this RTS/CTS exchange will refrain from transmissions during the data transmission period. This reduces the interference for the ongoing data transmissions. However, when we consider MIMO-based ad hoc networks, multiple RTS/CTS messages can be implemented simultaneously instead of blocking all the nodes. This is because, MIMO technology can enable multiple parallel transmissions in the same area by using proper channels. In other words, a destination node can accept many RTSs in the same time. Then multiple flows can be transmitted as separation is achieved in spatial domain.

In [13], the Distributed Scheduling for MIMO Ad hoc networks (DSMA) algorithm was proposed in which RTS/CTS exchanges were used to let the destination node know about the overall traffic condition in the network so that it can decide how many and which transmissions will be accepted via CTSs. A cross layer design model was presented where the physical layer and MAC layer directly exchange information to achieve better performance in the decoding process at the receiver. By considering a well known spatial multiplexing model, namely, the LAST-MUD (Layered SpaceTime

MultiUser Detection) model [55], which takes full advantage of rich scattering in a MIMO environment, to separate superimposed incoming transmissions, then a high data rate could be achieved. Since the superimposed incoming transmissions can be separated in the receiving nodes, the existence of multiple transmit links becomes completely transparent. In the MAC layer, RTS messages are first sent to request transmissions. Then the receivers respond with CTSs if they are available. Since SM allows multiple flows to be transmitted at the same time, RTSs may be addressed to multiple nodes. Since receiving nodes may receive a higher number of RTS messages than their capacity (depending on number of available antennas at the receiving nodes) they have to decide which flows will be accepted or denied and this information will be embedded into CTSs.

Instead of leaving the nodes to access the channel in an uncoordinated manner, the receivers are able to control the state of other nodes in the following frame. By scanning the backlog queue before composing the ACK (which will be sent after successful reception), the receiver nodes can embed the destination address in ACK to the nodes which they expect to receive from in the following frame.

2.2 Power and Rate Control in MIMO Wireless Networks

In a wireless network, mobile devices usually rely on a battery with a limited amount of energy. Therefore, minimization of transmission power can lead to efficient utilization of battery energy and hence longer battery life of mobile devices. Transmission power is thus one of the most critical resources in a wireless network. It can be conserved by controlling the transmission power of packets and/or by putting inactive nodes to *sleep* mode. It was shown in [20] that a MIMO-based transmission consumes much less transmission power than a SISO-based transmission for the same throughput. The transmission power determines the range over which the signal can be coherently received, and is therefore crucial in determining the performance of the network (throughput, delay, and energy consumption). For example, when the channel state of an individual channel is improved, the transmitter has to raise the transmission power to increase the data rate. Of course, rate adaptation can be also

achieved through a combination of variable spreading, coding, modulation, and code aggregation. But it was shown in [30] that, in the presence of channel state information at the transmitter as well as the receiver, by controlling both the rate and the power, the capacity of the channel can be significantly improved.

In a MIMO system, the MIMO links can be viewed as multiple parallel deterministic AWGN channels each corresponding to a block transmission under a specific channel state. With full availability of CSI, the capacity of the block fading channel can be found through optimal rate and power allocations for parallel AWGN channels based on their individual channel states. However, allocation of power and rate in fading channels is usually performed under some constraints imposed by practical considerations. For instance, the trade-off between the average transmission power, delay, and packet dropping probability for different transmission models over a fading channel with memory was presented in [22]. Using a Markov Decision Process (MDP) formulation, an optimal solution was obtained by using a relative value iteration and linear programming for both unconstrained and constrained problem, respectively.

As an extension of the above work, an approach based Q-learning (which is a recent form of Reinforcement Learning algorithm that does not need a model of its environment and works by learning an action-value function that gives the expected utility of taking a given action in a given state and following a fixed policy thereafter) was presented in [23] to solve the problem of rate and power adaptation under delay constraints. The problem was formulated as a CMDP and the solutions was obtained in an on-line fashion. This approach has an added advantage that transmission adaptation actions (that have to be negotiated between the transmitter and the receiver) can be performed less frequently than the power control actions. Also, rate control actions can be based on a more coarse quantization of the channel state than the power control actions.

2.3 Admission Control in MIMO Wireless Networks

Admission control is necessary in a wireless network to provide quality of service (QoS) for users. In wireless data systems, admission control is a challenging problem.

In particular, multimedia traffic can be very bursty, and is highly heterogeneous in terms of the QoS requirements, ranging from a small text message with weak requirements to high-data-rate video streaming with stringent delay requirements. Furthermore, wireless channels often exhibit time-varying fading. Such high variations in both channel conditions and traffic flows make admission control very challenging in wireless multimedia networks.

In a MIMO-based ad hoc network, there could be many ongoing flows at the same time, and admission control would be required so that the limited amount of transmission resource is not overwhelmed by too many ongoing connections. Admission control in an ad hoc network should be performed to achieve end-to-end QoS guarantee for different connections. A number of works in the literature focused on admission control in wireless mesh/ad hoc networks [37], [11]. However, the existing algorithms are not suitable for a multiple antenna environment. In [70], authors analyzed the probability density function of interference in a MIMO-based virtual group cell system. Based on this information, an interference-based admission control strategy with multi-level threshold was proposed. By using a multi-dimensional Markov model, performance of the proposed admission control strategy was evaluated. Another work in this area can be found in [49]. In this work, a game-theoretic model based on Q-learning was presented for distributed admission control in IEEE 802.11n based mesh networks using MIMO-OFDM technology. The objective of the admission control strategy is to maximize the utilities of all the routers along a routing path. The Nash equilibrium of the game was used to make the admission control decision.

2.4 Antenna Selection and Assignment in MIMO Wireless Mesh Networks

The transmit/receive antennas are one among the major resources in a MIMO mesh network. In a wireless mesh router, the number of antennas allocated to a flow for transmission determines the transmission rate for that flow, and therefore, determines the QoS performance (e.g., packet delay, packet loss) for that flow. The antenna allocation problem is therefore significant in the context of QoS provisioning in a

WMN. An antenna allocation scheme should consider the physical layer as well as the radio link layer parameters to achieve an efficient solution. That is, a cross-layer approach should be used to design an antenna selection and assignment scheme so that the spectrum utilization can be maximized as well as the QoS requirements can be satisfied for the different flows. The antenna selection and assignment (ASA) problem is focused in this thesis. A solution to this problem is presented in Chapter 3.

2.5 Cross-Layer Design in MIMO Systems: Perspectives and Challenges

In Section 2.1, the importance of cross-layer design for resource allocation in MIMO systems was highlighted. In this section, we will discuss the perspectives and challenges on cross layer design in MIMO systems. The objective of cross-layer design is to take full advantage of information at the different layers to achieve an efficient management of system resources. This is more complex and difficult in MIMO systems where multiple antennas are used to independently transmit data to increase the data rate (SM) or improve the reliability of data transmissions (SD). For instance, in the next section we will model a particular system where there are multiple antennas at the transmitter and pose the antenna assignment problem for different services (or users) as a MAC design problem. The data rate of each antenna remains however as a physical layer issue.

2.5.1 Complexity in Neighbor Discovery

CSI is the most important factor to decide the channel capacity of MIMO links. The most common way to obtain this information is by using the pilot signal which is embedded in the RTS messages. However, in a MIMO-based ad hoc network, MIMO can be used to increase the transmission range for a fixed data rate [29], specially when using a directional antenna. In an ad hoc network environment, some users may use omnidirectional antennas and the transmission range of different types of antenna could be different. Thus, identifying the correct set of reachable neighbors

with MIMO transmission would be very challenging. This issue is illustrated by a simple example in Figure 2.2.

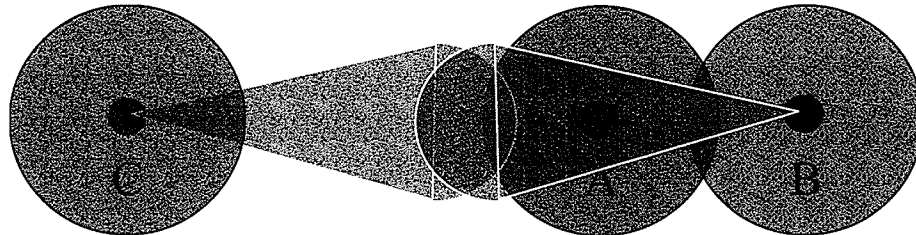


Figure 2.2. *Complexity in neighbor discovery.*

In this example, nodes A and B are neighbors when they are using an omnidirectional antenna or a directional antenna. Node C and A are not neighbours in omnidirectional antenna but they are neighbours when node C is using directional antenna and of course it is not B's neighbour when node B is using omnidirectional antenna. But they are neighbors when both node C and B are using directional antennas.

2.5.2 Difficulty in Obtaining the Channel State Information

Most of the works in literature assume perfect Channel State Information (CSI) knowledge at the transmitter. But in fact this information is rarely available because the random time-varying wireless medium makes it difficult and often expensive to obtain perfect CSI. Traditionally, receivers obtain CSI by using the training symbol (usually un-coded signal) embedded in the transmitted signal. Then they send this information back to the transmitter by using a perfect feedback channel. In MIMO closed-loop systems, CSI is degraded by the limited feedback resources, associated feedback delays, and scheduling lags, especially for mobile users with a small channel coherence time [35]. In MIMO open-loop systems, antenna calibration errors and turn-around time lags again limit CSI accuracy [54]. Therefore, the transmitter often only has partial channel information. Schemes exploiting partial CSI thus are both important and necessary. However, in MIMO systems where the high data rate is

exploited, the channel conditions during the time when the control message is sent back to the transmitter may quickly become obsolete. In an ad hoc network with high node mobility, obtaining the CSI in high data rate environment becomes very challenging. High mobility causes larger Doppler spread and faster channel time variation. In other words, larger Doppler spread results in higher temporal selectivity. A measure of the temporal selectivity is the channel coherence time, defined as the time interval over which the channel remains strongly correlated. The shorter the coherence time, the faster the channel changes with time.

2.5.3 Different Performance Objectives and Different MAC Design

Results in [73] emphasized the tradeoff between spatial multiplexing (transmit independent and separate stream in each antenna) and spatial diversity (transmit same information over multiple antenna) at the MAC layer. Depending on the objective of system design (reliability or high data rate) we can decide which model would be more suitable. With diversity, due to multiple replicas at the receiver side, BER at receiver side will be decreased which will result in improved transmission reliability, transmission range, and reduction in interference. With spatial multiplexing, a number of parallel independent streams can be sent at the same time, thus increasing the data rate. However, with a higher data rate, the reliability of data is decreased (due to the collision or loss of data).

2.5.4 Difficulty in Optimizing Resource Allocation in a High Mobility Environment

CSI is the most important parameter to recognize the network capacity and the performance of the network. Based on the achievable link SNR, a decision will be made to decide how many packets will be allocated to a specific link and the amount of allocated resources will thus be determined. This decision making has to be dynamic due to the fact that the channel state (in high mobility environment) in each sub-channel or each link varies dynamically, and a static scheduling scheme will not give the best performance. Also, in a MIMO-based wireless mesh/ad hoc network, the

fairness among data flows needs to be carefully considered to avoid spatial bias due to which users far from the gateway router may receive smaller amounts of resources compared to the users closer to the gateway router.

Chapter 3

Quality of Service in MIMO Wireless Mesh Networks: An Antenna Selection Approach

3.1 Introduction

In this chapter, we present a distributed antenna selection method for radio resource allocation for QoS provisioning in a mesh router. While using the MIMO links increases network throughput and connection reliability, the distributed scheme provides scalability and auto-configurability. The antenna selection scheme uses stochastic control algorithms for instantaneous delay control for a service differentiation model in a wireless mesh backbone using MIMO transmission. Each mesh node in the network is equipped with multiple antennas, and adaptive modulation is used in transmission links in a single-hop wireless scenario. In the physical layer, the MIMO channel considered here is a point to point wireless channel with t transmit and r receive antennas. The model considers two types of traffic, namely, QoS-sensitive traffic for real-time multimedia applications and best-effort (BE) traffic for applications such as web and e-mail. Two separate queues are used to accommodate the aggregated traffic from the QoS-sensitive and best-effort flows. This configuration is compatible with the *DiffServ* [12] model for service differentiation in which one queue is used for QoS-sensitive flows and another one is used for best-effort flows. Moreover, the multi-rate transmission feature in the physical layer, which can be achieved through AMC, is also taken into account and the multi-rate transmission is captured by using

multi-state Nakagami- m fading channel (fading occurring for multipath scattering with relatively larger time-delay spreads, with different clusters of reflected waves, the special case $m = 1$, the distribution reduces to Rayleigh fading). To capture the variations of the multi-state Nakagami fading channel, we employ the FSMC model [64]. The main objective of the proposed radio resource allocation (i.e., antenna selection and assignment (ASA)) approach is to maintain fairness of instantaneous delays between the flows. The antenna selection problem is presented as a stochastic optimization problem and we are able to obtain optimal solution for antenna allocated for each flow subject to the instantaneous delay fairness for traffic in both queues. Due to the consideration of instantaneous delays, which is represented by buffer occupancy, the problem is a dynamic stochastic optimization problem and can be stated as a Markov Decision Process (MDP).

The rest of the chapter is organized as follows. A survey of related works in the literature is presented in Section 3.2. The system model, assumptions, the problem definition and solution methodology are presented in Section 3.3. Section 3.4 presents the MDP formulation for the antenna selection and assignment problem and the corresponding solution. Simulation and numerical results are presented in Section 3.5.

3.2 Related Work

In [69], the problem of subcarrier and power allocation problems in a MIMO-OFDMA system was investigated in order to maximize the total system capacity subject to the total power and proportional rate constraints of each user. A greedy algorithm was proposed for resource allocation, which makes good use of the multi-antenna average channel gains and adopts an equal power allocation scheme to determine the number of subcarriers for each user. To reduce the complexity in resource allocation, subcarrier and power are allocated separately. Two main steps for subcarrier allocation is that the algorithm determines the number of subcarriers for each user by using a greedy-like scheme in the first step and assigns the subcarriers for each user by dividing the users into two groups in the second step. Power allocation among the assigned subcarriers for each user then is decided by using the multi-dimension

water-filling method.

An adaptive resource allocation algorithm for multiaccess MIMO/OFDM was proposed in [72] where the objective of the proposed algorithm was to maximize the system power efficiency given the QoS requirements of users. To reduce the complexity, a local search method was also presented. However, issues such as packet retransmissions due to channel errors and the impact of fading channels were not taken into account. [45] studied a joint problem of subcarrier allocation and beamforming to increase the downlink capacity, given that the power distribution is predetermined. A heuristic algorithm that sequentially inserts users in subcarriers was proposed without guaranteeing the minimum achievable data rate of each user. When QoS constraints were considered, the problem was further simplified by not allowing the reuse of subcarriers.

In the context of maintaining fairness among nodes in a wireless MIMO system, an optimal scheduler was proposed in [8], which optimizes users' diversity over antennas and provides high throughput while servicing users in a fair manner. A user utility function and cross-layer scheduler design was presented as a general solution for the Generalized Assignment Problem (GAP). The utility function was defined to control the throughput fairness tradeoff in an adaptive and efficient way. A similar work in [18] which combines spatial multiplexing with multiuser diversity, develops an optimal cross-layer scheduling mechanism in order to maximize capacity or to support proportional fairness. For optimal antenna assignment, the Hungarian algorithm was considered to utilize the characteristics of MIMO systems by adopting the graph theoretical approach. However, the optimal assignment was only achieved when the number of antennas is equal to the number of active users.

A similar work in MAC protocol design, i.e. Mitigating Interference using MultipleAntennas (MIMA-MAC) for MIMO ad hoc networks has been proposed in [52]. In this work, MIMA-MAC employs spatial multiplexing, with antenna subset selection for data packet transmission, while using the Alamouti space-time code for control packet transmission, to mitigate interference from neighboring nodes, to guarantee fairness between the traffic flows, and to increase the number of simultaneous traffic flows, resulting in an increase in the total network throughput. However, transmitters in this model use only a single antenna, whereas a receiver uses multiple antennas

and somehow this system did not take the full advantage of MIMO. On the other hand, two most common techniques, i.e. adaptive modulation and coding (AMC) and ARQ-based error recovery in the link layer, haven't been taken into account.

The closest work to our work is [23]. In [23] the problem of optimal rate control for a MIMO system is formulated as a finite state, average cost CMDP and it has been proved that the optimal policy is monotone. The authors proposed modified Q-learning algorithms, which treat the structure of the Q-factors and the optimal policy as constraints, to solve the inner step of the Lagrange dynamic programming formulation. In particular, retransmissions are not allowed, and hence the channel state does not affect the packet departure rate. In other words, channel states and buffer states are completely decoupled.

3.3 System Model, Assumptions, Problem Definition and Solution Methodology

MIMO system which exploits multiple transmit and receive antennas as a means to increase the information data rate. By taking advantage of the independence of the fading statistics of different users, multiuser diversity (MD) can be exploited to increase bandwidth utilization and decrease delay through simultaneous transmissions to a number of users. In a simultaneous transmissions environment, the question is how to schedule transmissions for the active users so that bandwidth utilization is maximized while at the same time a high degree of service fairness is achieved.

We assume a number of users having the same traffic requirement in one class and suppose that traffic from N classes will share B (bandwidth) resources in a mesh router. Traffic of class i arrives according to a Poisson process with parameter ζ_i . Each user demands b_i resources, where b_i is an integer. All the resources taken by a class i user are released simultaneously after an exponentially distributed service time with parameter μ_i . States are denoted with $c = (c_1, \dots, c_N)$, where c_i signifies the number of class users in the system. We assume that $\sum_{i=1}^N b_i c_i = B$, i.e., there is no waiting room. Therefore, users who do not find sufficient resources are automatically blocked. There is a distributed controller at each mesh router which can reject arriving users based on full state information. Each user who enters into the system gives rise

to a reward of r_i which is its instantaneous delay. More details on the states and the rewards will be provided in Section 3.4. For the sake of simplicity, in the rest of the chapter, we assume that there are only two classes of users: high priority (QoS-sensitive) and low priority (best-effort).

3.3.1 Physical and Link Layer Model

3.3.1.1 Signal Model

We consider a spatial multiplexing due to the fact that this physical layer is computationally easier to our model and also adapts with our objective, increasing the throughput of Wireless mesh backbone, than spatial diversity which has been used for increasing the *reliability*(see [73] for more details about the tradeoff between these two types of gains). MIMO wireless mesh backbone system here is equipped with multiple antennas. The channel considered is assumed to be a block fading channel which remains static during a time slot of length T . A MIMO transmission model similar to that in [75] is considered in which a flat fading MIMO channel with M transmitted and N received antennas is represented by a matrix \mathbf{H} with size $M \times N$.

Perfect channel information is assumed to be available at the transmitter side and rate and/or power information can be fed back to the transmitter. Let x be the $M \times 1$ transmit signal vector, y be the $N \times 1$ received signal vector and the channel matrix \mathbf{H} be an $N \times M$ matrix composed of independent complex Gaussian random variables. The zero-mean AWGN vector at the receiver, denoted by n , has a covariance matrix equal to the identity matrix scaled by σ^2 . For simplicity, we assume $\sigma^2 = 1$ and the variance of each component of \mathbf{H} equals to 1. The total power available to the transmitter is denoted by P_T . Then the received signal vector y is expressed as follows:

$$y = \mathbf{H}x + n. \quad (3.1)$$

Using the Singular Value Decomposition (SVD) technique, the channel transfer matrix \mathbf{H} can be diagonalized as follows:

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H \quad (3.2)$$

where \mathbf{U} and \mathbf{V} are unitary matrices; \mathbf{D} is a $M \times N$ diagonal matrix containing the singular values of \mathbf{H} which are real and non-negative; $[\sim]^H$ denotes complex conjugate

transpose operator. Note that $\{\lambda_i\}_{i=1}^m$ are eigenvalues of $\mathbf{H}\mathbf{H}^H$. For convenience, we define the eigenvalue vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_m]$.

By applying the transmitting weight matrix \mathbf{U}^H at the transmit side and the \mathbf{V} receiving weight matrix at the receive side, up to $m = \min(M, N)$ orthogonal sub-channels can be realized. Let $\tilde{\mathbf{y}} = \mathbf{U}^H \mathbf{y}$, $\tilde{\mathbf{x}} = \mathbf{V}^H \mathbf{x}$, and $\tilde{\mathbf{n}} = \mathbf{U}^H \mathbf{n}$. Then, the original channel is equivalent to the following channel:

$$\tilde{\mathbf{y}} = \mathbf{D}\tilde{\mathbf{x}} + \tilde{\mathbf{n}} \quad (3.3)$$

where $\tilde{\mathbf{n}}$ has the same distribution as \mathbf{n} .

From (3.3) we can see that the channel matrix \mathbf{H} has been decomposed into $m = \min(M, N)$ parallel eigen subchannels since \mathbf{D} is diagonal. The equivalent channel input and output are $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{x}}$, respectively. The subchannel gains are represented by $\boldsymbol{\lambda}$ which constitute a random process due to the randomness of the channel entries of \mathbf{H} . Therefore, the channel matrix is decomposed into m independent sub-channel as follows:

$$\tilde{y}_i = \lambda_i \tilde{x}_i + \tilde{n}_i, \quad 1 \leq i \leq m. \quad (3.4)$$

The SVD decomposition can be interpreted as follows: if the input is expressed in terms of a coordinate system defined by the columns of \mathbf{V} and the output is expressed in terms of a coordinate system defined by the columns of \mathbf{U} , then the input/output relationship is very simple. Equation (3.4) is a representation of the original channel (3.1) with the input and output expressed in terms of these new coordinates, where each λ_i corresponds to an *eigenmode* of the channel (also called an *eigenchannel*, and for more convenience, since now we call *eigenchannel* and *sub-channel* interchangeably). Each non-zero subchannel can support a data stream; thus, the MIMO channel can support the spatial multiplexing of multiple streams.

The joint probability density function (pdf) of these unordered eigenvalues ($\boldsymbol{\lambda}$) is given by [60]

$$f_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}) = (m!K_{m,n})^{-1} e^{-\sum_i \lambda_i} \prod_i (\lambda_i^{n-m}) \prod_{i < j} (\lambda_i - \lambda_j)^2, \quad \lambda_1 \geq \lambda_2 \dots \geq \lambda_m \quad (3.5)$$

where $m \triangleq \min(M, N)$, $n \triangleq \max(M, N)$, and $K_{m,n}$ is a normalizing factor. According to [74], the performance enhancement offered by transmit-beamforming in

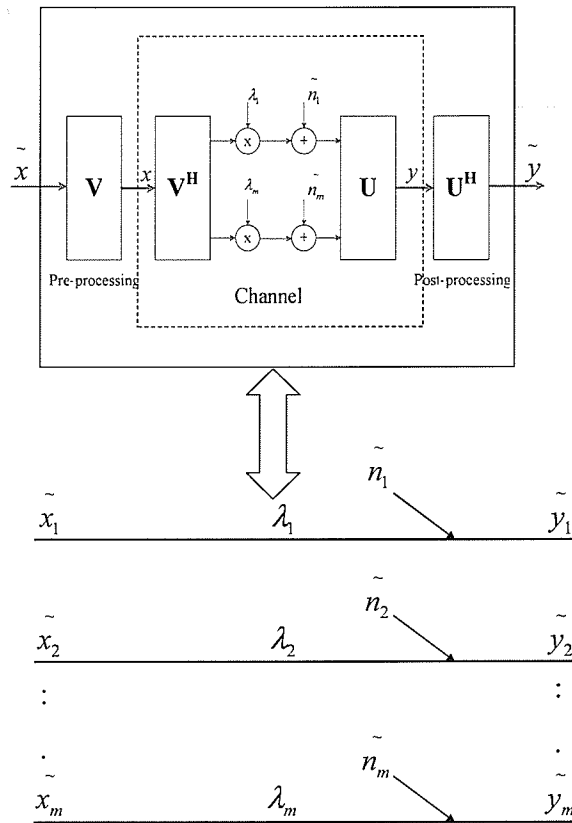


Figure 3.1. Converting a MIMO channel into a parallel eigenchannel by using Singular Decomposition Value technique.

MIMO channels is realized through the largest eigenvalue, i.e., λ_1 . Then the density of unordered eigenvalues can be expressed in closed form as

$$f_{\lambda_1}(\lambda_1) = \frac{1}{m} \sum_{i=0}^{m-1} \left(\frac{i!}{(i+d)!} \right) \sum_{l_1=0}^i \sum_{l_2=0}^i (-1)^{(l_1+l_2)} \times A_{l_1}(i, d) A_{l_2}(i, d) \lambda_1^{(l_1+l_2+d)} e^{-\lambda_1} \quad (3.6)$$

where $d \triangleq n - m$, and $A_l(i, d) \triangleq \frac{(i+d)!}{(i-l)!(d+l)!}$. Based on $f_{\lambda_1}(\lambda_1)$, let us define the following integral:

$$\Psi_1(c, x) \triangleq \int_x^{\infty} e^{-(c-1)\lambda_1} f_{\lambda_1}(\lambda_1) d\lambda_1 \quad (3.7)$$

Note that when $c = 1$, equation (3.7) becomes the cumulative distribution function

(cdf) of λ_1 . Therefore, the CDF of the largest eigenvalue can be expressed as follows:

$$\begin{aligned}\Psi_1(1, x) &= \int_x^\infty f_{\lambda_1}(\lambda_1) d\lambda_1 \\ &= \frac{1}{m} \sum_{i=0}^{m-1} \left(\frac{i!}{(i+d)!} \right) \sum_{l_1=0}^i \sum_{l_2=0}^i (-1)^{(l_1+l_2)} \\ &\quad \times A_{l_1}(i, d) A_{l_2}(i, d) \lambda_1^{(l_1+l_2+d)} e^{-\lambda_1} \Gamma(l_1 + l_2 + d + 1, x)\end{aligned}\quad (3.8)$$

where $\Gamma(n, x)$ is the upper incomplete Gamma function which is defined as

$$\Gamma(n, y) = \int_x^\infty y^{(n-1)} e^{-y} dy. \quad (3.9)$$

3.3.1.2 Channel Model

We will adopt K square quadrature-amplitude-modulation (QAM) constellations with size $M_k = 2^k$, $k = 1, \dots, K$ as the component modulation formats in each sub-channel. Generally, for an un-coded QAM with square constellation, such as 4-QAM, 16-QAM, the bit error rate (BER) expression over SISO Gaussian channels can be approximated by

$$BER_i \approx 0.2 \exp \left[\frac{-1.5\gamma_i}{2^k - 1} \right] \quad (3.10)$$

where γ_i is the received SNR (signal to noise ratio) per symbol of sub-channel i and k is the used modulation level. However, in practical systems, we will try to maintain a target BER for all sub-channels and this parameter mostly will be given. Let us assume that the transmit power at each sub-channel is P_i , and to reduce complexity, we allocate the total transmit power equally to each sub-channel. That is, $P_i = \frac{P}{m}$, where P is the total transmit power. For each channel realization, the received SNR can then be calculated as

$$\gamma_i = \frac{P\gamma_i}{m\sigma^2} = \gamma_0 \lambda_i. \quad (3.11)$$

Similar to single input single output (SISO) systems, MIMO systems need a set of coding and transmission parameters along with an algorithm to adapt these parameters to the variations of the environment. Such sets are often referred to as link adaptation modes. The employment of link adaptation techniques through AMC at the physical layer to enhance the spectral efficiency is very common in MIMO wireless systems [75] -[42]. Depending on the channel quality, the transmission mode

at the transmitter is adapted accordingly and, of course, this process essentially results in multi-rate transmission on wireless links. Basically, each modulation and coding scheme (MCS) is called one mode and corresponds to one particular interval of signal to noise plus interference ratio (SINR). Specifically, the SINR at the receiver is partitioned into a finite number of intervals with threshold values $X_0(= 0) < X_1 < X_2 < \dots < X_{K+1}(= \infty)$. Let X be the received SNR at the receiving side, transmission mode k is employed if $X_k \leq X < X_{k+1}$ ($k = 0, 1, 2, \dots, K$) which will be called channel state k in the sequel. For implementation, the receiver estimates the channel quality and transmits this channel state information (CSI) to the transmitter to choose the suitable transmission mode. Here we assume that the packet length is fixed. Then there are a finite number of transmission modes each of which corresponds to a unique modulation scheme [43]. These transmission mode thresholds can be obtained as follows:

$$X_k = \ln\left(\frac{BER_{tar}}{0.2}\right) \frac{(2^k - 1)}{-1.5} \quad (3.12)$$

where BER_{tar} is the target BER that we want to achieve in all sub-channels. The channel is said to be in state k if $X_k \leq \gamma < X_{k+1}$, and in this state, k bits are transmitted per symbol using 2^k -QAM which corresponds to transmission rate k , i.e. k packets will be transmitted when the channel is in state k . To avoid possible transmission errors, packets are not transmitted when $k = 0$.

From (3.8), the probability that modulation k is used in subchannel i can be obtained as follows:

$$\begin{aligned} P_r^{(i)}(k) &= P_r^{(i)}(\lambda_1 \in [X_k, X_{k+1})) = \int_{X_k}^{X_{k+1}} f_{\lambda_1}(\lambda_1) d\lambda_1 \\ &= \Psi_1(1, X_k/\gamma_0) - \Psi_1(1, X_{k+1}/\gamma_0) \end{aligned} \quad (3.13)$$

where $P_r^{(i)} = [P_r^{(i)}(0), P_r^{(i)}(1) \dots P_r^{(i)}(K)]$. For m available subchannels at MIMO link l , let $p(g)$ denote the probability that g packets are correctly received given by

$$p^{(l)}(g) = \bigodot_{i=1}^m P_r^{(i)} \quad (3.14)$$

where \bigodot denotes the discrete convolution and maximum number of packets that can be transmitted in one time slot over m subchannels is $N = m \times K$.

The number of transmitted packets can vary depending on the transmission rate and the packet error rate. According to results from [74], the average BER for modulation k can be approximated as follows:

$$\overline{BER}_k \approx a_k \left[\Psi_1(b_k \gamma_0 + 1, X_k/\gamma_0) - \Psi_1(b_k \gamma_0 + 1, X_{k+1}/\gamma_0) \right] \quad (3.15)$$

where a_k and b_k are the constant parameters depending on the modulation level k and can be obtained from table I in [32] when BER_{tar} is set to be 10^{-3} .

We assume that \overline{BER}_k is i.i.d with the others, packet size is PL (header + payload), and there are up to $PL_{correct}$ bits which can be corrected in a packet. Then the average packet error rate for modulation k can be obtained as follows:

$$\overline{PER}_k = 1 - \left(\sum_{i=0}^{PL_{correct}} \binom{PL}{i} (\overline{BER}_k)^i (1 - \overline{BER}_k)^{PL-i} \right) \quad (3.16)$$

Basically, we can calculate \overline{PER}_k over sub-channel i , since now, to reduce the complexity, we assume that \overline{PER} of MIMO link l is $\theta^{(l)}$.

Assuming that the channel is slowly fading (i.e., transitions occur only between adjacent states), the CSI is fed back to the transmitter to choose the suitable transmission mode. The state transition matrix for the FSMC can be expressed as follows:

$$\mathcal{H}^{(i)} = \begin{bmatrix} \xi_{0,0}^{(i)} & \xi_{0,1}^{(i)} & \cdots & 0 \\ \xi_{1,0}^{(i)} & \xi_{1,1}^{(i)} & \xi_{1,2}^{(i)} & \vdots \\ 0 & \ddots & \ddots & 0 \\ \vdots & \xi_{K-1,K-2}^{(i)} & \xi_{K-1,K-1}^{(i)} & \xi_{K-1,K}^{(i)} \\ 0 & \cdots & \xi_{K,K-1}^{(i)} & \xi_{K,K}^{(i)} \end{bmatrix}. \quad (3.17)$$

For the sake of simplicity, we assume when the channel is in state k , the transmitter transmits $b_k = k$ packets. Furthermore, we assume that $b_0 = 0$, i.e., there is no transmission when the channel state is 0 to avoid the high probability of transmission error), and $b_K = K$. To calculate the transition probability in matrix (3.17) from state k to k' ($k' \in \{k-1, k, k+1\}$), i.e. $\xi_{k,k'}^{(i)}$, we apply the results in [43], then we can obtain $\xi_{k,k'}$ for sub-channel i as follows:

$$\xi_{k,k+1}^{(i)} = \frac{N_{k+1}}{Pr^{(i)}(k) f_d} \quad (3.18)$$

$$\xi_{k,k-1}^{(i)} = \frac{N_k}{Pr^{(i)}(k) f_d} \quad (3.19)$$

$$\xi_{k,k}^{(i)} = \begin{cases} 1 - \xi_{k,k+1}^{(i)} - \xi_{k,k-1}^{(i)}, & \text{if } 0 \leq k \leq K \\ 1 - \xi_{0,1}^{(i)}, & \text{if } k = 0 \\ 1 - \xi_{K,K-1}^{(i)}, & \text{if } k = K \end{cases} \quad (3.20)$$

where f_d denotes the mobility-induced Doppler spread and N_k is the cross-rate of mode k [64]. Note that the differences among the transition probability matrix of these sub-channels is dependent on the average SNR, which is included in N_k .

3.3.2 Problem Definition and Solution Methodology

We consider a mesh router with two separate transmission queues: QoS-sensitive queue and best-effort (BE) queue. The transmission model between two routers in a wireless mesh backbone network is shown in Figure 3.2.

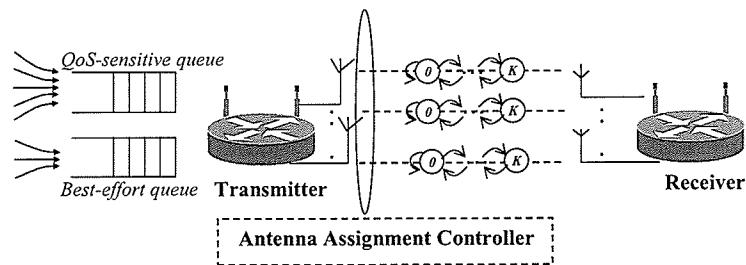


Figure 3.2. Transmission model between two mesh routers.

In this figure, the antenna assignment controller is responsible for assigning subsets of antennas to the two queues to differentiate services between these two queues. For this assignments it will exploit the estimated CSI in the physical layer. We can represent the operation of this antenna assignment controller as follows:

$$\begin{aligned} \text{Minimize:} \quad & J(w, z) := \lim_{n \rightarrow \infty} \sup \frac{1}{n} \sum_{t=0}^{n-1} \mathbb{E}^z [\mathcal{R}(s^{(t)}, a^{(t)})] \\ \text{Subject to:} \quad & d_1 = \Delta_2 d_2 \end{aligned} \quad (3.21)$$

where $\mathcal{R}(s^{(t)}, a^{(t)})$ is the reward of the whole system, included number of packets in both queues as well as channel states of m available subchannel, which is obtained by

giving the state s and taken action a at time slot t , and d_1, d_2 are the instantaneous delay of QoS-sensitive queue and best-effort (BE) queue, respectively. The details of the algorithm for this controller will be provided in Section 3.4.

At the receiver side, to recover the transmitted signals, we consider zero-forcing (ZF) detection [56] for its practicality and ability to exploit the MIMO channel into m parallel independent channels. When packets arrive at the receiver, it decodes the received packets and sends negative acknowledgments (NACKs) to the transmitter asking for retransmission if there are any erroneous packets. An error-free and instantaneous feedback channel is assumed here so that the transmitter knows exactly if there is any transmission error at the end of each service time slot. This assumption holds in many cases because the propagation delay and the processing time for the error detection code can be very small in comparison with the time slot interval.

The procedure to find the optimal solution for ASA is summarized as Figure 3.3.

Example: Considering a pair of mesh routers in a WMN network, we assume that both the transmitting node and receiving routers are equipped with 3 antennas thus there are 3 independent available subchannels to transmit data at the transmitter side. There are 2 types of traffic coming into transmitting node, i.e. real-time traffic and non-real-time traffic, with the average incoming packet is 1 packet/timeslot for both types of traffic. Each type of traffic is going into two separate queues. Real-time traffic and non-real time traffic require instantaneous delay (which is defined as the ratio between number of packets in queue and average incoming packets) of 1 time slot and 3 time slots (i.e. $d_1 = 1$ and $d_2 = 3$), respectively and the objective of fairness is to maintain the ratio $\frac{d_1}{d_2}$. We assume that at a particular time slot, among 3 available subchannels, 2 subchannels (subchannel 1 and 2) can serve 3 packets per time slot and the rest (subchannel 3) can serve 2 packets per time slot, thus total packets can be transmitted in this time slot is $3 + 3 + 2 = 8$ packets. The number of packets in real-time traffic queue and non-real-time traffic queue is 7 and 6, respectively. ASA can use all the available subchannels to transmit all packets from the real-time traffic queue. However, since the number of packets in real-time traffic queue is 7 and total link capacity is 8 packets, it causes wastage of bandwidth. Also, the QoS requirement for non-real-time traffic is not satisfied. A better allocation will be as follows: use subchannel 1 and 2 to serve packets coming from real-time traffic queue

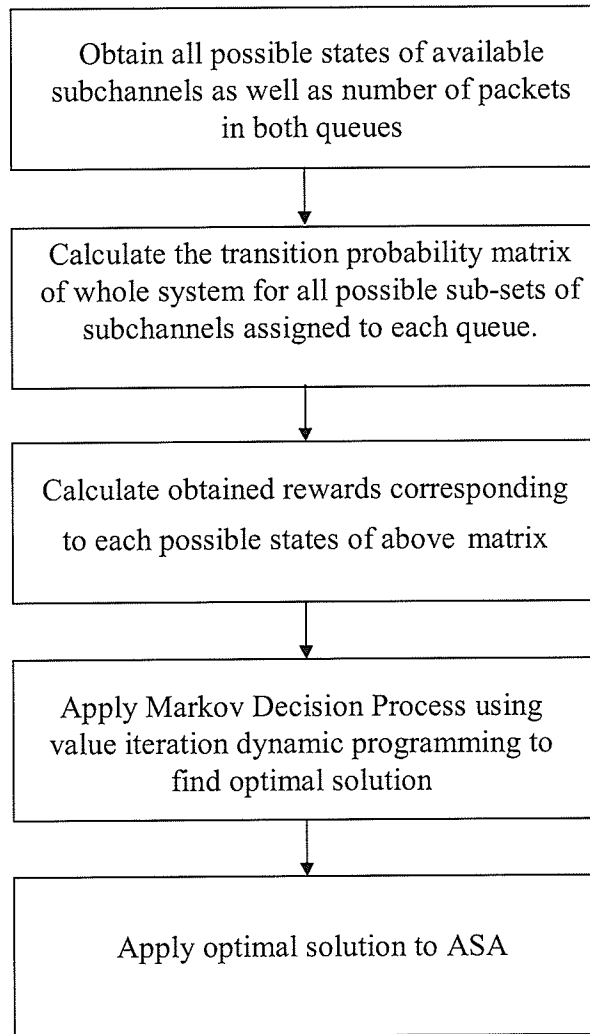


Figure 3.3. *The procedure to find the optimal solution for ASA.*

and use subchannel 3 for non-real-time traffic. Note that the instantaneous delay is calculated right after departure process (which is defined as the number of packets leaving queue). Then the instantaneous delay of real-time traffic is 1 time slot and non-real-time traffic is 4 time slots. Obviously, ASA do not really satisfies the exact ratio between the instantaneous delay of real-time traffic and non-real-time traffic but it nearly guarantees the fairness between these two types of flows.

Note that, the instantaneous delay d_i can be defined as [23]:

$$d_i = \frac{q_i}{\zeta_i} \quad (3.22)$$

where q_i is the number of packets left in queue i at the end of time slot t and $\bar{\zeta}_i$ is

the average number of incoming packets during time slot t which is obtained using (3.40).

3.4 Formulation of the Markov Decision Process Model

3.4.1 Packet Arrival Process

As has been mentioned before, to simplify the problem we consider only two types of traffic, i.e. real-time and non-real-time traffic, and the corresponding data packets are stored in two separate transmission queues. Let us assume that for each flow the packet arrival follows a truncated Poisson process, the packet arrival can be modeled as a Poisson process in which the probability of arrival can be obtained from (3.23), (3.24) and (3.25). The queueing problem for both classes of users can be modeled in discrete time with one time interval equal to one time slot. Real-time traffic will have higher priority (class one) than the other (class two). We assume that the packets arrive at the buffer according to independent Poisson processes with different packet arrival rates, i.e. ζ_c . The maximum number packet arrivals in one transmission queue in one time slot is A_c , where c is the index of classes of users. Packets of the same type are scheduled in a FIFO order. Then the probability of arrival of $a \in \{0 \dots A_c\}$ packets with mean ζ_c in time interval t is given by

$$f_a(\zeta_c) = \frac{e^{-\zeta_c t} (\zeta_c t)^a}{a!} \quad \text{for } a = 1, 2, \dots, A_c - 1 \quad (3.23)$$

$$f_{A_c}(\zeta_c) = 1 - \sum_{a=0}^{A_c-1} f_a(\zeta_c) \quad (3.24)$$

Then the arrival probability of each flow is obtained as follows:

$$\mathcal{AP}_c = \{f_0(\zeta_c), f_1(\zeta_c), \dots, f_{A_c}(\zeta_c)\}. \quad (3.25)$$

With this traffic modeling, there are at most A_c arriving packets in one time slot. Furthermore, packet transmissions in a time slot are assumed to finish before arriving packets enter the queue. We assume that the buffer size of each transmission queue is finite with the size of Q_c packets.

3.4.2 Packet Departure Process

Before we can calculate the packet departure process for each flow, we require to know the subset of antennas assigned to each flow. This antenna assignment will be performed by the Antenna Assignment Controller (AAC) which will be described in the next section. In a general system with C classes of users, let us assume that m_c is the set of antennas assigned to each class of users in the system, i.e. $\sum_{c=1}^C m_c = m$. Now, with only two types of users, let m_1, m_2 denote the subset of antennas assigned to class one and class two of services, respectively. It is clear that $m_1 \subseteq m$ and $m_2 = m \setminus m_1$. Moreover, if the number of packets in one queue is zero, another queue will automatically receive the entire service of all antennas. On the other hand, if both queues are backlogged, QoS-sensitive queue will have preemptive priority and the number of antennas allocated to it will be decided by the AAC. When class two packets return to the system, the service begins with a new independent service allocation.

Now let us define the following matrices:

- Let $\mathbf{H}_k^{(i)}$ ($k = 0, 1, \dots, K$) be the matrices of order $(K+1) \times (K+1)$ which are constructed by keeping the $(k+1)$ -st row of the channel transition probability matrix $\mathcal{H}^{(i)}$ in (3.17) and setting all other rows to $\mathbf{0}$, where i indicates the index of sub-channel. These matrices capture the number of packets leaving the queue in subchannel i when the channel state is k at the beginning of a particular time slot.
- Let $\mathbf{T}_{n,n'}^{(i)}$ be the matrices of order $(K+1) \times (K+1)$ where elements $(\mathbf{T}_{n,n'}^{(i)}(l_1, l_2))$ represent the probability that n' packets successfully leave the queue given that there were n packets in the queue where i indicates the index of the sub-channel, and channel state changes from state l_1 to state l_2 (l_1 and l_2 are indexes for the state associated with the changing in channel state of assigned set subchannels).
- Let $\mathbf{D}_t^{(m_c)}$ ($t = 0, 1, \dots, K \times m_x$) (where m_x is the number of antenna in the set of subchannels m_c) be the matrices of order $(K+1)^{m_x} \times (K+1)^{m_x}$ which are similar to $\mathbf{H}_k^{(i)}$, where t represents the number of packets leaving the queue using the set of subchannels m_c .

Then, we can calculate $\mathbf{D}_t^{(m_c)}$ and $\mathbf{D}^{(m_c)}$ as follows:

$$\mathbf{D}_t^{(12)} = \sum_{k=0}^t \mathbf{H}_k^{(1)} \otimes \mathbf{H}_{t-k}^{(2)} \quad (3.26)$$

$$\mathbf{D}_t^{(m_c)} = \sum_{\{k_1, k_2, \dots, k_{|m_c|} | \sum_j k_j = t\}} \bigotimes_{j=1}^{|m_c|} \mathbf{H}_{k_j}^{(j)} \quad (3.27)$$

$$\mathbf{D}^{(m_c)} = \bigotimes_j \mathcal{H}^{(j)} \quad (3.28)$$

- Let $\mathbf{T}_r^{(m_c)}$ be the matrices of order $(K+1) \times (K+1)$ whose elements $(\mathbf{T}_r^{(m_c)}(l_1, l_2))$ represent the probability that r packets are successfully leaving the queue when a set of sub-channels m_c is assigned to this queue, and channel state changes from state l_1 to state l_2 . It is clear that the maximum number of packets leaving the queue in a particular time slot in this set of subchannels is $|m_c| \times K$.

We can calculate $\mathbf{T}_r^{(m_c)}$ as follows:

$$\mathbf{T}_0^{(m_c)} = \mathbf{D}_0^{(m_c)} + \sum_{l=1}^{K \times |m_c|} p_{0,l}^{(l)} \mathbf{D}_l^{(m_c)} \quad (3.29)$$

$$\mathbf{T}_r^{(m_c)} = \sum_{l=1}^{K \times |m_c|} p_{r,l}^{(l)} \mathbf{D}_l^{(m_c)} \quad \text{for } l > 1 \quad (3.30)$$

$$(3.31)$$

Based on the average packet error rate, the probability that i packets are correctly received given that j packets are transmitted when channel state is k in one time slot can be obtained as follows:

$$p_{i,j}^{(k)} = \begin{cases} \binom{j}{i} \theta_k^{j-i} (1 - \theta_k)^i & \text{if } 0 \leq i \leq j \\ 0 & \text{otherwise} \end{cases} \quad (3.32)$$

where $\theta_k = \overline{PER}_k$.

3.4.3 State Space and Transition Probability Matrix

The state space of the proposed system model consists of buffer space \mathcal{Q} , incoming traffic space \mathcal{AP} , the channel state \mathcal{H} , i.e. $\mathcal{S} = \mathcal{Q} \times \mathcal{AP} \times \mathcal{H}$. In other words, the

discrete-time Markov Chain representing the system at time slot n has the following state space: $s_n = \{(q_n^{(c)}, ap_n^{(c)}, l_n^{(c)}), 0 \leq q_n^{(c)} \leq Q_c, 0 \leq ap_n^{(c)} \leq A_c, 0 \leq l_n^{(c)} \leq K\}$, where $q_n^{(c)}$ is the number of packets in the queue at the beginning of time slot n , $ap_n^{(c)}$ is the number of packet arrivals during time slot n , and $l_n^{(c)}$ is the set of channel states corresponding to the number of subchannels assigned to this flow during that timeslot, and let $l_n^{(x)}$ be the number of packets leaving queue corresponding to that assignment. The number of packets leaving the queue is equal to $\min\{q_n^{(c)}, l_n^{(x)}\}$.

Now, let us consider the transition matrix for the *QoS-sensitive queue*. The other matrix can be built in a similar way. Let \mathbf{P}_1 and (i, j, l) denote the transition matrix and a generic system state for this discrete time Markov Chain, respectively, what we have to obtain is the transition probability from state (i, j, l) to state (i', j', l') . For the same i and i' , the probabilities corresponding to these state transitions can be written in matrix blocks $\mathbf{p}_{i,k}$, which correspond to transitions in level i of the transition matrix. Thus, level i of the transition matrix represents the system state transitions where there are i packets in the queue before the transitions.

While the packet arrival probability is obtained from the truncated Poisson process (3.25) with maximum A_1 packets can arrive in one time slot, the departure process depends on the set of subchannels assigned to this flow, channel state of these subchannels during this time slot, and the packet error rate. With queue size Q_1 , the probability transition matrix \mathbf{P}_1 for the *QoS-sensitive queue* is defined as follows:

$$\mathbf{P}_1 = \begin{bmatrix} \mathbf{p}_{0,0} & \mathbf{p}_{0,1} & \cdots & \mathbf{p}_{0,A_1} & & & & & & \\ \mathbf{p}_{1,0} & \mathbf{p}_{1,1} & \mathbf{p}_{1,2} & \cdots & \mathbf{p}_{1,A_1} & & & & & \\ \vdots & \vdots & \ddots & \ddots & \ddots & & & & & \\ \mathbf{p}_{N,0} & \cdots & \mathbf{p}_{N,N-1} & \mathbf{p}_{N,N} & \mathbf{p}_{N,N+1} & \cdots & \cdots & \cdots & \cdots & \mathbf{p}_{N,\min\{N+A_1, Q_1\}} \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & \mathbf{p}_{i,i-N} & \cdots & \mathbf{p}_{i,i-1} & \mathbf{p}_{i,i} & \mathbf{p}_{i,i+1} & \cdots & \mathbf{p}_{i,\min\{i+A_1, Q_1\}} \\ & & & & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & & & & \mathbf{p}_{Q_1, Q_1-N} & \cdots & \mathbf{p}_{Q_1, Q_1-1} & \mathbf{p}_{Q_1, Q_1} \end{bmatrix} \quad (3.33)$$

where $N = m_1 \times K$ is the maximum number of packets that can leave the queue in one time slot. The matrix element $\mathbf{p}_{i,k}$ of matrix \mathbf{P}_1 is the probability that the number of packets in the QoS sensitive queue is i in the current time slot and it becomes k in the next time slot. Also, the elements inside the matrix $\mathbf{p}_{i,k}$ capture

the transitions in set of subchannels assigned to the queue at current time slot.

Now, the matrix blocks of the transition matrix in (3.33) can be written as follows:

$$\mathbf{p}_{i,j} = \sum_{\{a,r|u=j\}} f_a(A_c) T_r^{(mc)} \quad (3.34)$$

where $u = \min\{Q_1, \max\{0, i + a - r\}\}$.

It is clear that there are at most A_1 arriving packets, and at most N packets can be successfully transmitted in one time slot. Therefore, the transitions can go up by at most A_1 levels and go down by at most N levels. Note that, the maximum total packet transmission rate can be greater than the number of packets in the queue, and the decrease in the number of packets cannot be less than the number of packets in the queue. Therefore, the maximum number by which the number of packets in the queue can decrease is $N = \min(m_1 \times K, i)$.

Now, similar to the above process, with different assigned sets of subchannels to the *best-effort queue*, i.e., m_2 , the maximum number of packets leaving the queue is $M = m_2 \times K$. Then, with a different arrival rate $f_a(\zeta_2)$, and queue size Q_2 , the probability transition matrix \mathbf{P}_2 for the *best-effort queue* can be written as follows:

$$\mathbf{P}_2 = \left[\begin{array}{cccccccc} \mathbf{p}_{0,0} & \mathbf{p}_{0,1} & \cdots & \mathbf{p}_{0,A_2} & & & & \\ \mathbf{p}_{1,0} & \mathbf{p}_{1,1} & \mathbf{p}_{1,2} & \cdots & \mathbf{p}_{1,Q_2} & & & \\ \vdots & \vdots & \ddots & \ddots & \ddots & & & \\ \hline \mathbf{p}_{M,0} & \cdots & \mathbf{p}_{M,M-1} & \mathbf{p}_{M,M} & \mathbf{p}_{M,M+1} & \cdots & \cdots & \cdots & \mathbf{p}_{M,\min\{M+A_2,Q_1\}} \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & \mathbf{p}_{i,i-M} & \cdots & \mathbf{p}_{i,i-1} & \mathbf{p}_{i,i} & \mathbf{p}_{i,i+1} & \cdots & \mathbf{p}_{i,\min\{i+A_2,Q_1\}} \\ \hline & & & & & \ddots & \ddots & \ddots & \ddots & \\ & & & & & & \mathbf{p}_{Q_2,Q_2-M} & \cdots & \mathbf{p}_{Q_2,Q_2-1} & \mathbf{p}_{Q_2,Q_2} \end{array} \right] \quad (3.35)$$

where elements of the transition matrix in (3.35) can be defined similarly as in (3.34)

Based on (3.42) and (3.35), the transition probability matrix between the composite state

$$s_n = \{(q_n^{(1)}, ap_n^{(1)}, l_n^{(1)}), (q_n^{(2)}, ap_n^{(2)}, l_n^{(2)})\} \text{ and} \\ s'_n = \{(q_n'^{(1)}, ap_n'^{(1)}, l_n'^{(1)}), (q_n'^{(2)}, ap_n'^{(2)}, l_n'^{(2)})\}$$

when action y is taken, which decides the set of subchannels assigned to each

queue, i.e. m_1 and m_2 , is given by

$$\mathbf{P}^{(a)} = \mathbf{P}_1 \otimes \mathbf{P}_2. \quad (3.36)$$

3.4.4 Queueing Performance for Each Type of Service

With the assumption that m_1, m_2 denote the subset of antennas assigned to class one and class two services, respectively, we can obtain the QoS performance measures for each type of service.

3.4.4.1 Steady State Probability

To evaluate the QoS performance measures, the steady state probabilities for the system states are required. With the assumption that the size of the best-effort queue is finite, the steady state probability of the system $\boldsymbol{\pi}$ can be simply obtained by solving $\boldsymbol{\pi}^{(c)}\mathbf{P}^{(c)} = \boldsymbol{\pi}^{(c)}$ and $\boldsymbol{\pi}^{(c)}\mathbf{1} = 1$, where c is the index of class of users, $\mathbf{1}$ is a column vector of all ones with the same dimension as $\mathbf{P}^{(c)}$.

The steady state probability for the *QoS-sensitive queue* is a vector with size similar to matrix \mathbf{P}_1 . We can expand $\boldsymbol{\pi}^{(1)}$ as follows:

$$\boldsymbol{\pi}^{(1)} = [\pi_0^{(1)}, \pi_2^{(1)}, \dots, \pi_{Q_1}^{(1)}] \quad (3.37)$$

where $\boldsymbol{\pi}_i^{(1)} = [\pi_{i,0}^{(1)}, \pi_{i,1}^{(1)}, \dots, \pi_{i,m_1 \times K}^{(1)}]$. The steady state probability of i packets in the queue can be simply obtained as follows:

$$\boldsymbol{\pi}^{(1)}(i) = \sum_{j=i \times (m_1 \times K)}^{(i+1) \times (m_1 \times K)} \boldsymbol{\pi}^{(1)}(i, j). \quad (3.38)$$

3.4.4.2 Average Queue Length

The average number of packets in queue (i.e., average queue length) for the QoS-sensitive queue ($\bar{x}^{(1)}$) can be calculated as follows:

$$\bar{x}^{(1)} = \sum_{x=0}^{Q_1} x \times \boldsymbol{\pi}^{(1)}(x). \quad (3.39)$$

3.4.4.3 Packet Dropping Probability

The buffer overflow probability (or packet loss probability) can be calculated as the ratio between the average number of dropped packets at queue i due to overflow (denoted as \overline{O}_i) and the average number of incoming packets in one time slot (denoted as $\overline{\zeta}_i$), which can be obtained as follows:

$$\overline{\zeta}_i = \sum_{j=1}^{A_i} f_j(\zeta_j) \times j. \quad (3.40)$$

Then the buffer overflow probability for queue i can be written as

$$P_l^{(i)} = \frac{\overline{O}_i}{\overline{\zeta}_i} \quad (3.41)$$

To calculate the average number of dropped packets due to overflow at queue i , we need to consider the number of packet arrivals as well as the number of successfully transmitted packet(s) in the same time slot. Given that there are j packets in the queue and there are i arriving packets, the number of dropped packets is $\max\{0, i - (Q_1 - j)\}$. It is clear that, buffer overflow only happens when $i + j > Q_1$. To capture the average number of dropped packets, we consider a *fake* matrix \mathbf{P}_1^{fake} , which has the same packet arrival as well as packet departure process as matrix $\mathbf{P}^{(1)}$ but has buffer size of $Q_1 + A_1$. By constructing blocks of submatrices as shown in (??), (??), and (??), where we considered both the packet arrival and the packet departure events, we can obtain a *fake* transition matrix as follows:

$$\mathbf{P}_1^{fake} = \left[\begin{array}{cccc|cccc} \mathbf{q}_{0,0} & \mathbf{q}_{0,1} & \cdots & \mathbf{q}_{0,A_1} & & & & & \\ \mathbf{q}_{1,0} & \mathbf{q}_{1,1} & \mathbf{p}_{1,2} & \cdots & & \mathbf{q}_{1,A_1} & & & \\ \vdots & \vdots & \ddots & \ddots & & \ddots & & & \\ & \ddots & \ddots & \ddots & & \ddots & & & \\ \hline & & \cdots & \cdots & \mathbf{p}_{i,i+A_1-2=Q_1-1} & \mathbf{q}_{i,i+A_1-1=Q_1} & \mathbf{q}_{i,Q_1+1} & 0 & \cdots \\ & & \cdots & \cdots & & \mathbf{q}_{i,i+A_1-2=Q_1} & \mathbf{q}_{i,Q_1+1} & \mathbf{q}_{i,Q_1+2} & \cdots \\ & & & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & & \mathbf{q}_{Q_1,Q_1-1} & \mathbf{q}_{Q_1,Q_1} & \mathbf{q}_{Q_1,Q_1+1} & \cdots & \mathbf{q}_{Q_1,Q_1+A_1} \end{array} \right] \quad (3.42)$$

where \mathbf{q}_{i,Q_1+m} captures the situation when there are m dropped packets. Note that submatrix \mathbf{p}_{i,Q_1+m} captures the transition of channel states for subchannels assigned to this queue.

$$\mathbf{q}_{i,j} = \sum_{\{a,r|\max\{0,i+a-s\}\}} f_a(A_1) T_r^{(m_1)} \quad (3.43)$$

where $0 \leq a \leq A_1$, $0 \leq r \leq |m_c| \times K$, $i \in 0, Q_1$ and $j \in 0, Q_1 + A_1$.

Now, let π' define as follows:

$$\pi' = \pi^{(1)} \mathbf{P}_1^{fake} \quad (3.44)$$

$$\pi' = [\pi'_0, \pi'_1, \dots, \pi'_{Q_1+A_1}] \quad (3.45)$$

Let z_j be the probability that j packets are dropped

$$z_0 = \sum_{k=0}^{Q_1} \pi'_k e \quad (3.46)$$

$$z_1 = \sum_{k=0}^{Q_1} \pi'_{Q_1+j} e \quad \text{for } j = 1, 2, \dots, A_c \quad (3.47)$$

$$(3.48)$$

Then the average number of dropped packets can be obtained as follows:

$$\bar{O}_1 = \sum_{j=0}^{A_c} j z_j \quad (3.49)$$

3.4.4.4 Queue Throughput

The throughput (in packets/time slot) for the QoS queue and the best-effort queue is obtained as $\eta_1 = \bar{\zeta}_1(1 - P_i^{(1)})$.

3.4.4.5 Average Delay for one Packet

Using the effective arrival rate, the average delay for one packet in the QoS sensitive queue (\bar{D}_1) can be obtained by applying Little's law as follows: $\bar{D}_1 = \frac{\bar{x}^{(1)}}{\eta_1}$.

3.4.5 Markov Decision Process Model

The Markov Decision Process (MDP) is a mathematical tool which provides a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of the decision maker. The theory of MDP enables us to find the optimal distributed solution for a discrete time stochastic

control problem. Simultaneously, the state spaces and the actions in our model correspond to a discrete time process, which allows us to formulate a MDP-based model. A MDP model is a basic and generic model which is completely described through its state spaces, actions, transition probabilities and reward functions.

Our objective is to find a solution for antenna assignment for each flow (i.e., how many antenna and which antenna for each flow) subject to the fairness constraint between two flows. For this, the following MDP model can be formulated:

- **State space:** At each time slot, the system occupies a state among a set of finite states which is called the state space. In our model, the state space is $\mathcal{S} = \mathcal{Q} \times \mathcal{AP} \times \mathcal{H}$ which was represented in Section 3.4.3.
- **Action space:** The set of available actions \mathcal{A} is called the action space. In our system model, the action refers to antenna assignment to each flow depending on the number of packets in each queue and the channel quality of each subchannel under the constraint of fairness in instantaneous delays for both flows.

Let \mathcal{A} denote the finite set called the action set. The action in the resource control problem is interpreted as the composite bit-loading allocation for the individual transmit antennas. As has been discussed through (3.4), our model can be decomposed into m independent subchannels, and hence the number of total possible actions can be obtained as follows:

$$a_{(total)} = \sum_{i=0}^m \binom{m}{i}. \quad (3.50)$$

With each action, there should be another transition matrix \mathbf{P} , such as matrix $\mathbf{P}^{(a)}$ in (3.36) corresponding to action a .

Let $a^{(n)} \in \mathcal{A}$ denote the action taken by the decision maker (i.e., the Antenna Assignment Controller) at time slot n . Then, depending on the action taken, the transmitter will know exactly how many and which antennas are allocated to which flows. Note that, the number and index of subchannels allocated to each flow will decide the number of packets leaving the corresponding queue. Therefore, with different antenna assignments we will have different number of packets in the queues in time slot $(n + 1)$. Consequently, we will have different instantaneous delays for the two queues.

Now, let us define function $\Phi(a)$, which represents the number of packets leaving each queue when action a is taken, as follows:

$$\Phi^{(c)}(\mathbf{a}) = \sum_{i=1}^{m_c} \phi_i \quad (3.51)$$

where m is the total subchannel after using SVD in a MIMO system and ϕ_i represents the number of packets retrieved in each subchannel assigned to this queue if action a is taken. Then the total number of packets leaving both the queues, when action a is applied, can be presented as follows:

$$\Phi(\mathbf{a}) = \sum_{i=1}^c \Phi^i. \quad (3.52)$$

- **Transition probability:** When an action is chosen, the next system state is determined by the transition probability. The evolution of the MDP based model here is Markovian with transition probabilities given by

$$Pr(s_i, s_j, a) = Pr[s^{(t+1)} = s_j, s^{(t)} = s_i, a^{(t)} = a] \quad (3.53)$$

where $s_i, s_j \in \mathcal{S}$, $a \in \mathcal{A}$. Note that the process to calculate this transition probability was presented in Sections 3.4.3 and ??.

- **Reward:** The decision maker receives a reward when it chooses an action among the action set. In our model, this reward is related to the instantaneous delay of each queue.

Let us assume that the reward function $\mathcal{R}: \mathcal{S} \times \mathcal{A}$ is given. Then the average cost of a particular policy \mathbf{w} , for a given initial condition z , is defined as

$$J(\mathbf{w}, z) := \lim_{n \rightarrow \infty} \sup \frac{1}{n} \sum_{t=0}^{n-1} \mathbb{E}[\mathcal{R}(s_{\mathbf{w}}^{(t)}, a_{\mathbf{w}}^{(t)}) | Z = z]. \quad (3.54)$$

A policy \mathbf{w}^* is then called *optimal* if $J(\mathbf{w}^*, z) < J(\mathbf{w}, z)$, subject to minimize the instantaneous delays in both queues for all policies \mathbf{w} and any initial state z .

Let the reward $\mathcal{R}(s_{\mathbf{w}}^{(t)}, a_{\mathbf{w}}^{(t)})$ be the instantaneous reward when policy \mathbf{w} is applied. In other words, \mathcal{R} is the reward that the system receives when taking action $a^{(t)}$ in state $s^{(t)}$ at time slot t . $\mathcal{R}(s_{\mathbf{w}}^{(t)}, a_{\mathbf{w}}^{(t)})$ is defined as follows:

$$\mathcal{R}(s_{\mathbf{w}}^{(t)}, a_{\mathbf{w}}^{(t)}) = \mathbb{E}[\mathcal{R}(s_{\mathbf{w}}^{(t)}, s'_{\mathbf{w}}^{(t+1)}, a_{\mathbf{w}}^{(t)})] = Pr(s_{\mathbf{w}}, s'_{\mathbf{w}}, a_{\mathbf{w}}) \times \mathcal{R}(s_{\mathbf{w}}^{(t)}, s'_{\mathbf{w}}^{(t+1)}) \quad (3.55)$$

Now, let the instantaneous reward be defined as

$$\mathcal{R}(s_{\mathbf{w}}^{(t)}, s'_{\mathbf{w}}^{(t+1)}) = (d_1 - \Delta_2 d_2)^2 \times (d_1 - \Delta_3 d_3)^2 \times \dots \times (d_1 - \Delta_c d_c)^2 = \prod_{i=2}^c (d_1 - \Delta_i d_i)^2. \quad (3.56)$$

where d_i is the instantaneous delay of queue i at the end of time slot t and Δ_i is the weight of the instantaneous delay we want to maintain compared with the instantaneous delay for the incoming packets arriving into the *QoS-sensitive queue*. From (3.56), we obtain $\frac{d_1}{\Delta_2} = \frac{d_1}{\Delta_3} = \dots = \frac{d_1}{\Delta_i}$, when $\mathcal{R}(s_{\mathbf{w}}^{(t)}, s'_{\mathbf{w}}^{(t+1)})$ is minimized.

Again, our goal is to solve the average cost optimal control problem (3.54) by constructing a deterministic policy \mathbf{w} with minimal average cost. To construct an optimal policy, value iteration is the most common approach used in practice. The idea is to consider a finite time problem with the following value function

$$V_n(s) = \min \mathbb{E} \left[\sum_{t=0}^{n-1} \mathcal{R}(s(t), a(t)) + V_0(s(n)) \right] \quad (3.57)$$

where $a(t)$ is the sequence of actions $\{a(t) : t \in \mathbb{Z}_+\}$ which is adapted and determined by some policy. Note that, $a(t)$ can only depend on the history of states $\{s(0), \dots, s(t)\}$ and the policy should minimize (3.57). We can form a deterministic policy $\mathbf{w} = w^0(s(0)), w^1(s(1)), w^2(s(2)), \dots$, where for each i function w^i maps the system states to corresponding actions, i.e., $\mathcal{S} \mapsto \mathcal{A}$, with $w^i(s) \in \mathcal{A}(s)$ for each s .

The function V_0 is considered as the penalty term and in the standard value iteration is performed assuming $V_0 \equiv 0$. A deterministic policy is a Markov policy for which $w^i = w$ for all i , for some fixed state feedback law w . This deterministic policy can be obtained based on the *value iteration algorithm* [16]. If the value function V_n is given, the action $w^n(s)$ is defined as

$$w^n(s) = \arg \min_{a_{\mathbf{w}^n} \in \mathcal{A}(s)} \left[Pr(s_{\mathbf{w}^n}, s'_{\mathbf{w}^n}, a_{\mathbf{w}^n}) \theta V_n(s') + \mathcal{R}(s, a) \right] \quad s \in \mathcal{S}. \quad (3.58)$$

For each n the following equation has to be satisfied:

$$\begin{aligned} V_{n+1}(s) &= \mathcal{R}(s, \mathbf{w}^n) + Pr(s, s', \mathbf{w}^n) \theta V_n(s') \\ &= \min_{a \in \mathcal{A}(s)} \left[\mathcal{R}(s, a) + Pr(s, s', a) \theta V_n(s') \right] \\ &= \min_{a \in \mathcal{A}(s)} \sum_{s'} Pr(s, s', a) \left[\mathcal{R}(s, s') + \theta V_n(s') \right] \end{aligned} \quad (3.59)$$

which then makes it possible to calculate the next function \mathbf{w}^{n+1} , where θ ($0 \leq \theta \leq 1$) is the discount rate. The complete procedure of *value iteration algorithm* can be expressed as follows:

Algorithm 3.4.1: VALUE ITERATION ALGORITHM($V_0(s) \equiv 0$)

repeat

$\Delta \leftarrow 0$

for each $s \in \mathcal{S}$

do $\begin{cases} v \leftarrow V(s) \\ V(s) \leftarrow \min_{a \in \mathcal{A}(s)} \sum_{s'} Pr(s, s', a) [\mathcal{R}(s, s') + \theta V_n(s')] \\ \Delta \leftarrow \min(\Delta, |v - V(s)|) \end{cases}$

until $\Delta < \vartheta$

In the above algorithm, ϑ is a small positive value and the output is a deterministic policy $\mathbf{w}^n(s)$ such that (3.58) is satisfied. After the deterministic policy $\mathbf{w}^n(s)$ is obtained, at the beginning of time slot, the AAC checks the states in the previous time slot (i.e., channel states of all subchannel, number of packet in queue at the end of previous time slot) and then makes a decision based on the above policy.

The operation of the AAC for the situation where there are only two kinds of packets in the system can be described as shown in Figure 3.4, where $State_1$ and $State_2$ can be calculated as follows:

$$State_1 = q_2 \times size(P_2) \times K^{m_c} - \left[(K - r_1^{(2)}) \times K^{m_c-1} + (K - r_2^{(2)}) \times K^{m_c-2} + \dots + (K - r_{m_c}^{(2)}) \right] - q_1 \quad (3.60)$$

$$State_2 = q_1 \times size(P_1) \times K^{m_1} - \left[(K - r_1^{(1)}) \times K^{m_1-1} + (K - r_2^{(1)}) \times K^{m_1-2} \dots + (K - r_{m_1}^{(1)}) \right] - q_2 \times size(P_2) \times K^{m_2} - \left[(K - r_1^{(2)}) \times K^{m_2-1} + \dots + (K - r_{m_2}^{(2)}) \right] \quad (3.61)$$

where $r_j^{(i)}$ is rate of channel number j^{th} of queue i .

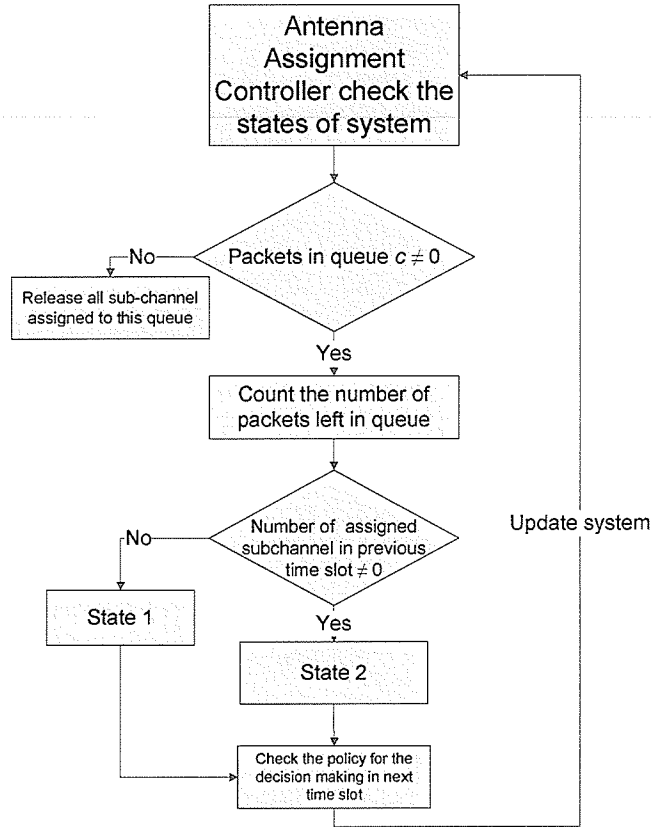


Figure 3.4. Algorithm for the Antenna Assignment Controller.

3.5 Performance Evaluation

3.5.1 Complexity of Policy Evaluation

Before presenting the numerical and simulation results, let us first evaluate the complexity of the MDP-based solution for the antenna selection and assignment problem. Value iteration works by producing successive approximations of the optimal value function. According to [5], the complexity of the value-iteration algorithm with full backups, per iteration, is represented in $O(AS)^2$.

3.5.2 Parameter Setting

In this section, we present typical numerical results considering an uncoded wireless system with three transmission modes, i.e., adaptive modulation with only 3 transmission rates ($K = 3$). First, we obtain the queueing performance of these two queues with the assumption that at a random timeslot, the AAC decides to assign one and two subchannel(s) to *best-effort queue* and *QoS-sensitive queue*, respectively.

To reduce the computational complexity of finding the optimal solution, in the simulation model, the size of both queues is assumed to be 15 packets. We assume that $c_{(n)} = n$, time slot interval is $T_s = 2ms$, and the BER_{target} is set to 10^{-3} , the packet size is 255 bits and block code is used for forward error correction in which errors up to 2 bits can be corrected. We also assume that there are at most two separate subchannels after using SVD to obtain these separate subchannels. For a fading channel, we assume a Nakagami $-m$ channel with parameter $m = 1.1$. An infinite-persistent ARQ is used to ensure reliable packet transmission.

3.5.3 Queueing Performance

In this subsection, we present some typical numerical results of queueing performance. For the purpose of illustration, we present some queueing performance for one particular scenario with an arbitrarily chosen channel assignment. Figure 3.7 and 3.8 present the average delays and packet dropping probability versus packet arrival rate to a QoS-sensitive queue referred to this scenario. Note that this numerical results are only for the queueing performance, not related to MDP solution. We assume that at a particular time slot, there are 3 available subchannels at the transmitter side, and we assign two subchannels to the QoS-sensitive queue and the other to best-effort queue. Among the two subchannels assigned to QoS-sensitive queue, we fix the average SNR to one channel to 10 dB and vary the average SNR of the other subchannel. The subchannel assigned to best-effort queue has a fixed average SNR of $\gamma_2 = 10$ dB.

3.5.4 Simulation Methodology

The simulation results are obtained for both types of services for the proposed ASA policy. Given the system and channel parameters, the channel transition matrix \mathbf{P}_1 ,

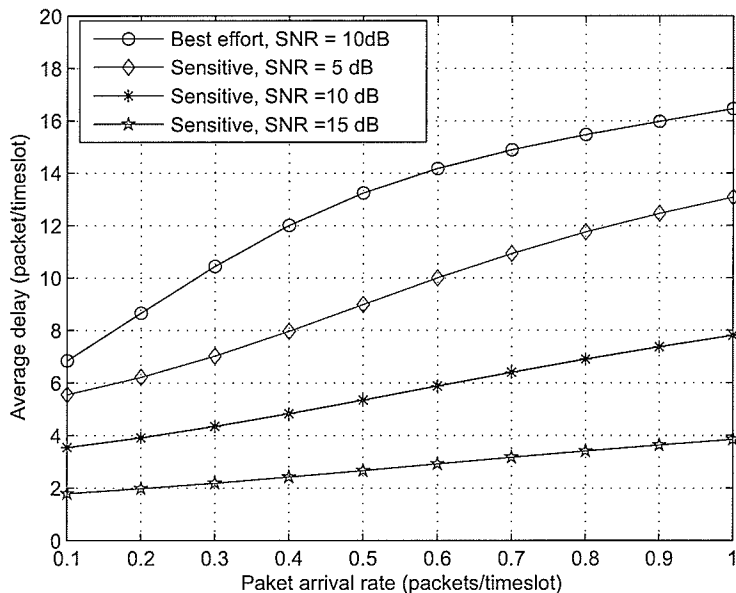


Figure 3.5. Average packet delay under different packet arrival rate with $\zeta_1 = \zeta_2$ and subchannel assigned to best-effort queue has $\gamma_2 = 10dB$ (from analysis).

\mathbf{P}_2 , and \mathbf{P} are calculated. The simulation run time is chosen to be 2×10^6 time slots, in which the channel states of all users are generated based on their channel states in the previous time slot and the corresponding channel state transition probabilities (i.e. $\mathcal{H}^{(i)}$). The number of packets transmitted during any service time slot is determined by the channel state and also PER in this state. The number of packets left in the queue is updated at every time slot by considering packet arrival and the number of successfully transmitted packets. Based on this information, the AAC will calculate the system *State* of system at a previous time slot using the algorithm in Figure 3.4. Then, the AAC applies this *State* into policy \mathbf{w} to find the corresponding action.

The weighted round-robin (WRR) scheduler in our simulation uses both the subchannels to transmit packets from both queues, where the *QoS-sensitive queue* and the *best-effort queue* receive two and one service slots in one cycle, respectively. One cycle is defined to be the smallest interval with time slot assignments and repeats periodically.

The state space in this case contains a total of $16 \times 3 \times 16 \times 3 = 2304$ states, with

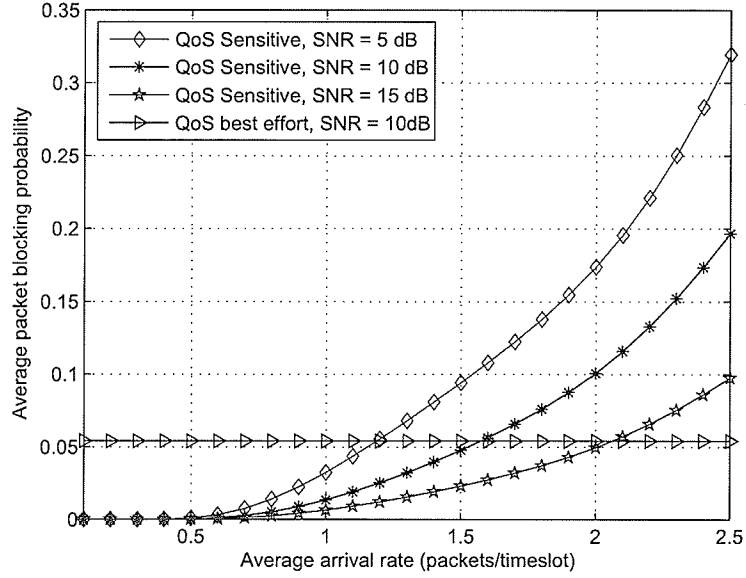


Figure 3.6. Average packet blocking probability under different packet arrival rate with $\zeta_2 = 0.8$ and subchannel assigned to best-effort queue has $\gamma_2 = 10dB$ (from analysis).

the corresponding actions shown in Figure 3.9, where

- *Action number one:* The best channel assigned to the high priority flow, the rest for the other.
- *Action number two:* Opposite to the above situation.
- *Action number three:* Assign both channels to the high priority flow and no channel is assigned to the low priority flow.
- *Action number four:* Opposite to the above situation.

Figure 3.11 shows the packet blocking probability versus the average SNR of one subchannel while the other subchannel has a fixed average SNR 5 dB. Obviously, the packet dropping probability of for the QoS-sensitive queue is higher compared to the scenario where two subchannels are assigned to this queue continuously, and it is lower than the scenario where one subchannel is assigned to this queue. Also, there is a significant improvement in packet dropping probability for best-effort traffic and

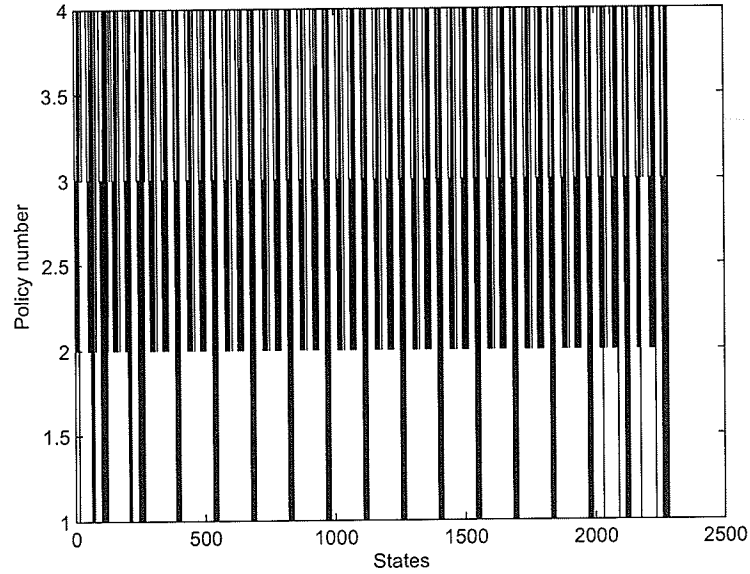


Figure 3.7. *Relation between states and the deterministic policy, where $\zeta_1 = 0.8, \zeta_2 = 0.5, \gamma_1 = 15dB$ and $\gamma_2 = 5dB$.*

it is more likely to guarantee fairness between these two flows compared to the case where a fixed subchannel is assigned to best-effort traffic.

Figure 3.13 shows the average delays of both types of traffic when the arrival rate of packets coming to the best-effort queue is set to be a fixed value, i.e., $\zeta_2 = 0.7$. When the packet arrival rate for the QoS-sensitive queue is smaller than the arrival rate for the best-effort queue (i.e., $\zeta_1 < 0.4$), the average delay of packets in the best-effort queue is smaller than that for the packets in the QoS-sensitive queue. This is because, since the packets coming to the best-effort queue have higher arrival rate than those coming to the QoS-sensitive queue, the ASA scheme decides to allocate more radio resources to the best-effort queue. In other words, during that period, the best-effort traffic has higher priority than the QoS-sensitive traffic.

When the packet arrival rate to the QoS-sensitive queue is higher than the other, ASA starts to share more resource for sensitive queue to satisfy the QoS requirements of both flows. At the diamond shaped points in Figure 3.13, where arrival rate of packets coming to the QoS-sensitive queue equals to the packet arrival rate for the

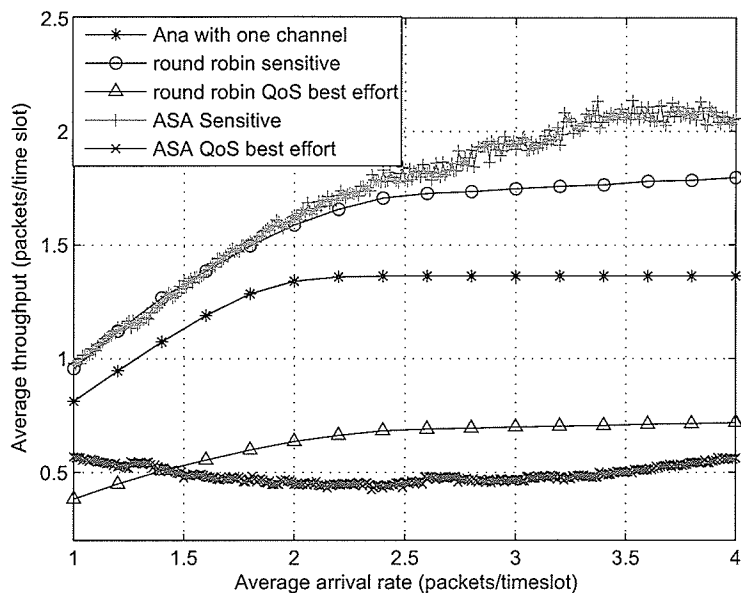


Figure 3.8. *Packet throughput under different packet arrival rate (for $\zeta_2 = 0.8$, and $\gamma_2 = 5$ dB).*

best-effort queue, the average delay is observed to be 5.7374 and 9.0253 corresponding to QoS-sensitive and best-effort traffic, respectively. These obtained results conform to our objective, where we set the instantaneous delay requirement for QoS-sensitive traffic to be 6 timeslots and $\frac{d_1}{d_2} = \frac{2}{3}$. ASA allocates more resources to best-effort traffic since its delay exceeds the target delay requirement, and of course the delay for QoS-sensitive traffic will increase in this case.

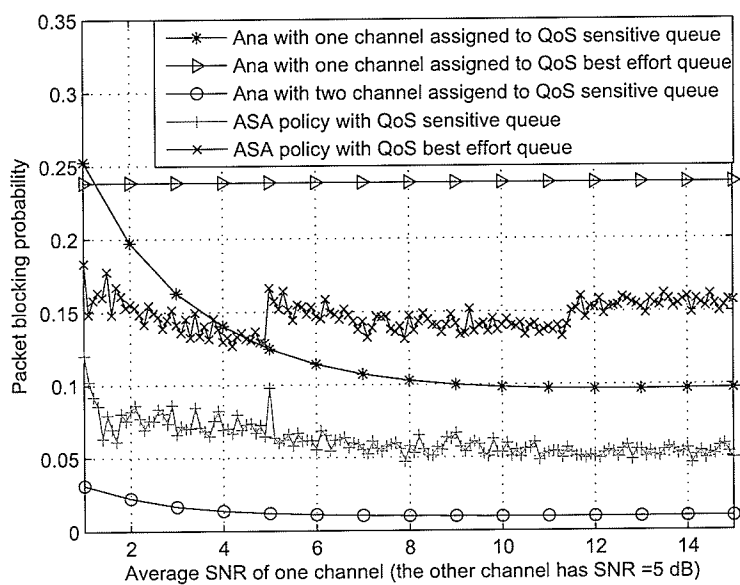


Figure 3.9. Packet blocking probability under different SNR (for $\zeta_1 = \zeta_2 = 0.5$, and $\gamma_2 = 5$ dB).

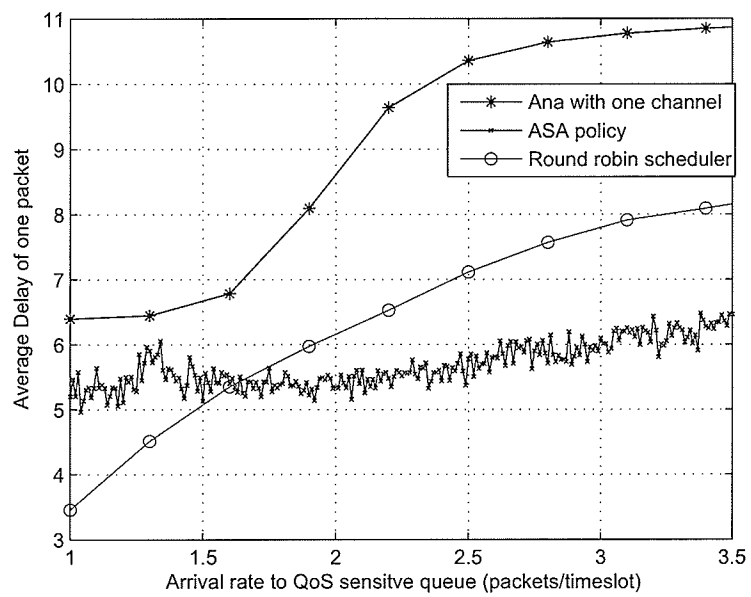


Figure 3.10. *Impact of connection arrival rate on average delay for QoS-sensitive queue (for $\zeta_2 = 0.5$ and $\gamma_2 = 5$ dB).*

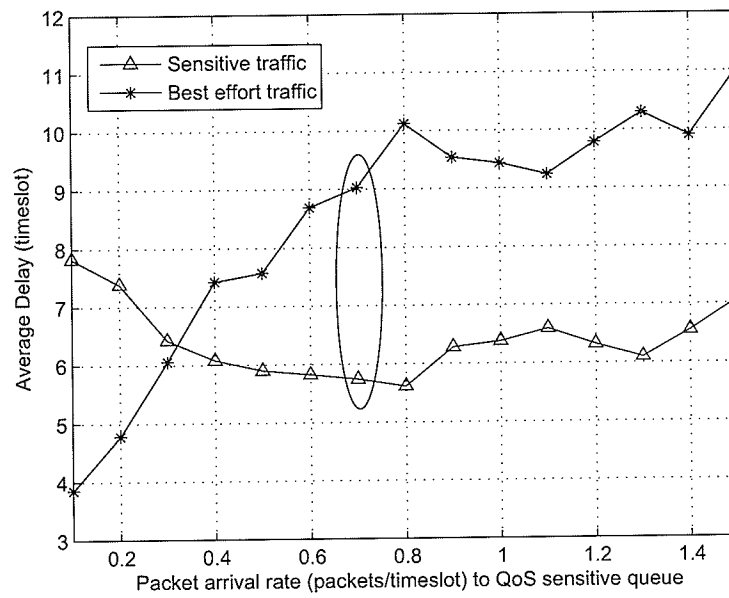


Figure 3.11. Average delays for best-effort and QoS-sensitive traffic (for $\zeta_2 = 0.7$ and $\gamma_1 = 9$ dB $\gamma_2 = 5$ dB).

Chapter 4

A Tandem Queue Model for Performance Analysis in MIMO Wireless Mesh Networks

4.1 Introduction

An important research problem for MIMO-based WMN is to quantify the impacts of different factors on QoS performance measures so that wireless protocols can be designed and/or tuned in an optimal manner. Two of the most important performance measures are end-to-end packet dropping probability and average delay which can only be obtained by solving the tandem system of queues along the routing path of a connection [61]. There are a few tandem queueing models proposed in the literature. Tandem systems of two queues were modeled in discrete time in [47]. The end-to-end delay for TDMA and ALOHA multiple access schemes were approximately derived in [68] for constant bit rate traffic. Developing a general model to analyze tandem queueing systems under realistic arrival process and wireless channel assumptions is still an open research issue.

In this chapter, we present an exact queueing model to solve the tandem queue system for a MIMO-based wireless mesh backbone where batch arrival and multi-rate transmission in the physical layer, and automatic repeat request (ARQ)-based error recovery in the link layer are taken into account. Note that, multi-rate transmission in the physical layer using adaptive modulation and coding (AMC) and ARQ-based error recovery in the link layer are widely used techniques for most of the current wireless

standards,[1]-[3]. For multihop communications in a wireless mesh backbone, it is well known when the size of network increases (counted on the number of hops from source nodes to destination nodes), the network performance degrades significantly. The reason is that the end to end reliability sharply drops as the scale of network increases. By using MIMO links, instead of SISO links, we will show that there is a significant improvement in end to end performance. Because the computational complexity of the exact model is very high, we propose a decomposition approach for the tandem queue system. The proposed decomposition approach allows traffic to arrive at each node besides the relayed traffic. Using the decomposition approach, we can calculate some performance measures, namely, the end-to-end packet dropping probability and average delay. The model is validated and typical numerical results are presented.

The rest of this chapter is organized as follows. Section 4.2 presents the system model. An exact tandem queue model is described in section 4.3, and section 4.4, the decomposition approach is presented in section 4.5. Numerical results are presented in section 4.6 and the last section provides a summary of this chapter.

4.2 System Model and Assumptions

We consider a multihop wireless network, which can be considered as a tandem system, with L concatenated nodes where traffic coming out of each node is fed into the next node in the chain. This kind of tandem system models the operation of wireless mesh backbone networks where data traffic arriving at the source node is transmitted hop by hop to the destination node. The sequence of nodes that the traffic flow traverses is obtained from a routing algorithm.

We assume that each node in this wireless mesh backbone maintains one queue for each link emanating from the node where traffic flows from different connections traversing through the link are buffered for transmission in a FIFO manner. A multihop network model with two connections is shown in Figure 4.1 where for convenience, we show only one queue at each node. In general, data traffic stored in each buffer may come from different connections. For a particular connection, the tandem system of queues along its routing path is illustrated in Figure 4.2. Note that, traffic from

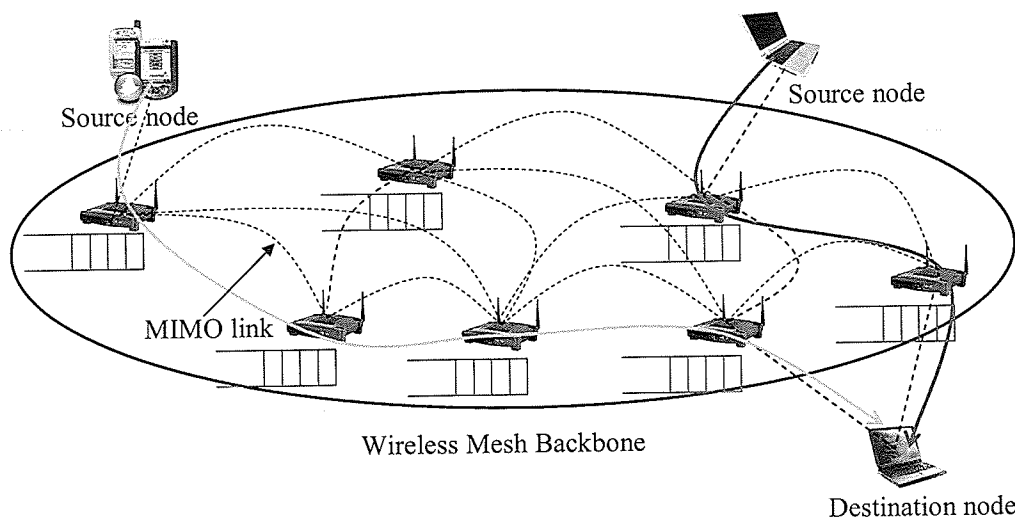


Figure 4.1. A multihop wireless network with multiple ongoing connections.

other connections (called exogenous traffic in this chapter) may arrive at any one of the queues of the tandem system.

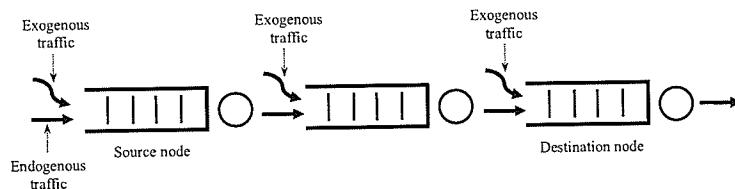


Figure 4.2. A tandem system of queues.

4.2.1 The physical and link layer models

The physical and link layer models are the same as those described in Chapter 3. However, instead of choosing the assigned solution for different flows, in this chapter we use all available subchannels to transmit data. Recall chapter 3, from (3.8), the probability that modulation k is used in subchannel i can be obtained as follows:

$$\begin{aligned}
 Pr^{(i)}(k) &= Pr^{(i)}(\lambda_1 \in [X_k, X_{k+1})) = \int_{X_k}^{X_{k+1}} f_{\lambda_1}(\lambda_1) d\lambda_1 \\
 &= \Psi_1(1, X_k/\gamma_0) - \Psi_1(1, X_{k+1}/\gamma_0).
 \end{aligned} \tag{4.1}$$

For m available subchannels at MIMO link l , let $p(g)$ denote the probability that g packets are correctly received given by

$$p^{(l)}(g) = \bigodot_{i=1}^m Pr^{(i)} \quad (4.2)$$

where \bigodot denotes the discrete convolution and maximum number of packets that can be transmitted in one time slot over m subchannels is $N = m \times K$.

Basically, we can calculate \overline{PER}_k (3.16) over sub-channel i , since now, to reduce the complexity, we assume that \overline{PER} of MIMO link l is $\theta^{(l)}$

Based on the average packet error rate, assuming that the transmission outcomes of consecutive packets are independent, the probability that i packets are correctly received given that j packets were transmitted over link l can be calculated as follows:

$$\beta^{(l)}(i, j) = \binom{j}{i} \left(\theta^{(l)} \right)^{j-i} (1 - \theta^{(l)})^i, \quad (4.3)$$

Note that the difference in departure process between SISO and MIMO here is depended on the number of available antennas at transmitter and receiver sides, where SISO has only one channel while MIMO might have multiple channels. The difference here depends on the number of antennas at the transmitter and receiver side, which is presented by parameter M and N in (3.6) in chapter 3, i.e. SISO has only one antenna at both transmitter and receiver sides or $M = N = 1$. This difference will effect to the service time distribution of SISO and MIMO. Clearly, MIMO technology offers a substantial performance improvement and MIMO system also does not require additional transmit power or receive SNR to deliver such performance gains.

4.3 An Exact Tandem Queue Model

We assume that the packets arrive at the buffer according to independent Poisson processes with different packet arrival rates, i.e. ζ . The maximum number of packets arriving in one transmission queue in one time slot is M , then the probability of arrival of $i \in \{ 0 \dots M \}$ packets with mean ζ in time interval t is given by

$$a_i(\zeta) = \frac{e^{-\zeta t} (\zeta t)^i}{i!}. \quad (4.4)$$

This arrival probability can be expressed as follows:

$$a = \{a_0(\zeta), a_1(\zeta), \dots, a_M(\zeta)\}. \quad (4.5)$$

For simplicity, we define this arrival probability as $a_i = \{a_i\}_{i=0}^M$. We will consider a simple tandem system with two queues as in Figure 4.3 in this section. The more general case with L queues ($L > 2$) is considered in the next subsection. Let $q_i(t)$ be the number of packets in queue i at time slot t , the random process $X(t) = \{q_1(t), q_2(t)\}$, ($0 \leq q_1(t) \leq Q_1, 0 \leq q_2(t) \leq Q_2$) forms a discrete time Markov chain (MC). For notational convenience, we omit the time index t in the related variables when it does not create confusion in the sequel. Let (x, y) be the generic system state (i.e., $q_1 = x, q_2 = y$) and $(x_1, y_1) \rightarrow (x_2, y_2)$ be the system transition from state (x_1, y_1) to state (x_2, y_2) . Now, we derive the transition probabilities for the underlying MC. Specifically, we need to find the transition probability $\Pr\{(x_1, y_1) \rightarrow (x_2, y_2)\}$.

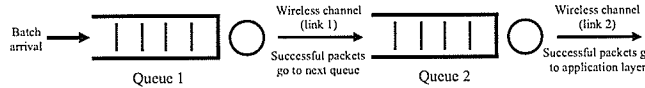


Figure 4.3. A tandem system with two queues.

Note that the number of packets transmitted on each link in any time slot is the minimum of the number of packets in the corresponding queue and the transmission capability of the channel, i.e., $\min\{q_i, i\}$, where i denotes channel capacity of MIMO link allocated to this queue. Let the maximum number of packets that can be transmitted in one time slot be N (i.e., $N = m \times K$). Then, the number of packets in each queue can be reduced at most by N . Because there are at most M packets arriving at queue one (from the source node) and at most N packets enter queue two (due to successful transmissions in all possible channels allocated to queue one) in one time slot, the number of packets can increase at most by M for queue one and by N for queue two.

Hence, if we write the probabilities of transitions $(x_1, *) \rightarrow (x_2, *)$ in a matrix block \mathbf{A}_{x_1, x_2} , the probability transition matrix of the MC $X(t)$ can be written as in (4.7). The order of matrix block \mathbf{A}_{x_1, x_2} is $(Q_2 + 1) \times (Q_2 + 1)$ and its (y_1, y_2) -th element is $\mathbf{A}_{x_1, x_2}(y_1, y_2) = \Pr\{(x_1, y_1) \rightarrow (x_2, y_2)\}$.

Let s denote the number of packets arriving at queue one in a particular time slot and the channel capacity of link 1 k packets, we need to find the conditions under which a general transition $(x_1, y_1) \rightarrow (x_2, y_2)$ occurs. The number of packets in queue one after accepting the newly arrived packets is $\min(x_1 + s, Q_1)$ and the number packets transmitted in link 1 is $\min(x_1, k)$ and among these transmitted packets, w packets are correctly received at the receiving end (i.e., these w successfully transmitted packets will enter queue 2). Thus, we have $x_2 = \min(x_1 + s, Q_1) - w$. Similarly, link 2 can transmit l packets in one time slot, and h packets among $\min(y_1, l)$ transmitted packets are correctly received at the receiving side then $y_2 = \min(y_1 + w, Q_2) - h$. Hence, we can calculate $\mathbf{A}_{x_1, x_2}(y_1, y_2)$ by including all cases where these two conditions hold in the following sum

$$\mathbf{A}_{x_1, x_2}(y_1, y_2) = \sum_s \sum_g \sum_l \sum_w \sum_h a_s p^{(1)}(k) p^{(1)}(l) \times \beta^{(1)}(\min \{x_1, k\}, w) \times \beta^{(2)}(\min \{y_1, l\}, h). \quad (4.6)$$

where all possible cases such that $x_2 = \min(x_1 + s, Q_1) - w$ and $y_2 = \min(y_1 + w, Q_2) - h$ are included in sum.

$$\mathbf{P} = \left[\begin{array}{cccccccc} \mathbf{A}_{0,0} & \mathbf{A}_{0,1} & \cdots & & \mathbf{A}_{0,M} & & & \\ \mathbf{A}_{1,0} & \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \cdots & & & \mathbf{A}_{1,M} & \\ \vdots & \vdots & \ddots & \ddots & & & \ddots & \\ \mathbf{A}_{N,0} & \mathbf{A}_{N,1} & \mathbf{A}_{N,2} & \cdots & \cdots & & & \mathbf{A}_{N,N+M} \\ \vdots & \vdots & \ddots & \ddots & \ddots & & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots \\ & & \mathbf{A}_{Q_1-1, Q_1-1-N} & \mathbf{A}_{Q_1-1, Q_1-N} & \mathbf{A}_{Q_1-1, Q_1-N} & \cdots & \mathbf{A}_{Q_1-1, Q_1} & \\ & & & & & \mathbf{A}_{Q_1, Q_1-N} & \cdots & \mathbf{A}_{Q_1, Q_1-1} & \mathbf{A}_{Q_1, Q_1} \end{array} \right] \quad (4.7)$$

Now, we are ready to derive the steady state probabilities for MC $X(t)$. Let π be the steady state probability vector of the MC $X(t)$. We have

$$\pi \mathbf{P} = \pi, \quad \pi \mathbf{1} = 1 \quad (4.8)$$

where $\mathbf{1}$ is a column vector of all ones with the same dimension as π , which is $(Q_1 + 1)(Q_2 + 1)$. We can expand π as follows:

$$\pi = [\pi_0, \pi_1, \pi_2, \cdots, \pi_{Q_1}]$$

where π_i is a row vector of dimension $Q_2 + 1$, which can be further expanded as $\pi_i = [\pi_{i,0}, \pi_{i,1}, \pi_{i,2}, \dots, \pi_{i,Q_2}]$, where $\pi_{i,j}$ is the probability that the queueing system is in state (i, j) . Given the steady state probability vector π which is calculated using (4.8), we can derive the following important end-to-end QoS measures.

4.3.1 End to End Packet Dropping Probability

The packets can be lost due to transmission errors on wireless links or due to overflow. The buffer overflow probability for queue k of the tandem can be calculated as a ratio between the average number of dropped packets due to overflow at queue k (denoted as \bar{O}_k) and the average number of packets arriving at queue k in one time slot (denoted as \bar{A}_k). Hence, the buffer overflow probability for queue k can be written as $P_l^{(k)} = \frac{\bar{O}_k}{\bar{A}_k}$.

Note that the average number of packets arriving at queue one in one time slot is $\bar{A}_1 = \sum_{i=1}^M i a_i$. To calculate the average number of dropped packets due to overflow at queue one, let us define z_i be the marginal probability that there are i packets in queue one. We have $z_i = \pi_i \mathbf{1}_{Q_2+1}$, where $\mathbf{1}_{Q_2+1}$ is a column vector of all ones with dimension $Q_2 + 1$. The average number of dropped packets due to overflow at queue one can be calculated as

$$\bar{O}_1 = \sum_{i=1}^M \sum_{j=Q_1-M}^{Q_1} a_i z_j \max\{0, i + j - Q_1\}$$

where $\max\{0, i + j - Q_1\}$ is the number of dropped packets (if any) given there are j packets in queue one and i arriving packets.

Now we calculate the buffer overflow probability at queue two. We first determine the arrival probability for packets entering queue two due to successful transmissions from queue one. In fact, the number of packets arriving at queue two are those successfully transmitted over link one. Note that, in this case we do not consider the exogenous traffic, then the probability that i packets arrive at queue two can be approximated as

$$b_i \approx \sum_{k=0}^{Q_1} \sum_{l=0}^{K \times m} z_k p^{(1)}(l) \beta^{(1)}(\min\{k, l\}, i).$$

The average arrival rate to queue two can be calculated as $\bar{A}_2 = \sum_{i=1}^N i b_i$. To calculate the average number of dropped packets due to overflow at queue two, let us

define s_i to be the marginal probability that there are i packets in queue two, which can be calculated as $s_i = \sum_{j=0}^{Q_1} \pi_{j,i}$. Similar to queue one, the average number of dropped packets due to overflow at queue two can be approximated as

$$\bar{O}_2 \approx \sum_{i=1}^N \sum_{j=Q_2-N}^{Q_1} b_i s_j \max\{0, i + j - Q_2\}$$

Finally, the end-to-end packet dropping probability can be approximated as

$$P_l \approx 1 - (1 - P_l^{(1)})(1 - P_l^{(2)}) \quad (4.9)$$

where the loss due to overflow at both buffers and due to channel errors are taken into account.

4.3.2 Average Delay

The end-to-end delay is the sum of delays that any packet experiences in all queues and links along its routing path. We ignore the transmission delays and only include queueing delay in the calculation. Using Little's law, the end-to-end average delay can be written as

$$D = \frac{\sum_{i=1}^{Q_1} i z_i}{A_1(1 - P_l^{(1)})} + \frac{\sum_{i=1}^{Q_2} i s_i}{A_2(1 - P_l^{(2)})} \quad (4.10)$$

where the numerator of each term is the average queue length of each queue and the denominator is the average arrival rate considering packet loss due to overflow.

4.4 General Case ($L > 2$)

We consider a more general tandem system with more than two queues in this section. Only traffic flow from the considered connection is taken into account as in the previous subsection. The more general case as shown in Figure 4.2 will be investigated in Section 4.5.2. Now, the tandem system has L queues ($L > 2$) which are concatenated to each other as a chain. The buffer size of queue i is assumed to be Q_i packets. Similar to the previous subsection, let $q_i(t)$ be the number of packets in queue i at time slot t ($i = 1, 2, \dots, L$), the random process $Y(t) = \{q_1(t), q_2(t), \dots, q_L(t)\}$, ($0 \leq q_1(t) \leq Q_1, 0 \leq q_2(t) \leq Q_2, \dots, 0 \leq q_L(t) \leq Q_L$) forms a discrete-time Markov chain (MC).

Obviously, a similar approach as in Section 4.3 can be pursued to obtain the transition probabilities for this MC. The number of state transitions for this MC, however, grows exponentially with the number of queues in the tandem system. In fact, the order of the transition probability matrix \mathbf{P} is $\prod_{i=1}^L (Q_i + 1) \times \prod_{i=1}^L (Q_i + 1)$. Therefore, the computational complexity is very high for a large number of queues and a large buffer size. However, theoretically we can follow a similar procedure as in Section 4.3 to derive the transition probabilities, obtain the steady state probability vector and all end-to-end performance measures.

4.5 Solution of the Tandem Queueing Model: Decomposition Approach

We present a decomposition approach to solve the general tandem queue where the computational complexity grows only linearly with the number of queues in the system. For ease of reference, buffers (queues) along the routing path are numbered in an increasing sequence of integers with the buffer of the source node denoted as buffer (queue) one.

To the best of our knowledge, a Jackson network is a queueing network with L queues satisfying 3 properties:

- The servers in each of the L queues are independent of each other.
- The external arrivals (if any) at each queue are independent processes with different rates.
- The network uses random routing.

The idea of decomposition approach here is basically similar to the idea of solution for Jackson queueing networks where the service times are not exponential, buffer sizes are finite, and any arriving packet which sees the full buffer will be dropped instead of waiting for space in the next queue when the buffer at next queue is full.

4.5.1 Technical Approach

We first consider the simple case where only queue one in the tandem accepts traffic from the higher layer, other queues receive packets which are successfully transmitted

from the previous queue in the chain. The more general case with exogenous traffic from other connections as shown in Figure 4.2 will be discussed in Section 4.5.2. For notational convenience, we assume that i packets arrive at queue k with probability $a_i^{(k)}$ ($a_i^{(2)} = b_i$ for the two queue case considered in Section 4.3.1). Note that the maximum batch size (maximum number of packets arriving in one time slot) captured in $a_i^{(k)}$ is N (i.e., number of packets leaving this queue) while the maximum batch size captured in $a_i^{(1)}$ is M (i.e., $a_M^{(1)}$ is the probability of M packets arriving queue).

We observe that the behavior of queue $i + 1$ does not impact queue i in the chain but the reverse is not true. This is because the outcomes (i.e., successfully transmitted packets) from queue i are fed into queue $i + 1$ and packets leaving queue i which cannot be accommodated at queue $i + 1$ are simply dropped, and packets do not wait at queue i for space in queue $i + 1$ when the buffer at $i + 1$ is full. Thus, instead of forming the Markov chain which captures the queue length dynamics of all queues we could find the queue length dynamics for one queue at a time where its input is the output of the previous queue in the chain (except for queue one).

Specifically, we pursue the following steps. Form the MC for queue one, and calculate its steady state probability vector. Based on the steady state probabilities, calculate the packet arrival probabilities to the next queue. These arrival probabilities are used to derive the arrival probabilities for the next queue in the chain. This procedure is repeated until we obtain the solutions for the last queue of the tandem system.

Obviously, using this decomposition approach the joint steady state probability vector could not be found as in Section 4.3. However, the steady state probability vector for each queue in the chain is what we need to calculate the desired queueing performance measures. Essentially, the presented procedure requires to solve L separate queues each of which accepts batch arrival traffic and serves packets also in batches. Let us consider a particular queue k of the chain and form the MC $X_k(t) = \{q_k(t)\}$, ($0 \leq q_k(t) \leq Q_k$) where $q_k(t)$ denotes the number of packets in queue k with an arrival process described by $a_i^{(k)}$.

The transition probabilities for this MC can be found as follows. Let us consider a general transition probability $\Pr \{x_1 \rightarrow x_2\}$. Let s be the number of packets arriving at queue k and let the number of packets leaving this queue during the considered

time slot be l . Let us also assume that i packets among $\min\{x_1, c_l\}$ packets are correctly received, then a number of packets at queue k in the next time slot is $x_2 = \min\{x_1 + s, Q_k\} - i$. Thus, the transition probability $\Pr\{x_1 \rightarrow x_2\}$ can be approximated as

$$\Pr\{x_1 \rightarrow x_2\} \approx \sum_l \sum_s \sum_i a_s^{(k)} p^{(k)}(l) \times \beta^{(k)}(\min\{x_1, l\}, i) \quad (4.11)$$

where all combinations of l , s and i such that $x_2 = \min\{x_1 + s, Q_k\} - i$ are included in the sum.

Given the transition probabilities, we can easily calculate the steady state probability vector of this MC $\pi^{(k)} = [\pi_0^{(k)}, \pi_1^{(k)}, \dots, \pi_{Q_k}^{(k)}]$, where $\pi_i^{(k)}$ denotes the probability that there are i packets in queue k . As in Section 4.3, the buffer overflow probability at queue k can be calculated as $P_l^{(k)} = \frac{\bar{O}_k}{\bar{A}_k}$. The average arrival rate of traffic to queue k can be written as $\bar{A}_k = \sum_{i=1}^{B^{(k)}} i a_i^{(k)}$, where $B^{(k)}$ is the maximum batch size of the arrival process.

The probability that i packets are successfully transmitted at queue k and arrive at queue $k+1$ can be written as

$$a_i^{(k+1)} = \sum_{j=0}^{Q_k} \sum_{l=0}^N \pi_j^{(k)} p^{(k)}(l) \beta^{(k)}(l) (\min\{j, l\}, i). \quad (4.12)$$

These arrival probabilities are used to derive the queueing solution for queue $k+1$ as in the presented procedure. And the average number of dropped packets due to overflow at queue k can be calculated as

$$\bar{O}_k = \sum_{i=1}^{B^{(k)}} \sum_{j=Q_k-B^{(k)}}^{Q_k} a_i^{(k)} \pi_j^{(k)} \cdot \max\{0, i + j - Q_k\}. \quad (4.13)$$

Finally, the end-to-end loss probability can be approximated as

$$P_l \approx 1 - \prod_{i=1}^L (1 - P_l^{(i)}) \quad (4.14)$$

and the end-to-end average delay can be written as

$$D = \sum_{k=1}^L \frac{\sum_{i=1}^{Q_k} i \pi_i^{(k)}}{\bar{A}_k (1 - P_l^{(k)})}. \quad (4.15)$$

4.5.2 Tandem Queue with Exogenous Traffic

In this subsection, we generalize the queueing model further by allowing buffers of the tandem system to accommodate exogenous traffic of other connections traversing the links of the considered routing path. At each queue of routing path, traffic coming from other connections also enqueues and shares the capacity of the wireless link as shown in Figure 4.2. Traffic from the considered connection will be called endogenous traffic and that from other connections will be called exogenous traffic in the sequel.

Beside the endogenous traffic as has been considered in subsection 4.5.1, we assume that i packets from exogenous traffic sources arrive at queue k with probability $e_i^{(k)}$ ($i = 0, 1, \dots, M, k = 1, \dots, L$). Hence, both endogenous and exogenous traffic arrive at each queue of the chain. The same procedure presented in subsection 4.5.1 can be applied in this case where the aggregate traffic is multiplexed by endogenous and exogenous traffic sources. Let $f_i^{(k)}$ denote the probability that i packets of the aggregate traffic arrive at queue k in one time slot. Then, we put $f_i^{(k)}$ into a row vector as $f^{(k)} = [f_0^{(k)}, f_1^{(k)}, \dots, f_{B^{(k)}}^{(k)}]$, where $B^{(k)}$ is the maximum batch size. Similarly, putting arrival probability for endogenous traffic $a_i^{(k)}$ and exogenous traffic $e_i^{(k)}$ into vectors $a^{(k)}$ and $e^{(k)}$, respectively, we have

$$f^{(k)} = a^{(k)} \odot e^{(k)} \quad (4.16)$$

where \odot denotes the convolution operation. To calculate the queueing performance measures, $f_i^{(k)}$ is considered as arrival probabilities at each queue of the chain (instead of $a_i^{(k)}$ as in Section 4.5.1). If exogenous traffic corresponding to some connections leaves a queue in the middle of the chain, we have to exclude that traffic in calculating the queueing solution for the next queue. We can approximately exclude that traffic by turning off the traffic from that connection entering the tandem system at some previous queue.

4.6 Validation of Decomposition Approach and Typical Numerical Results

We validate the proposed decomposition approach for the tandem queue model and present some illustrative numerical results. We consider wireless networks employing

M-QAM adaptive modulation without coding using six transmission modes for all transmission links, i.e. $k = 0, 1, 2, \dots, 5$. We assume that in channel state k , k packets are transmitted in one time slot, the time slot interval $T_s = 2$ ms.

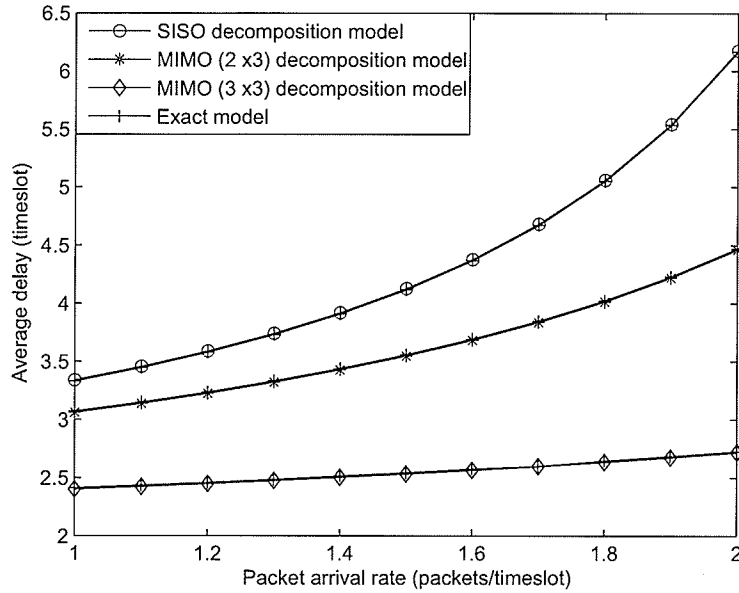


Figure 4.4. Impact of packet arrival rate on average delay (for SISO: $\gamma_0 = 15$ dB, for 2×3 MIMO: $\gamma_0 = 5$ dB, for 3×3 MIMO: $\gamma_0 = 3$ dB, $Q_1 = Q_2 = 50$, $L = 2$).

The variations of end-to-end packet dropping probabilities and average delay with endogenous arrival rate where there is no exogenous traffic are shown in Figure 4.4 and Figure 4.5 for average PER = 0.001. We show results obtained from both the exact and proposed decomposition queueing models in these figures. As is evident, the proposed decomposition model provides a very accurate queueing solution but it has much lower computational complexity compared with the exact model. Figure 3.12 and 4.5 also show the impacts of PHY layer design on the queueing performance. Specifically, with more conservative PHY design, i.e. more antennas at transmitter and receiver sides, there is a significant improvement in both the average delay and the end-to-end packet dropping probability.

The average delay versus packet arrival rate with different exogenous traffic in the tandem with $L = 2$ is shown in Figure 4.6. Here, besides endogenous traffic, we allow

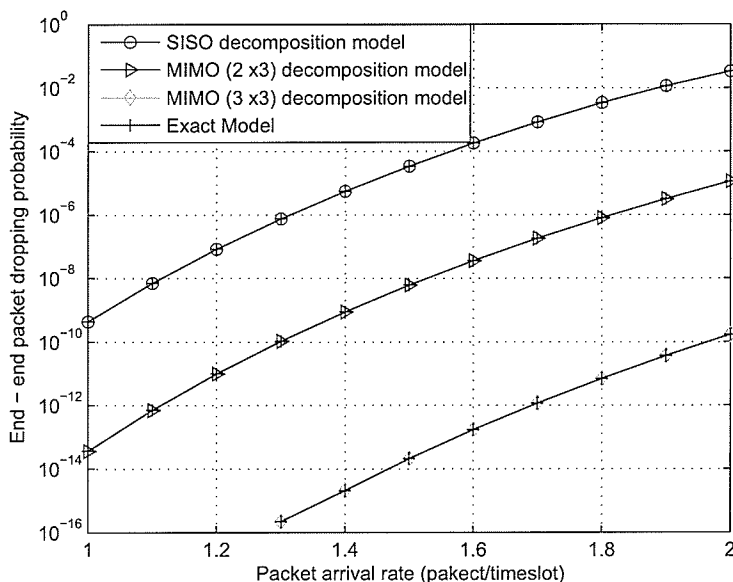


Figure 4.5. Impact of packet arrival rate on end - end packet dropping probability (for SISO: $\gamma_0 = 15$ dB, for 2×3 MIMO: $\gamma_0 = 5$ dB, for 3×3 MIMO: $\gamma_0 = 3$ dB, $Q_1 = Q_2 = 50$, $L = 2$).

exogenous traffic from one other connection to arrive at queue two of the tandem and share wireless capacity with endogenous traffic in the onward direction. The average delays are shown for different arrival rates of exogenous traffic. Moreover, when multiple antennas are equipped in each node, the end-to-end packet dropping probability decreases rapidly compared to the SISO scenario. For example, in 3×3 MIMO scenario packet dropping probability of packet arrival rate from 0.1 to 1.3 ≈ 0 .

Figures 4.7 and 4.8 show the end-to-end packet dropping probability and average delay versus packet arrival rate and exogenous traffic entering each queue has packet arrival rate $\zeta = 0.8$. We observe that both end-to-end packet dropping probability and average delay increase rapidly with the number of hops (H) and packet arrival rate. Therefore, the long routing paths may increase the chances of QoS requirement being violated significantly.

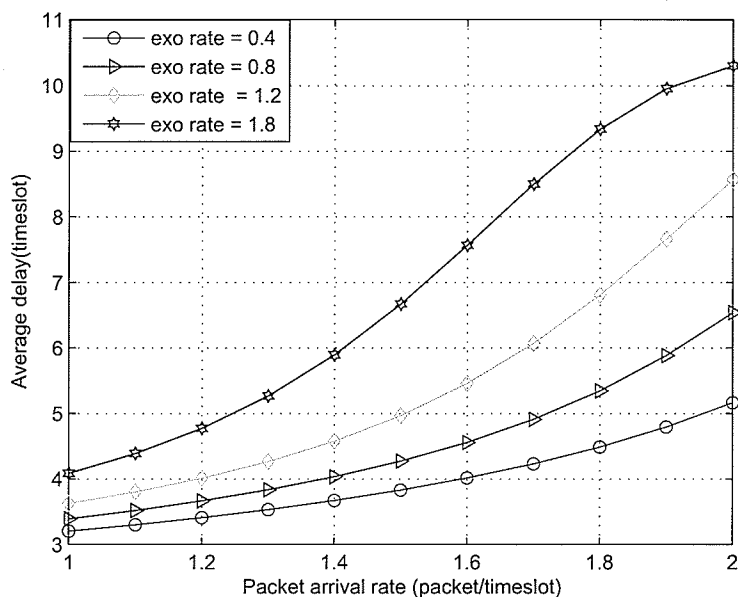


Figure 4.6. Impact of packet arrival rate on average delay in 2×2 MIMO with $\gamma_0 = 5$ dB, $Q_1 = Q_2 = 50$, $L = 2$, with different exogenous traffic.

4.7 Summary

We have proposed a tandem queueing model for wireless mesh backbone networks. We have presented both exact and approximated decomposition approaches to solve a general tandem queue system. The proposed decomposition approach achieves very accurate queueing performance measures with much lower computational complexity compared to the exact approach. We also observe that by using MIMO links, instead of SISO links, there is a significant improvement in end to end performances. The proposed queueing model can be used to perform QoS routing and admission control in a multihop wireless mesh backbone network.

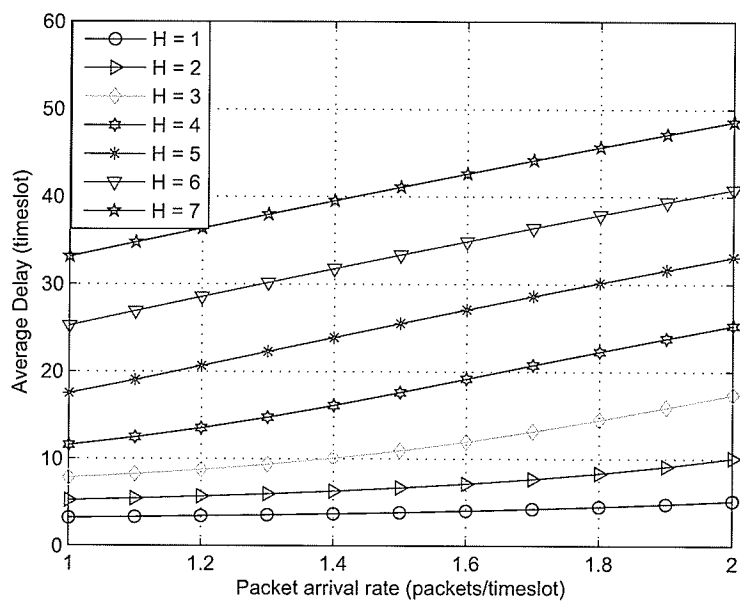


Figure 4.7. Impact of packet arrival rate on average delay in 2×2 MIMO with $\gamma_0 = 5$ dB, $Q_k = 50$, with different number of hops, i.e., $H = L - 1$.

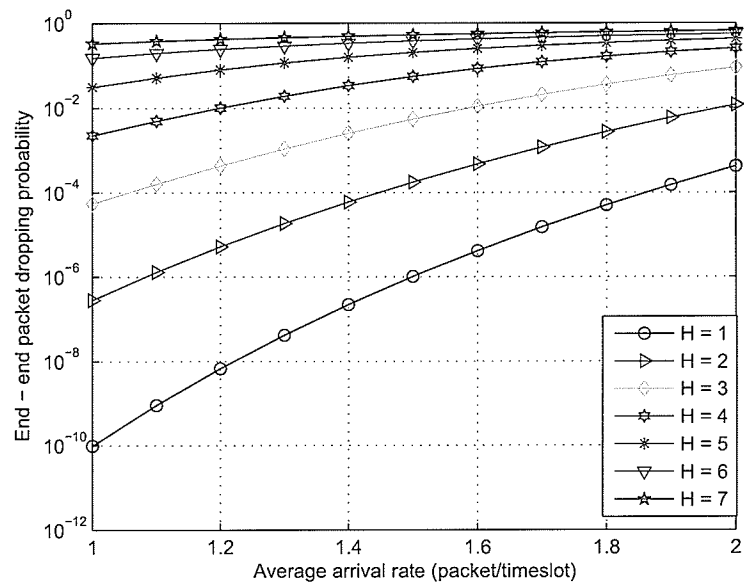


Figure 4.8. *Impact of end - end packet dropping probability on average delay in 2×2 MIMO with $\gamma_0 = 5$ dB, $Q_k = 50$, with different number of hops, i.e., $H = L - 1$.*

Chapter 5

Conclusion

This chapter provides a summary of the works presented in this thesis and outlines a few directions for future research.

5.1 Summary

We have presented an introduction and background on MIMO and its application on wireless mesh networks. The advantage of using MIMO in a wireless mesh backbone instead of a single antenna has been emphasized. We also presented a survey on the issues and approaches to resource allocation in MIMO based wireless networks. These issues include MAC protocol design, admission control, power and rate control, cross layer design in MIMO wireless networks. Two analytical models, i.e. tandem queue system and its application in wireless mesh backbone based MIMO and an Antenna selection and assignment approach, have been developed for wireless mesh networks considering realistic physical and link layer designs. In particular, link adaptation techniques in the physical layer and ARQ-based error recovery in the link layer have been taken into account in different analytical models. Both single-hop radio link level design problems and end-to-end research issues have been considered.

The following provides a summary of the works presented in this thesis:

- *Antenna Selection and Allocation in MIMO based wireless mesh networks:* We have presented a scheme for antenna selection and allocation to achieve service differentiation between QoS-sensitive and best-effort traffic in MIMO based wireless mesh networks. This scheme, which is based a Markov Decision Process formulation of the antenna selection problem, prioritizes two different traffic

types and satisfies the requirement of QoS fairness between these two traffic types. Both channel state information and channel memory are exploited for wireless transmission. From the analytical model, various QoS measures can be obtained. The performance evaluation results have shown that the solution based MDP can provide nearly the target level of QoS to both the QoS-sensitive and the best-effort traffic.

- *Performance Analysis of a MIMO-Based Wireless Mesh Backbone Network:* For performance analysis of a multihop wireless mesh backbone network, we have proposed a tandem queue model which takes the QoS requirements such as the end-to-end packet dropping probability and end-to-end average delay into account. Both exact and approximate tandem queueing models considering link adaptation and ARQ error protection have been proposed. By using spatial multiplexing MIMO in the physical layer, a number of parallel independent streams can be sent at the same time, thus increasing the data rate. The results from the analysis reveal that using MIMO links, instead of SISO links, is a key technique to improving the end to end performance.

5.2 Future Research Works

The thesis has developed some systematic methods for cross-layer analysis and resource allocations in MIMO-based wireless mesh networks. The results are also applicable to resource allocation problems in MIMO-based wireless ad hoc or sensor networks. The long-term goal of this research is to derive structural results and adaptive algorithms for MIMO-based wireless ad hoc/mesh/sensor networks. For future research, the work in this thesis can be extended in several directions. In what follows, we propose a few research problems related to the work in this thesis.

- *Reduction in computational complexity:* As it has been mentioned in Section 3.5.1, although, in theory, MDPs can be solved in polynomial time in the size of state space and action space, this only holds true for so called *flat* representations of the system in which the states are explicitly enumerated. Due to complexity in computations required to investigate policy, the system, in fact, takes a considerable amount of memory and time to implement the optimal policy. Thus

a stochastic approximation algorithm which reduces the computational complexity is necessary for a practical system where the buffer size and transmission modes are in a large number.

There are some works in the literature which focused on *how to reduce the computational complexity*. In [48], an optimal approach to find a transmission scheduling policy that minimizes an expected total cost, which is the sum of accumulated transmission costs and a data loss cost, has been proposed. The concept of *supermodularity* was applied to the dynamic programming equations to prove that the optimal transmission policy is a *threshold policy* in the residual transmission time and the buffer state. Thus if a MDP has a *threshold policy*, one only needs to compute the *threshold* to implement the optimal policy. However, the action space in this work is quite simple, i.e., there are only two actions where action 1 corresponds to ‘transmit’ and action 2 corresponds to ‘no transmit’. Therefore, this approach cannot be directly applied to our model.

The concept of *supermodularity* [21] was used to prove the monotone structure of the optimal resource allocation policy. However, the proofs and the algorithm to obtain the monotone results in [21] cannot be generalized for the case of a Markovian system as considered in this thesis.

Another possible approach to reduce the computational complexity of natural MDP might be the *policy bound* for MDP. In [44], the authors proved that an MDP can be approximated to generate a *policy bound*, i.e., a function that bounds the optimal policy from below or from above for all states. This work also presents sufficient conditions for several computationally attractive approximations to generate rigorous policy bounds. These approximations include approximating the optimal value function, replacing the original MDP with a separable approximate MDP, and approximating a stochastic MDP with its deterministic counterpart. An example in fishery management problem [44] showed that the number of calculations per iteration reduces from 10^{60} states to 10^{20} states.

In future works, we might count on Q-learning and gradient based stochastic approximations to solve the problem. Q-learning is a *model-free* method that is

extremely useful for solving infinite horizon MDPs, where the transition probabilities that characterize how the system evolves may not be available. The principle of Q-learning is to gather information about the transition probabilities, as well as costs, through sampled trajectories of the system [65]. Thus, the agent is required to learn by interacting with the environment. Using the learning algorithm, the expected cost of performing different actions at different system states can be estimated, and the agent can choose the best action accordingly.

- *Opportunistic resource allocation with partially observable states:* In this thesis, we have made the assumption that perfect quantized CSI is available to both transmitter and receiver. In a practical system, channel states can only be observed in noise, and perfect quantized CSI is not easy to obtain. A prospective extension of the work here is to consider the case when the channel state cannot be observed perfectly. In such a situation, the opportunistic resource allocation problem can be formulated as a partially observable MDP (POMDP). The motivation for the POMDP formulation is the obvious fact that POMDPs are a closer approximation to the real system. However, the drawback is that POMDPs are much harder, and typically infeasible, to solve optimally. As a result, analyzing a POMDP will require some sorts of approximation eventually.
- *Resource allocation in MIMO wireless mesh networks in multiple-hop scenarios:* In Chapter 3, we have considered the resource allocation problem from the perspective of a single mesh router. In Chapter 4, we have presented a queueing theoretic performance analysis model for a multihop transmission scenario. Constructing an optimal resource allocation strategy involving multiple mesh routers which optimizes the end-to-end system performance is the key to successful implementation of wireless mesh networks. Optimization and game theory might be important tools to investigate the solution. Also, due to the distributed nature of wireless mesh networks, both in wireless clients and routers, routing and congestion control together with resource allocation and physical layer design issues continue to be challenging research issues.

Bibliography

- [1] *cdma2000 High rate packet Data Air Interface Specification*, 3GPP2, C.S0024, Version 4.0, Oct. 2002.
- [2] IST-MEMBRANE: *Multi-Element Multihop Backhaul Reconfigurable Antenna Network* (www.ist-membrane.org).
- [3] *Physical Layer Aspects of UTRA High Speed Downlink Packet Access (Release 4)*, 3GPP, 3G TR25.848 V4.0.0, Mar. 2001.
- [4] <http://users.ece.utexas.edu/~rheath/research/mimo/adhoc/>
- [5] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, 237 - 285, 1996.
- [6] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communication Magazine*, vol. 43, no. 9, pp. 23-30, Sept. 2005.
- [7] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, pp. 1451-1458, vol. 16, no. 8, Oct. 1998.
- [8] G. Aniba and S. Aissa, "Adaptive scheduling for MIMO wireless networks: Cross-layer approach and application to HSDPA," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 259-268, 2007.
- [9] L. Badia, M. Zorzi, and A. Gazzini, "A model for threshold comparison call admission control in third generation cellular systems," *Proceedings of IEEE International Conference on Communications 2003*, vol. 3, pp. 1664-1668, May 2003.
- [10] H. Balakrishnan, V. Padmanabhan, S. Seshan, and R. H. Katz, "A comparison of mechanisms for improving TCP performance over wireless links," *IEEE/ACM Transactions on Networking*, Dec. 1997.
- [11] R. Bhatia and L. Li, "Throughput optimization of wireless mesh networks with MIMO links," *Proceedings of Infocom 2007*, pp. 2326-2330, May 2007.
- [12] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," *IETF, RFC 2475*, Dec. 1998.
- [13] P. Casari, M. Levorato, and M. Zorzi, "DSMA: An access method for MIMO ad hoc networks based on distributed scheduling" *Proceedings of IWCMC*, July 2006.

- [14] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath Jr., "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Communication Magazine*, vol. 40, no. 6, pp. 108-115, June 2002.
- [15] A. Chan, K. Tsang, and S. Gupta, "TCP (Transmission Control Protocol) over wireless links," in *Proceeding of IEEE Vehicular Technology Conference*, May 1997.
- [16] R.-R. Chen and S. Meyn, "Value iteration and optimization of multiclass queueing networks," in *Proceedings of IEEE Conference on Decision and Control*, 1998.
- [17] J. Chen and K. Wong, "Multiuser MIMO-TDMA with statistical feedback," *Proceedings of International Conference on Information, Communications Signal Processing*, 10-13 Dec. 2007.
- [18] Y. Choi and S. Bahk, "Downlink scheduling with fairness and optimal antenna assignment for MIMO cellular systems," in *Proceeding of IEEE Globecom*, vol. 5, pp. 3165-3169, Nov. 2004.
- [19] R. R. Choudhury, X. Yang, R. Ramanathan, and N. H. Vaidya, "Using directional antennas for medium access control in ad hoc networks," *Proceedings of ACM MOBICOM*, Sept. 2002.
- [20] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO in sensor networks," *IEEE Journal on Selected Areas of Communications*, vol. 22, pp. 1089-1098, Aug. 2004.
- [21] C. Derman, G. Lieberman, and S. Ross, "Optimal system allocations with penaltycost," *Management science*, vol. 23, no. 4, pp. 399-403, Dec. 1976.
- [22] D. V. Djonin, and V. K. Bhargava, "Optimal and suboptimal packet scheduling over time-varying flat fading channels" *Proceeding of IEEE International Conference on Communications*, France, Jun. 2004.
- [23] D. V. Djonin, and V. K. Bhargava, "Q-Learning algorithms for constrained decision processes with randomized monotone policies: Application to MIMO transmission control," *IEEE Transactions on Signal Processing*, vol. 55, issue 5, no. 2, pp. 2170-2181, May 2007.
- [24] M. Dohler and H. Aghvami, "On the approximation of MIMO capacity," *IEEE Transaction on Wireless Communication*, vol. 4, no. 1, pp. 30-34, Jan. 2005.
- [25] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," *ACM Mobicom*, Sept. 2004.
- [26] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, pp. 41-59, 1996.
- [27] G. J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, and R. A. Valenzuela,

- “Analysis and performance of some basic space-time architectures,” *IEEE Journal Selected Areas on Communnication*, vol. 21, pp. 303-320, 2003.
- [28] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, “Simplified processing for high spectral efficiency wireless communication employing multi-element arrays,” *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1841-1852, Nov. 1999.
- [29] B. Friedlander, “Using MIMO to increase the range of wireless systems,” *Proceedings of Conference on Signals, Systems and Computers*, Oct. 2005.
- [30] A. Goldsmith and P. Varaiya, “Capacity of fading channels with channel side information,” *IEEE Transaction on Information Theory*, vol. 43, pp. 1986-1992, 1997.
- [31] Z. He, Q. Zhang, and V. Iversen Trunk, “Reservation in multi-service networks with BPP traffic,” *EuroNGI Workshop 2006*, Sitges, Spain, 2006.
- [32] K. J. Hole, H. Holm, and G. E. Øien, “Adaptive multidimensional coded modulation over flat fading channels,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 7, pp. 1153-1158, July 2007.
- [33] D. Hong and S. S. Rappaport, “Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures,” *IEEE Transactions on Vehicular Technology*, Aug. 1986.
- [34] E. Hossain and K. K. Leung (Eds.), *Wireless Mesh Networks: Architectures and Protocols*, Springer-Verlag, 2007.
- [35] A. Hottinen, O. Tirkkonen, and R. Wichman, *Multi-antenna Transceiver Techniques for 3G and Beyond*, Wiley Sons, Inc., 2003.
- [36] L. Hu, “Distributed code assignments for CDMA packet radio networks,” *IEEE/ACM Trans. Networking*, vol. 1, pp. 668-677, Dec. 1993.
- [37] M. Hu and J. Zhang, “MIMO ad hoc networks: Medium access control, saturation throughput, and optimal hop distance,” *Journal of Communications and Networks, Special Issue on Mobile Ad Hoc Networks*, pp. 317-330, Dec. 2004.
- [38] M. Hyung, G. Pan, K. Olesen, and R. Donald “Peak power characteristics of single carrier FDMA MIMO precoding system” *Proceedings of IEEE VTC*, Oct. 2007.
- [39] R. Jain and E.W. Knightly, “A framework for design and evaluation of admission control algorithms in multi-service mobile networks,” in *Proceedings of Annual Joint Conference of the IEEE Computer and Communications Societies 1999*, pp. 1027-1035, March 1999.
- [40] J. Jun, “Capacity estimation of wireless mesh networks,” *Masters Thesis*, Depart-

- ture of Electrical and Computer Engineering, North California State University, Jul. 2002.
- [41] J. Jun and M.L. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Wireless Communication*, vol. 10, no.1, pp. 8-14, Oct. 2003.
- [42] Y.-D. Kim, I. Kim, J. Choi, J.-Y. Ahn, and Y. H. Lee, "Adaptive modulation for MIMO systems with V-BLAST detection," in *Proceedings of IEEE VTC 2003-Spring*, vol. 2, pp. 1074-1078, Apr. 2003.
- [43] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1746-1755, Sept. 2004.
- [44] William S. Lovejoy, "Policy bounds for Markov decision processes," *Operations Research*, vol. 34, no. 4, pp. 630-637, Aug. 1986.
- [45] I. Koutsopoulos and L. Tassiulas, "Adaptive resource allocation in SDMA-based wireless broadband networks with OFDM signaling," *Proceeding of IEEE INFOCOM*, vol. 3, pp. 1376-1385, Jun. 2002.
- [46] P. H. Lehne and M. Pettersen, "An overview of smart antenna technology for mobile communication systems," *IEEE Communications Surveys*, pp. 2-13, vol. 2, no. 4, 1999.
- [47] J. A. Morrison, "Two discrete-time queues in tandem," *IEEE Transactions on Communications*, vol. 27, no. 3, pp. 563-573, Mar. 1979.
- [48] M. Ngo and V. Krishnamurthy, "On the optimality of threshold scheduling policies for video transmission in Markovian fading wireless channels with channel-aware ARQ," *Proceedings of IEEE Global Telecommunications Conference (GlobeCom06)*, 2006, San Francisco.
- [49] D. Niyato and E. Hossain, "Radio resource management in MIMO- OFDM based wireless infrastructure mesh networks: Issues and approaches," *IEEE Communications Magazine*, vol. 45, no. 11, pp. 100-107, Dec. 2007.
- [50] C. Pan, Y. Cai, and Y. Xu, "Multipacket reception in MIMO-OFDM systems," *Proceedings of IEEE International Symposium on Communications and Information Technology*, vol. 2, pp. 1156-1159, Oct. 2005.
- [51] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single-node case," *IEEE/ACM Transactions on Networking*, vol. 1, no. 1, pp. 344-357, June 1993.
- [52] M. Park, S. Nettles, and R. W. Heath, Jr., "Improving throughput and fairness for MIMO ad hoc networks using antenna selection diversity," *Proceeding of IEEE Global Telecommunications Conference*, vol. 1, pp. 3363-3367, Dallas, TX, Nov. 29 - Dec. 3, 2004.

- [53] A. J. Paulraj, D. A. Gore, R.U. Nabar, and H. Boelcskei, "An overview of MIMO communications - A key to gigabit wireless," *Proceeding of the IEEE*, vol. 92, no. 2, Feb. 2004, pp. 198-218.
- [54] A. Paulraj and C. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49-83, Nov. 1997.
- [55] S. Sfar, R. D. Murch, and K. B. Letaief, "Layered spacetime multiuser detection over wireless uplink systems," *IEEE Transaction on Wireless Communications*, vol. 2, no. 4, pp. 653-668, July 2003.
- [56] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multi-user MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, Feb. 2004, pp. 461-471.
- [57] S. C. Swales, M. A. Beach, D. J. Edwards, and J. P. McGeehan, "The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 39, pp. 56-67, 1990.
- [58] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transaction on Information Theory*, pp. 1456-1467, vol. 45, no. 5, Jul. 1999.
- [59] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744-765, Mar. 1998..
- [60] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transaction on Telecommunications*, vol. 10, pp. 585-595, Dec. 1999.
- [61] C.-K. Toh, M. Delwar, and D. Allen, "Evaluating the communication performance of an ad hoc wireless network," *IEEE Transaction on Wireless Communications*, vol. 1, no. 3, pp. 402-414, July 2002.
- [62] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.
- [63] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Transaction on Wireless Communications*, vol. 4, no. 5, pp. 2318-2328, Sept. 2005.
- [64] H. S. Wang and N. Moayeri, "Finite-state Markov channel: A useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163-171, Feb. 1995.
- [65] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3, pp. 279-292, 1992.
- [66] J. H. Winters, "On the capacity of radio communication systems with diversity

- in Rayleigh fading environments," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 871-878, Jun. 1987.
- [67] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," *Proceedings of URSI International Symposium on Signals, Systems, and Electronics*, Sept.-Oct. 1998.
- [68] M. Xie and M. Haenggi, "Delay performance of different MAC schemes for multihop wireless networks," *Proceeding of GLOBECOM'05*, Dec. 2005.
- [69] J. Xu, J. Kim, W. Paik, and J. Seo, "Adaptive resource allocation algorithm with fairness for MIMO-OFDMA system," *Proceedings of IEEE Vehicular Technology Conference*, 2006.
- [70] W. Ying, S. Xiaodong, and Z. Ping, "Admission control in MIMO-based virtual group cell systems" *Proceedings of Vehicular Technology Conference 2004*, Sept. 2004.
- [71] H. Zhai, X. Chen, Y. Fang, "How well can the IEEE 802.11 wireless LAN support quality of service?," *IEEE Transaction on Wireless Communications*, vol. 4, no. 6, pp. 3084-3094, Nov. 2005.
- [72] Y. Zhang and K. Letaief, "Adaptive resource allocation for multiaccess MIMO/OFDM systems with matched filtering," *IEEE Transactions on Communications*, vol. 53, no.11, pp. 1810-1816, 2005.
- [73] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.
- [74] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?," *IEEE Transaction on Wireless Communication*, vol. 3, no. 4, pp. 1285-1294, Jul. 2004.
- [75] Z. Zhou, B. Vucetic, M. Dohler, and Y. Li, "MIMO systems with adaptive modulation," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 5, pp. 1828 - 1842, Sept. 2005.
- [76] M. Zorzi, J. Zeidler, A. Anderson, A. L. Swindlehurst, M. Jensen, S. Krishnamurthy, B. Rao, and J. Proakis, "Cross-layer issues in MAC protocol design for MIMO ad hoc networks," *IEEE Wireless Communications*, Special Issue on Smart Antennas, Aug. 2006.
- [77] R. Pabst "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communication Magazine*, vol. 42, no. 9, pp. 80-89, Sept. 2004.

VITA

Surname: An

Given Names: Thai Nguyen

Place of Birth: Vietnam

Date of Birth: Aug. 16, 1981

Educational Institutions Attended

Hanoi University of Technology (HUT) 1999 to 2004

Degrees Awarded

B.E. HUT 2004

Honors and Awards

Telecommunication Research Labs (TRLabs) Research Scholarship 2006-2008