CROSS-LAYER DESIGN OF ADMISSION CONTROL POLICIES IN CODE DIVISION MULTIPLE ACCESS COMMUNICATIONS SYSTEMS UTILIZING BEAMFORMING

by

Wei Sheng

A thesis submitted to the

Department of Electrical and Computer Engineering

in conformity with the requirements

for the degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

August 2008

Abstract

To meet growing demand for wireless access to multimedia traffic, future generations of wireless networks need to provide heterogenous services with high data rate and guaranteed quality-of-service (QoS). Many enabling technologies to ensure QoS have been investigated, including cross-layer admission control (AC), error control and congestion control.

In this thesis, we study the cross-layer AC problem. While previous research focuses on single-antenna systems, which does not capitalize on the significant benefits provided by multiple antenna systems, in this thesis we investigate cross-layer AC policy for a code-division-multiple-access (CDMA) system with antenna arrays at the base station (BS). Automatic retransmission request (ARQ) schemes are also exploited to further improve the spectral efficiency.

In the first part, a circuit-switched network is considered and an exact outage probability is developed, which is then employed to derive the optimal call admission control (CAC) policy by formulating a constrained semi-Markov decision process (SMDP). The derived optimal policy can maximize the system throughput with guaranteed QoS requirements in both physical and network layers.

In the second part, a suboptimal low-complexity CAC policy is proposed based on an approximate power control feasibility condition (PCFC) and a reduced-outage-probability

algorithm. Comparison between optimal and suboptimal CAC policies shows that the suboptimal CAC policy can significantly reduce the computational complexity at a cost of degraded performance.

In the third part, we extend the above research to packet-switched networks. A novel SMDP is formulated by incorporating ARQ protocols. Packet-level AC policies are then proposed. The proposed policies exploit the error control capability provided by ARQ schemes, while simultaneously guaranteeing QoS requirements in the physical and packet levels.

In the fourth part, we propose a connection admission control policy in a connection-oriented packet-switched network, which can guarantee QoS requirements in physical, packet and connection levels. By considering joint optimization across different layers, the proposed optimal policy provides a flexible way to handle multiple QoS requirements, while at the same time, maximizing the overall system throughput.

Acknowledgments

First and foremost, I would like to thank my supervisor, Dr. Steven Blostein, for his excellent guidance, encouragement, patience and support while doing this research.

My appreciation goes to the thesis defense committee members: Dr. J. Mark from University of Waterloo, Dr. H. Hassanein, Dr. P. J. McLane and Dr. C.E. Saavedra, for their taking time to review my thesis and for their comments and suggestions with respect to this thesis.

My friends at Queens University have provided great encouragement and assistant through the years. Special thanks to Jinpeng Wang, Jing Gai, Minhua Ding, Neng Wang, Yi Song, Constantin Siriteanu and Yu Cao for their help, friendship and the wonderful time we shared.

I would like to thank my parents and other family members for their support and understanding. I could never thank my mother and my mother-in-law enough for their efforts in taking care of my baby while I am doing this research.

Finally, I am especially grateful to my husband Yang Lu and my baby boy David for every hard time they headed me through and for every day they have been in my life.

Contents

List of Tables

List of Figures

Acronyms

AC	Admission Control
ACK	Acknowledgement
AoA	Angle-of-Arrival
AOP	Average Outage Probability
ARQ	Automatic Retransmission Request
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BS	Base Station
BPSK	Binary Phase Shift Keying
CAC	Call Admission Control
CDMA	Code Division Multiple Access
CS	Complete Sharing
DS	Direct Sequence
FER	Frame Error Rate
FDMA	Frequency Division Multiple Access
GSMP	Generalized Semi-Markov Process
HSUPA	High Speed Uplink Packet Access
LDPC	Low-Density Parity-Check
LMMSE	Linear Minimum Mean Square Error

LOS	Line-Of-Sight
LP	Linear Programming
MAC	Media Access Control
MAI	Multiple Access Interference
MC	Multicarrier
MCRL	MultiCriterion Reinforcement Learning
MDP	Markov Decision Process
MIMO	Multiple-Input Multiple-Output
MS	Mobile Station
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiplexing Access
PAC	Packet Access Control
PCFC	Power Control Feasibility Condition
PDF	Probability Density Function
PER	Packet Error Rate
PIC	Parallel Interference Cancellation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
ROP	Reduced-Outage-Probability
RRM	Radio Resource Managment
SIC	Successive Interference Cancellation
SIR	Signal-to-Interference Ratio
SMDP	Semi-Markov Decision Process
SNR	Signal-to-Noise Ratio
TDMA	Time Division Multiple Access

UMTS	Universal Mobile Telecommunications System
VA	Voice Activity
WCDMA	Wideband CDMA
WSOP	Worst State Outage Probability

List of Important Symbols

J	Number of classes
M	Number of antennas
W	System bandwidth
η_0	Spectral density for AWGN
κ	Voice activity indicator
p_v	Voice activity factor
K	Number of active users
$\mathbf{a}_i$	Array response vector for user i
$\mathbf{w}_i$	Beamforming weight for user i
R_j	Data rate for class j
γ_j	Target SIR for class j
λ_j	Arrival rate for class j
μ_j	Departure rate for class j
B_j	Buffer size for class j
n_s^j	The number of class j sessions
n_q^j	The number of class j sessions in the queue
$\mathbf{s}$	System state
S	System state space
$\mathbf{a}$	Action

$A_{\mathbf{s}}$	Feasible action space for state **s**
ρ_j	Target PER for class j
α_{dec}	Decrease factor
α_{inc}	Increase factor
ρ_{av}	Average outage probability constraint
ρ_w	Worst-state outage probability constraint
$r_{a,j}$	Packet arrival rate for class j user
$r_{d,j}$	Packet departure rate for class j channel
L_j	Maximum number of retransmissions for class j packet
$(\cdot)^t$	Matrix or vector transpose
$(\cdot)^*$	Complex conjugate
$(\cdot)^H$	Matrix or vector conjugate transpose
$\|\boldsymbol{x}\|$	A norm of a vector $\boldsymbol{x}$
$\upsilon(.)$	Maximum eigenvalue

Chapter 1

Introduction

1.1 Motivation

Recently, there has been significant growth in the use of wireless communications. The success of the second-generation (2G) mobile systems, such as GSM, IS-95 and US-TDMA (IS-136), prompted the development of third-generation (3G) mobile systems [44]. 3G systems were designed to provide high-data-rate multimedia mobile services with varied quality-of-service (QoS) requirements. The standard requirements specify a data rate of 384 Kb/s for outdoor devices moving at high speeds, and 2 Mbps for devices moving at pedestrian speeds [22]. During the evolution from 2G to 3G, a range of wireless networks and systems, including General Packet Radio Service (GPRS), cdma2000 (www.tiaonline.org), wideband CDMA (WCDMA) [58], WiFi, WiMax, HomeRF, Bluetooth and infostations [22], have been developed. Researchers are currently developing programs for beyond 3G networks.

Future generations of wireless networks will enable heterogeneous services, such as voice, data, wireless broadband access, video chat, high definition TV content, digital video broadcasting (DVB) and other streaming services, with QoS constraints and a variety of data rates that may even reach up to the order of a gigabit per second [22].

The quality-of-service (QoS) and high data rate requirements for future wireless networks [29], together with the rapidly increasing number of mobile subscribers and the demand for multimedia services, pose new technical challenges. The limited radio resource and hostile wireless communication environment [66], such as multipath fading, interference and user mobility, degrade the QoS and as a result further increase the challenges. Therefore, improving spectrum utilization subject to QoS guarantees is a major design objective for future wireless networks. To achieve this goal, call admission control (CAC) is increasingly becoming important, which represents a good compromise between high resource utilization and satisfactory service provisioning [88].

Various CAC approaches for controlling QoS are proposed in the literature, e.g., [19] [27] [41] [45] [56], and a comprehensive survey on CAC policies is provided in [4]. In order to guarantee the QoS requirements in different layers while simultaneously maximizing the long-term system throughput, it is necessary to perform a joint optimization over the physical layer and the upper layers, i.e., the CAC policy should be designed across layers.

There has been previous research on cross-layer CAC policy design. To mention a few, in [24] and [73], optimal call admission control policies at the network layer and the power control in physical layer are discussed for an integrated voice/data DS-CDMA system with a linear minimum mean-square estimation (LMMSE) receiver. In [88], an effective bandwidth based CAC scheme is proposed for the uplink of a CDMA cellular system that supports heterogeneous data traffic with self-similarity [88]. In [99], optimal admission control schemes are proposed in CDMA networks with variable bit rate packet multimedia traffic. In [52], a multicriterion reinforcement learning (MCRL)-based adaptive admission control method is proposed for a low-density parity-check (LDPC) multi-rate multiuser system, in which the admission control problem with multiple QoS constraints is formulated as a multicriterion decision problem, and hence can be solved by the MCRL algorithms.

These algorithms integrate the AC policy design with a specific physical-layer signal model, and as a result, are able to guarantee QoS requirements while optimizing system performance across physical and upper layers, which leads to an improved spectral efficiency. However, the above mentioned work on cross-layer CAC design only considers single antenna systems, which lack the performance benefits provided by multiple antenna systems [61] [79].

Antenna arrays are one of the key techniques that can mitigate the multipath fading and interference, and as a result can help to achieve the requirements for high speed data services in 3G and beyond wireless systems [61]. It has been proven that with multiple antennas at the transmitter and/or receiver side, spatial diversity as well as capacity gain can be achieved [5] [11] [26] [33] [39]. When designing CAC policy across different layers, the significant performance gain in the physical layer can lead to a significant performance gain in the upper layers, and as a result, the overall system throughput can be dramatically improved. Currently, in the literature, cross-layer CAC design and multiple antenna systems are investigated separately, and hence the benefits from both techniques are not fully employed.

Another technique to mitigate fading and interference, which leads to an increased capacity, is automatic retransmission request (ARQ) [36] [74] [90]. An ARQ scheme retransmits an incorrectly received packet until it is correctly received or the maximum number of retransmissions is reached. ARQ provides an alternative way in improving the system throughput and is widely adopted in wireless networks. However, to the best of our knowledge, no admission control design in the literature incorporates ARQ, which lacks a powerful error control capability.

Although antenna arrays, ARQ and cross-layer CAC design are very effective in improving the system performance as well as the spectral efficiency, they are designed individually in the existing literature. To fully exploit the benefits provided by these techniques, we investigate the cross-layer admission control problem in the presence of both antenna arrays and error control schemes. The objective is to investigate admission control (AC) policies which maximizes the overall system throughput, while simultaneously guaranteeing QoS requirements. The system throughput is defined as the number of correctly received sessions per unit time. For a circuit-switched network, the term *session* denotes a call, while for a packet-switched network, the term *session* represents a packet.

1.2 Thesis Overview

This thesis includes seven chapters that investigate the cross-layer admission control problem for CDMA beamforming systems. Chapter 2 briefly reviews the background and the related literature, and Chapter 7 summarizes the results and indicates possible future directions. The main body of this thesis consists of Chapters 3-6, which are organized as follows:

Chapter 3 investigates how to develop an optimal cross-layer CAC policy for multiple antenna systems. With multiple antennas at the base station (BS), spatial filtering is employed at the receiver to suppress interference, which results in a fluctuating signal-to-interference ratio (SIR), leading to a non-zero outage probability in the physical layer. In this chapter, an exact approach is studied to control the outage probability. Based on this exact approach, an optimal admission control policy is proposed by formulating a constrained semi-Markov decision process (SMDP). The proposed CAC policy can maximize the overall system throughput while simultaneously guaranteeing QoS requirements in both physical and network layers.

The above optimal CAC policy requires high computational complexity. In Chapter 4, an approximate approach is studied to control the outage probability, which includes a linear approximate power control feasibility condition (PCFC) and a separate reduced-outage-probability (ROP) algorithm. Based on this approximate approach, a suboptimal call admission control (CAC) policy is proposed. Compared with the optimal CAC policy, the suboptimal CAC policy can dramatically reduce the complexity with slightly degraded performance.

In the above two chapters, CAC policies are proposed for circuit-switched networks in which the resource requirements for each accepted user remain unchanged during the whole connection. Circuit-switched networks feature the first and the second generation of wireless communications. With the significant growth of the internet and increasing demands for wireless data services, packet-switching technology is currently employed to provide multimedia services to mobile users [88]. In Chapter 5, we investigate admission control policies for a packet-switched network, in which admission control is performed at the packet level and a connection is not necessary. Admission control policies block packets instead of blocking the whole connection, and as a result, can efficiently utilize resources for bursty traffic. In this chapter, to take into account the impacts of a truncated ARQ scheme, a novel semi-Markov decision process (SMDP) formulation is required. An optimal AC policy as well as a low-complexity suboptimal AC policy are then discussed.

In Chapter 6, we investigate the admission control problem in a more complicated connection-oriented packet-switched network. In contrast to the packet-switched network discussed in Chapter 5, in which the connection is not established and the QoS requirements in the connection level are ignored, in Chapter 6, a connection is employed and connection level QoS requirements are also taken into account. We propose an optimal connection admission control policy which employs the benefits provided by both multiple antennas and ARQ schemes. The proposed policy is capable of maximizing the system throughput while simultaneously satisfying all the QoS requirements in the physical layer as well as packet and connection levels.

In this thesis, we focus on a single-cell system, in which the uplink and downlink are treated in one cell. User mobility, handoff and backbone networks are ignored.

Throughout this thesis, a code-division-multiple-access (CDMA) system is considered, which has shown promise in mitigating the multipath fading and interference, and as a result achieves a high capacity. We here consider a CDMA system because of its strong interaction among different layers, while for frequency-division-multiple-access (FDMA) and time-division-multiple-access (TDMA) systems, user capacity is determined by fixed resources such as frequency and time slots, and therefore, CAC design for FDMA/TDMA systems can be performed relatively independently of the physical layer design. For some multiple access systems, such as orthogonal frequency-division multiple access (OFDMA), there may still exist a strong interaction across different layers. For example, user capacity in an OFDMA system depends on QoS requirements, system parameters and resource allocation schemes. We remark that the proposed AC policies for CDMA multiple antenna systems in this thesis can be further extended to FDMA, TDMA, OFDMA as well as other multiple access systems provided that the user capacity region, i.e., the maximum number of users that the system can accommodate, is available.

1.3 Contributions

The primary contributions of this thesis are as follows:

- An exact approach is provided for beamforming systems to ensure the physical layer

QoS, and based on this exact approach, an optimal maximum-throughput CAC policy is proposed which guarantees QoS requirements in both physical and network layers. While cross-layer CAC design and multiple antenna systems are extensively studied in the literature, it is the first time that these two aspects are jointly considered, so that the benefits provided by both techniques can be fully exploited. In contrast to existing cross-layer CAC policies, which only optimize network layer performance, our proposed optimal CAC policy can also optimize the system throughput, which represents overall system performance across different layers.

- An approximate approach is provided to ensure physical layer QoS, and based on this approximate approach, a low-complexity suboptimal CAC policy is proposed. While the optimal CAC policy requires high computational complexity, especially for this system under consideration that lacks a closed-form analytical expression for outage probability, the proposed suboptimal CAC policy can dramatically reduce complexity. This low-complexity suboptimal policy can also be applied to more general systems, which provides a simple yet effective approach to an otherwise very complicated problem.

- The packet admission control problem is formulated as a novel semi-Markov decision process (SMDP) by considering the impacts of ARQ, and based on the formulated SMDP, packet admission control policies are then derived. While ARQ is widely employed in practical wireless systems to mitigate transmission errors, in the literature there is no semi-Markov decision process formulation which incorporates ARQ. Our formulated SMDP makes it possible to employ the powerful semi-Markov decision process model to solve the packet-level AC problem for systems utilizing ARQ.

- An optimal connection-level admission control policy is designed for packet switched networks. This policy provides a novel framework for joint optimization among multiple antennas in the physical layer, ARQ schemes in the data-link layer and cross-layer connection admission control design in the network layer. As a result, multiple QoS requirements can be handled more flexibly to achieve maximum system throughput.

Chapter 2

Background

This chapter briefly reviews the pertinent background and related literature.

2.1 Multiple Access

In communication networks, a multiple access scheme allows several sessions to share the same communication channel.

Frequency division multiplexing access (FDMA) and time division multiplexing access (TDMA) are two well-known multiple access approaches which are widely used in narrowband systems such as GSM and IS-136. In FDMA or TDMA, the available channel is divided into several sub-channels which occupy non-overlapping frequency bands or time slots. Each sub-channel is assigned to each user upon request. The narrowband network using FDMA or TDMA can be simplified and approximated by a collection of point-to-point non-interfering links, and the physical-layer issues are essentially point-to-point ones [83].

FDMA and TDMA systems suffer from some weaknesses. For example, all users are assumed to transmit continuously, which is not true for circuit-switched voice and bursty traffic transmission. Also, TDMA and FDMA systems have hard capacity limits, which depend on the number of frequency bands or time slots. To mitigate these weaknesses, CDMA and OFDMA are proposed and are widely used in current and future envisioned wireless networks, in which all transmitted signals are spread across the available bandwidth. The key feature of these systems is universal frequency reuse: the same frequency band is used in every cell [83], and different users are not necessarily occupying orthogonal sub-channels.

For a CDMA system, which is based on direct-sequence spread-spectrum, a user's information stream is modulated by pseudonoise sequences. Each communication will be allocated the entire spectrum all of the time. CDMA uses codes to identify individual transmission sessions. In CDMA systems, interference is the most significant factor in determining system capacity and call quality. Any techniques which can suppress interference can increase capacity. Therefore, CDMA systems have a soft capacity.

In an OFDMA system, on the other hand, a user's information is spread by hopping in the time-frequency grid and the transmissions within a cell can be kept orthogonal. However, adjacent cells share the same bandwidth and inter-cell interference exists [83].

In the above, we have discussed dedicated channel assignment methods, in which each user can be assigned a different channel for some period of time. However, some users do not require continuous transmission, so dedicated channelization can be extremely inefficient from a resource utilization viewpoint. An alternative to overcome this disadvantage is random access [59]. In random access, the multiple users compete for a set of channels [63]. The signals from different users may be transmitted simultaneously over the same channel. Since these signals are not distinguished by specific time slot, frequency band, code sequences or spatial filtering via beamforming, the receiver cannot separate them. As a result, when more than one user attempts to use the same channel simultaneously, these transmissions collide and interfere with one another. When a collision occurs, the information is lost and must be re-transmitted. To resolve conflicts, and minimize re-transmissions and delay, protocols are needed to handle the random access and re-transmission. Some typical protocols are Aloha [2], slotted Aloha [17], as well as CSMA/CD (carrier sense multiple access with collision detection) [48]. Discussion of the stability issues of random access protocols can be found in [3].

2.2 CDMA

In FDMA and TDMA systems, the available channel is partitioned into independent single-user sub-channels, and as a result, a system designed for single-user communications is directly applicable and no new problems are encountered [63]. For a CDMA system, interference mitigation and power control for one user impact the performance of other users as well. Therefore, for CDMA systems, there are strong interactions among the design for different users. In the following, we briefly discuss some related background on CDMA.

2.2.1 Single-user detection and multiuser detection

In CDMA, the transmitted signals from multiple users occupy the same time slots and frequency bands, and are distinguished by non-orthogonal code sequences. At the transmitter, the signal is spread by unique spreading codes, and then transmitted in a channel, which is below noise level. The receiver uses a correlator to despread the desired signal. Spreading codes are noise-like pseudo-random codes. The spreading factor is the ratio of the chip rate to baseband information rate.

The receiver can perform independent detection for each user, or joint detection for multiple users. For independent detection, the receiver only knows the code sequence of the desired user, and regards the signals from all the other users as interference. Independent signal detection schemes are easy and simple to implement. However, the independent detection, in combination with tight power control, is only optimal under a white Gaussian noise model for the multiple-access interference (MAI) [22]. Furthermore, it cannot increase system capacity with an increase of users [63]. To effectively mitigate interference and improve channel capacity, significant research has occurred on joint detection, also known as multi-user detection [62] [86] [87] [89]. In multiuser detection, the code sequences for all the multiple users are available at the receiver, and the receiver employs the underlying structure of the received spread signals to mitigate the MAI, and as a consequence, improve the system capacity [89]. An optimal multi-user receiver has exponential computation complexity [63], so sub-optimal multi-user detection methods have been studied, including the de-correlating detector [64], MMSE detector [63], successive interference cancellation (SIC) and multistage interference cancellation (MIC) detectors [15] [60] [84].

2.2.2 Voice activity

Voice activity [35] is one of the very important advantages for CDMA, which can be employed to mitigate the interference and hence increase the capacity.

Voice activity implies that a voice user may transit between an active state (ON-state) and an inactive state (OFF-state). When the user is at OFF state, i.e., the user is in a silent period, transmission is suppressed for that user, and the resources allocated to that user can be temporarily released to other users. Voice activity factor, which represents the time percentage that a voice user is active, is typically chosen from 35% to 40% [13].

With considerations of voice activity, a voice user can be modeled as an ON/OFF

Markov model. The transition probability from an ON state to an OFF state is denoted by α, and the transition probability from OFF state to ON state is denoted by β. The stationary probability that a voice user is in ON state can be obtained by [99]

$$P_v = \frac{\beta}{\alpha + \beta}$$

and the stationary probability that a voice user is in OFF state is $1 - p_v$. Among the K_1 voice users, the number of state ON users has a Binomial distribution with success rate p_v.

2.2.3 Power control

Different from TDMA/FDMA, in which power control can be performed user by user, power control in CDMA system must be jointly performed for multiple users, since CDMA systems are interference-limited and suffer from a phenomenon known as the near-far effect where strong users significantly degrade the performance of the weak users [97].

Reverse link power control methods in 3G WCDMA and cdma2000 include open loop and closed loop. For open loop power control, a mobile adjusts the transmitted power according to its received level from the base station, while closed loop power control includes inner loop and outer loop power control. Inner loop power control aims to keep the mobile as close to its target SIR as possible. The uplink outer loop power control is responsible for setting a target SIR.

In this thesis, we discuss a signal-to-interference ratio (SIR)-based power control in which the transmitted power for each user is adjusted adaptively to achieve a target SIR. With a temporally matched filter receiver, i.e., single user detection, the achieved signal-to-interference ratio at the base-station (BS) for a desired user k can be obtained as

$$SIR_k = \frac{W}{R_k} \frac{P_k h_k^2}{\sum_{i \neq k} P_i h_i^2 + \eta_0 W} \tag{2.1}$$

where W and R_k denote the bandwidth and data rate for desired user k, P_i and h_i denote the transmitted power and the channel gain for user i, respectively, and η_0 denotes the one-sided power spectral density of background additive white Gaussian noise (AWGN).

To reduce the interference to other cell, power control scheme aims to minimize the total transmitted power from all the users while satisfying the QoS requirements in terms of SIR. As shown in [68], an optimal power solution satisfying the above requirements should achieve the target SIR with equality, i.e.,

$$\gamma_k = \frac{W}{R_k} \frac{P_k h_k^2}{\sum_{i \neq k} P_i h_i^2 + \eta_0 W} \tag{2.2}$$

where $k = 1, .., K$, and γ_k denotes the target SIR for user k.

By grouping the above K equations, we have the following matrix form

$$[I_K - Q]\mathbf{p} = \mathbf{u} \tag{2.3}$$

where I_K is a $K-$dimensional identity matrix, power vector $\mathbf{p} = [P_1 h_1^2, .., P_K h_K^2]^t$, $(.)^t$ denotes transpose,

$$Q = \begin{bmatrix} 0 & \frac{\gamma_1 R_1}{W} & \ldots & \frac{\gamma_1 R_1}{W} \\ \frac{\gamma_2 R_2}{W} & 0 & \ldots & \frac{\gamma_2 R_2}{W} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\gamma_K R_K}{W} & \frac{\gamma_K R_K}{W} & \ldots & 0 \end{bmatrix} \tag{2.4}$$

and $\mathbf{u}$ is a diagonal matrix with the i^{th} element as $\eta_0 \gamma_i R_i$.

The optimal power solution can be obtained by solving the above K equations [68]

$$P_k = \frac{\eta_0 W}{h_k^2 (1 + \frac{W}{\gamma_k R_k})[1 - \sum_{i=1}^{K} \frac{1}{1 + \frac{W}{\gamma_i R_i}}]} \tag{2.5}$$

where $k = 1, .., K$.

The positivity of the power solution implies the following power control feasibility condition

$$\sum_{i=1}^{K} \frac{1}{1+\frac{W}{\gamma_i R_i}} < 1 \qquad (2.6)$$

which limits the maximum number of users that a system can accommodate under the QoS constraints.

We remark that if the condition in (2.6) holds, we say the system is feasible [94]. Inequality (2.6) is thus referred to as the power control feasibility condition (PCFC). With this condition, a positive power solution is always available which can satisfy QoS.

In a practical system, a central power control scheme according to (2.5) may not be easy to implement, and in this case a distributed power control scheme can be employed, in which the transmitted power for user i at time instant $k+1$ can be iteratively updated. According to the Foschini-Miljanic algorithm [34], an iteration function is given as follows

$$P_i(k+1) = \frac{\gamma_i}{SIR_i(k)} P_i(k) \qquad (2.7)$$

where $P_i(k)$ and $SIR_i(k)$ denote the transmitted power and the received SIR for user i at time instant k, respectively. For a feasible system, the Foschini-Miljanic algorithm in (2.7) converges from any initial power to the desired power in (2.5) [94].

In summary, CDMA is interference limited system, and the capacity can be increased by suppressing the interference. By employing multiuser detection, voice activity, power control, antenna arrays and any other interference mitigation techniques, much higher system capacity can be achieved than that in FDMA and TDMA.

2.3 Base Station Beamforming

A beamforming performs spatial filtering to separate signals that have overlapping frequency content but originate from different spatial locations [85]. The objective is to estimate the signal arriving from a desired direction in the presence of noise and interfering signals [85]. With beamforming at the base station (BS), interference can be dramatically suppressed, and as a result, the physical layer performance, in terms of signal-to-interference ratio (SIR), can be improved. In this section, we briefly review the pertinent background and literature on beamforming, and then illustrate how the physical layer performance can be improved by employing antenna arrays at the BS.

2.3.1 Antenna arrays at the BS

To perform beamforming, knowledge of the array response vector is required at the BS, which contains the relative phases of the received signals at each array element [93]. For example, with an M-element circularly antenna array at the BS, the array response vector for user i, denoted by $\mathbf{a}_i$, can be written as [93]

$$\mathbf{a}_i = \left[\frac{1}{M}e^{j\frac{\pi\cos(\theta_i)}{2\sin(\frac{\pi}{M})}}, \frac{1}{M}e^{j\frac{\pi\cos(\theta_i-2\pi/M)}{2\sin(\frac{\pi}{M})}}, .., \frac{1}{M}e^{j\frac{\pi\cos(\theta_i-2\pi(M-1)/M)}{2\sin(\frac{\pi}{M})}}\right]^t \tag{2.8}$$

which $j=\sqrt{-1}$, $(.)^t$ denotes transpose, and θ_i denotes the angle of arrival (AoA) for user i. The AoAs for different users are assumed to be independent and identically uniformly distributed in $[0,2\pi]$.

At the BS, a beamforming receiver consists of an array of small non-directional antenna elements, which can simulate a large directional antenna. By varying the amplitudes and phases of the elements in this array, the main beam of this synthesized directional antenna can be controlled [40]. The combined relative amplitude and phase shift for an antenna element is expressed as complex-valued weight or beamforming weighting coefficient.

Under the assumption that the distance between the desired mobile and the base station is large relative to the carrier wavelength, the incoming signals from that mobile can be treated as plane waves. By further assuming that the distance between adjacent antennas is half of the wavelength, the beam pattern can be derived. Denote $\mathbf{a}_i$ as the array response vector for a mobile i with direction of arrival θ_i, $\mathbf{w}_k$ as the beamforming weight vector for a desired mobile k with direction of arrival θ_k, and M the number of antenna elements in this array. Once the array response vector is obtained for a particular geometry, the beamforming pattern can be created as follows [93],

$$\phi_{ik}^2 = \left|\mathbf{w}_k^H \mathbf{a}_i\right|^2 \tag{2.9}$$

where $(.)^H$ denotes the conjugate transpose, and ϕ_{ik}^2 is the fraction of interferer i's signal passed by a desired k user's beamforming weights of the antenna array.

We remark that the above beam pattern can be modified to include mutual coupling and scattering [93]. For a desired user k, when mutual coupling and scattering are taken into account, Equation (2.9) becomes [93]

$$\phi_{ik}^2 = \left| \frac{(\mathbf{Z}^{-1}\mathbf{w}_k)^H}{\|\mathbf{Z}^{-1}\mathbf{w}_k\|} \frac{(\mathbf{Y}_i \mathbf{a}_i)}{\|\mathbf{Y}_i \mathbf{a}_i\|} \right|^2 \tag{2.10}$$

where $\|.\|$ denotes norm, $\mathbf{Z}^{-1}$ is the inverse of mutual impedance matrix $\mathbf{Z}$ [93], and $\mathbf{Y}_i$ is a diagonal matrix with elements $\{\upsilon_i r_{i1}, .., \upsilon_i r_{iM}\}$, in which υ_i denotes the path loss and shadowing effects factor for user i, and r_{im} represents Rayleigh fading random variables for user i at array element m , where $m = 1, .., M$, which depends on the given angle spread, $\triangle$. The detailed calculation of ϕ_{ik}^2 can be found in [93].

In this thesis, to highlight the cross-layer design across different layers, we consider an environment without mutual coupling and scattering for simplicity. However, cross-layer CAC design for a beamforming system with mutual coupling and scattering can be extended straightforward by using beam pattern in (2.10).

The above beam-pattern has a main lobe directed towards θ_k. Therefore, the signal of the desired mobile is easily passed through the beam-pattern while signals from the interfering mobiles located at other angles-of-arrival are suppressed [93].

There are different ways to choose beamforming weights according to what criterion is used. Some commonly used criteria for adaptive beamforming include minimum mean-square error (MMSE), maximum-SIR and minimum-variance.

2.3.2 Literature review

The use of beamforming in wireless communications has received a lot of interest. Optimum combining was studied in [91], and conventional fixed beamforming techniques are studied in [80]. Power control in beamforming wireless networks has been discussed in [32] and [96]. In [93], the performance of CDMA systems employing antenna arrays is investigated under more realistic signal propagation assumptions, where the performance degradation in digital beamforming due to the combination of mutual coupling, scatter, and imperfect power control and its impact on uplink CDMA system capacity is quantified. In [32], the joint problem of power control and beamforming is considered, in which an algorithm is provided for computing the transmission powers and the beamforming weight vectors. In [98], two commonly used receiver processing-based interference management methods: multiuser detection and receiver beamforming have been studied. In [85] an overview of beamforming is provided from a signal processing perspective. In [57], the capacity improvement of multicell CDMA cellular system with BS antenna array is studied for both the downlink and the uplink. In [76], the behavior of smart antennas is explored in

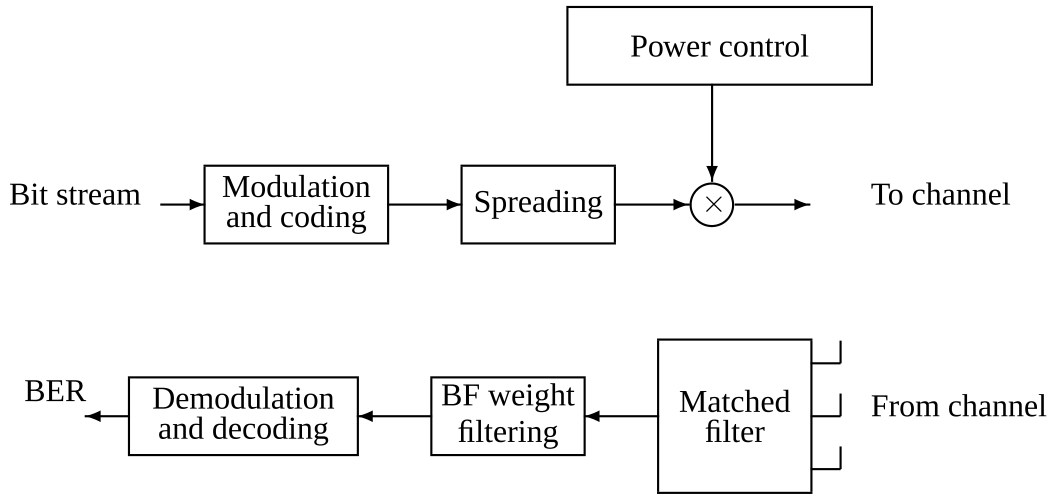

Figure 2.1. Transmitter and receiver structure for a CDMA beamforming system.

power controlled CDMA systems by analyzing and comparing the performance of optimal beamforming and spatially matched filter beamforming.

2.3.3 Employing beamforming to improve SIR

Beamforming can improve the SIR as well as channel capacity. In the following, we illustrate this point by giving a simple example.

Consider a CDMA beamforming system which has M antennas at the BS and a single antenna for each user. A temporal matched-filter receiver is employed at the base station. Suppose there are K active users in the system, and a channel with slow fading is assumed.

The transmitter-receiver structure is presented in Figure 2.1. The source bit stream is coded and modulated to an information symbol stream $b_i(t)$, which has a symbol rate of R symbols/s. The symbol stream is then spread to a wideband sequence with chip rate of R_c symbols/s. For user i, the wideband sequence, denoted by $s_i(t)$, is given by $s_i(t) =$

$\sum_n b_i(n)c_i(t-nT)$, where $b_i(n)$ is the coded symbol stream, and $c_i(t)$ is the spreading sequence. The spread signal, $s_i(t)$, multiplied by $\sqrt{P_i}$, is then transmitted over the fading channel, where P_i denotes the transmitted power for user i which is decided by power control scheme.

We assume the signature sequences of the interfering users appear as mutually uncorrelated noise. As shown in [32], the received signal-to-interference ratio (SIR) for a desired user k can be written as

$$SIR_k = \frac{W}{R_k} \frac{p_k \phi_{kk}^2}{\sum_{i \neq k} p_i \phi_{ik}^2 + \eta_0 W} \tag{2.11}$$

where W and R_k denote the bandwidth and data rate for user k, respectively, and the ratio $\frac{W}{R_k}$ represents the processing gain; $p_i = P_i h_i^2$ denotes the received power for user i, and η_0 denotes the one-sided power spectral density of background additive white Gaussian noise (AWGN); the parameter ϕ_{ik}^2 is defined in (2.9), which captures the effects of beamforming. In this thesis, we consider a spatially matched filter receiver, i.e., $\mathbf{w}_k = \mathbf{a}_k$.

The achieved SIR is a random process depending on the realizations of AoA as well as beamforming weights. With an increased M, ϕ_{ik}^2 is reduced, which leads to an improved SIR. Therefore, increasing the number of antennas at the BS can suppress the interference, and as a result, increase capacity.

2.4 Layered Architecture

Traditionally, a wireless network is organized as a series of relatively independent layers. The purpose of each layer is to offer certain services to the higher layers, shielding those layers from the details of how the offered services are actually implemented [78]. The layered architecture makes a network easy to standardize and flexible to update.

Layered architecture has been very successful for wire-line networks, and is the default

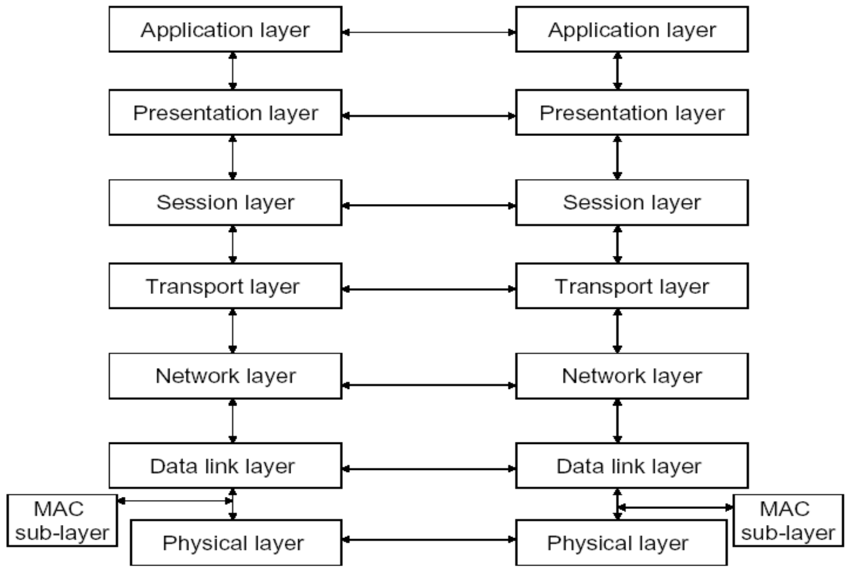

Figure 2.2. Open-system-interconnection (OSI) layered architecture

architecture for wireless networks [46]. A well-known and widely used architecture is the open system interconnection (OSI) model. The seven-layer OSI structure is shown in Figure 2.2. In this thesis, we mainly focus on the designs across the lower three layers. The detail for the other layers can be found in [9].

The physical layer, which is the bottom layer in the OSI architecture, provides transmission, reception and processing of signals [51]. This layer aims to transmit bits over a communication channel. In wireless networks, the physical layer combats fading with channel coding, spread-spectrum, and multiple antennas [51].

The QoS in the physical layer can be represented by a target bit-error-rate (BER) or packet-error-rate (PER), which can be equivalently mapped to a target SIR requirement. In a wireless communication network, we must allow for outage, defined as the probability that a target SIR, or equivalently, a target bit-error-rate (BER) or target packet-error-rate (PER), cannot be satisfied. Therefore, in this thesis, the QoS measurement in the physical layer is represented by a target outage probability. We consider two types of outage probability constraints: worst-state-outage-probability (WSOP) constraint, denoted by ρ_w, and average-outage-probability (AOP) constraint, denoted by ρ_{av}. While the WSOP constraint is very conservative which ensures that at any time instant and at any system state this outage probability constraint cannot be violated, the AOP constraint is less restrictive and only ensures a long-run average outage probability.

The data-link layer lies immediately above the physical layer, which can realize reliable transmission of groups of bits. The data entering this layer is first broken into groups of bits, e.g., packet, and overhead bits are added in each group (frame) to ensure error-free transmission. For multi-access communications, there is a need for a sub-layer to manage the media access control (MAC), which lies in the data-link layer [9]. Packet-level access control is performed at the MAC layer.

The layer above the data-link is the network layer, which controls network-related operations. This layer deals with routing, admission control and base station assignment (handoff). In current cellular wireless networks, the network layer assigns mobile stations to access points (base stations), and once a mobile is assigned to an access point, all communication occurs with that access point. Therefore, routing is not considered to be a major problem in cellular wireless networks [51].

The network-layer QoS requirements can be represented by blocking probability and connection delay. In the data link layer, packet-level QoS requirements can be characterized by packet delay and packet loss probability.

As we mentioned previously, layered architecture is the default architecture for wireless networks. It is being realized that the original layering concept is not very efficient for wireless networks, due to the fact that wireless channels are dynamically time-varying, and have inherent coupling between different layers [28]. Furthermore, the future wireless communications aim to provide heterogeneous service while satisfying QoS constraints in all the layers. Obviously, it is difficult for traditional independent-layer-design in wireless network to meet all the QoS constraints without interactions across layers, or meet the QoS constraints at a high cost (energy, bandwidth, etc.). This motivates cross-layer design for wireless networks [22].

2.5 Call Admission Control

When a user requests to access the network, the system decides if the user can be accepted. This is known as call admission control (CAC).

In cellular wireless communications, CAC is performed by an admission controller at the mobile switching center (MSC) [100]. Traffic of admitted calls is then controlled by other radio resource management (RRM) techniques such as scheduling, handoff, power, and rate control schemes [4]. Without loss of generality, in this thesis, the term *user* is not distinguished from the term *call*. For example, when a user generates a voice call and a data call at the same time, it is considered as two *users*, or equivalently, two *calls*.

CAC in wireless networks has been receiving a lot of attention, e.g., [6] [8] [19] [23] [27] [30] [37] [38] [41] [45] [49] [95], due to the growing popularity of wireless communications and the central role that CAC plays in QoS provisioning in terms of the signal quality, call blocking and dropping probabilities, packet delay and loss rate, and transmission rate [4].

There are many ways to categorize the current CAC policies. For example, CAC can be classified into parameter-based policies and measurement-based policies [4]. For a parameter-based policy, the incoming call is admitted or rejected based on some predictive or analytical assessment of the QoS constraints, while for a measurement-based policy, the incoming call may start transmission by transmitting some probing packets or using reduced power, and then deciding if the call can be admitted based on the QoS measurements during the initial transmission attempt [4].

The various CAC policies proposed for wireless networks can also be categorized into complete-sharing (CS) based CAC policies, threshold-based CAC policies, and cross-layer CAC policies [73]. In this thesis, we employ this approach to categorize the CAC policies.

For CS-based CAC policies, the incoming user can be admitted if and only if the physical layer SIR constraints can be satisfied, while for threshold-based CAC, the incoming user can be accepted if a system performance related threshold is not exceeded. A cross-layer CAC policy, which is very different from the above two CAC approaches, makes the admissibility decisions by solving a constrained optimization problem across layers.

Various CS-based and threshold-based CAC approaches for controlling signal quality are proposed in the literature. In [19] [27] [41] [45] [56], the incoming call is admitted if the interference level is less than a predefined threshold value. In [37] [45] [55] [92], the admission is based on the number of users or resource utilization factor. In [30], the maximum number of admissible users is determined using the effective bandwidth concept. In [6] [20], the incoming call is admitted if a feasible power allocation is determined. In [47] [49] [61], the total transmitted/received power is used as the admission criterion.

For complete-sharing-based CAC policy, e.g., [27] [45] [56], the incoming user is accepted if and only if the physical layer performance, e.g., signal-to-interference (SIR) ratio, can be satisfied, in which the users with heterogenous services are not distinguished, and QoS in network and packet levels are ignored.

For threshold-based policies, e.g., [55] [92], a performance related threshold for each traffic class is specified. Whenever this threshold is not exceeded, the incoming user can be accepted. Although threshold-based policies can distinguish different traffic classes by selecting thresholds to achieve bounds on outage and blocking probabilities [65], the threshold-setting process for the different classes is somewhat ad hoc [65], and may perform poorly in practice [73].

Recently, the admission control problem of ensuring multiple quality-of-service (QoS) requirements in different layers is receiving much attention, which motivates the cross-layer CAC design. In [24] and [73], CAC policies at the network layer and the power control at the physical layer are jointly designed for a DS-CDMA system with a linear minimum mean-square estimation (LMMSE) receiver. In [88], an effective bandwidth based CAC scheme is proposed for a CDMA cellular system that supports heterogeneous data traffic with self similarity [88]. In [99], optimal CAC schemes are proposed in CDMA networks with variable bit rate packet multimedia traffic. In [52], a multicriterion reinforcement learning (MCRL)-based adaptive admission control method is proposed, in which the admission control problem with multiple QoS constraints is formulated as a multicriterion decision problem.

Cross-layer CAC policies can be determined by optimizing some objective function subject to signal quality constraints, and are usually solved by semi-Markov decision process (SMDP). In this thesis, we will focus primarily on SMDP-based cross-layer AC design.

2.6 Cross-Layer SMDP-based CAC Policy

Before we discuss how to formulate and solve the CAC problem by a SMDP, we first need to introduce key definitions related to Markov chains and Markov decision processes (MDP).

If the future probabilistic behavior of a process depends only on the present state of the process and is not influenced by its history, the process has a Markovian property [81]. A

Markov chain [67] is a random sequence in which a Markovian property holds, i.e., the current state information is sufficient to predict the future stochastic process. A Markov chain can be employed to represent a dynamic system provided that the next state of the system depends only on the current state. In practical systems, we may have a dynamic system evolving over time where the state transition probabilistic law depends on not only the state, but also on the decision taken at that state. With sequentially made decisions, costs are also incurred which can represent QoS requirements and performance. With fixed decision epoches, the above system model is referred to as a Markov decision process (MDP). In many optimization problems, such as the AC problem investigated in this thesis, the times between consecutive decision epoches are not identical but are random. In this case, the system can be modeled by a semi-Markov decision process (SMDP).

In summary, a MDP is a Markov chain with action-dependent transition probabilities, while a semi-Markov decision process is a random process that changes state in accordance with a Markov decision process but takes a random amount of time between transitions. Semi-Markov decision processes (SMDP) occur widely in economics and operations research [65]. For the AC problem we investigate, we track arrival and departure processes whose instants of initiation are Poisson distributed and whose durations are independent and exponentially distributed [65]. Then the system state, represented by the number of users in progress at any time, is a semi-Markov decision process.

A semi-Markov decision process can be employed to model the cellular system [21], and the CAC problem can be formulated as a SMDP by indicating the following components: state, decision epoch, action, dynamic statistics and policy. The SMDP components are presented in the following and summarized in Table 2.1. The detailed SMDP formulation can be found in [81].

State space, decision epoch and action space

In the admission problems, the discrete-value (finite) state at time t can be expressed to include the number of accepted users,

$$\mathbf{s}(t) = [n_q^1(t), n_s^1(t), .., n_q^J(t), n_s^J(t)] \tag{2.12}$$

where $n_s^j(t)$ and $n_q^j(t)$ represent the number of simultaneously transmitting users (number of servers) and the number of users waiting in the queue of class j at time t, respectively, $j = 1, .., J$, and J is the total number of user classes. Note that there may be other parameters in the system state definition. In the above, $\mathbf{s}(t)$ denotes the system state at time t. By dropping the time index, (2.12) also defines a system state s.

The state space is defined as the set of all possible $\mathbf{s}$ which satisfies some QoS requirements. We remark that for the SMDP-based CAC policies there are two ways to guarantee the QoS requirements: one is to restrict the state space, and another is to add constraints to the optimization problem.

Decision epochs are defined as the instances when the stochastic process $\mathbf{s}(t)$ changes state. For an admission control problem, decision epoches include the time instances that arrivals and departures occur.

At each decision epoch, an action is chosen that determines how the admission control will perform at the next decision moment [24]. In general, an action can be defined as

$$\mathbf{a} = [a_1, d_1, .., a_J, d_J]$$

where a_j denotes the action for class j if an arrival occurs, $j = 1, .., J$. If $a_j = 0$, the new arrival is placed in the buffer provided that the buffer is not full, or is blocked if the buffer is full; if $a_j = 1$, the arrival is admitted which can be transmitted immediately, and the number of servers of class j is increased by one. The quantity d_j denotes the action for class j if a departure occurs. If $d_j = 0$, no users that are queued in the buffer are made active, i.e., transmitted, and the number of servers in class j is decreased by one; if $d_j = 1$, maintain the number of servers by admitting the user at the head of the buffer as an active user.

The action space can be defined as the set of all possible actions,

$$A = \{\mathbf{a} : \mathbf{a} \in \{0,1\}^{2J}\}$$

and for any $\mathbf{s} \in S$, the admissible action space $A_{\mathbf{s}}$ is defined as [24]

$$\{\mathbf{a} \in A : a_j = 0, \text{if } \mathbf{s} + (0,0,..,\underbrace{0,1},..,0,0) \notin S, \text{ and } (a_1,..,a_J) \neq (0,0,..0) \text{ if } \mathbf{s} = (0,..,0)\} \tag{2.13}$$

which ensures that after taking this action, the next transition state is still in the feasible space S.

State dynamics

The state dynamics of a SMDP are completely specified by stating the transition probabilities and the expected holding time. The transition probability, denoted by $p_{\mathbf{sy}}(\mathbf{a})$, is defined as the probability that the state at the next decision epoch is $\mathbf{y}$ if action $\mathbf{a}$ is selected at the current state $\mathbf{s}$. The expected holding time, denoted by $\tau_{\mathbf{s}}(\mathbf{a})$, is the expected time until the next decision epoch after action $\mathbf{a}$ is chosen in the present state $\mathbf{s}$ [24].

Policy

For any given state $\mathbf{s} \in S$, an action $\mathbf{a}$, which decides if a new call at the next decision epoch can be accepted (transmitted), is selected according to a specified policy R. A stationary policy R is a function that maps the state space into the admissible action space, where the class of admission policies can be defined as [99]

$$\mathbf{R} = \{R : S \rightarrow A | R_{\mathbf{s}} \in A_{\mathbf{s}}, \forall \mathbf{s} \in S\}.$$

Different policies incur different costs. The average cost criterion employed in this thesis for a given policy R and initial state $\mathbf{s}_0$ is given as follows:

$$J_R(\mathbf{s}_0) = \lim_{t\to\infty} \frac{1}{T} E\{\int_0^T c(\mathbf{s}(t), \mathbf{a}(t))dt\}$$

where $c(\mathbf{s}(t), \mathbf{a}(t))$ can be interpreted as the expected cost until the next decision epoch [24].

An optimal policy R^* should minimize or maximize the above average cost function $J_R(\mathbf{s}_0)$ for any initial state $\mathbf{s}_0$ subject to QoS constraints, which can be obtained by analyzing and solving the above SMDP. There are several approaches to solve SMDP, such as value iteration, policy iteration and linear programming.

2.7 Data-Link Layer Analysis: Automatic Retransmission Request

Due to multi-path fading and multiple access interference, a wireless channel has strong error prone characteristics. An efficient way to mitigate transmission error is to apply automatic retransmission request (ARQ) protocol. An ARQ scheme requests retransmissions for those packets received in error. Since retransmissions are activated only when necessary, ARQ is quite effective in improving system throughput relative to using only forward error coding (FEC) at the physical layer [54].

ARQ protocols can be categorized as stop-and-wait ARQ, Go-back-n ARQ, selective repeat ARQ and ARPanet ARQ, etc [9]. Since only finite delays and buffer sizes can be afforded in practice, the maximum number of ARQ retransmissions, denoted by L_{max}, has to be bounded, which can be specified by dividing the maximum allowable system delay over the round trip delay required for each retransmission [54]. To minimize delays and buffer sizes in practice, in this thesis, we employ a truncated ARQ protocol which has been

Table 2.1. Components of a SMDP.

SMDP components	Definition
State	State is the parameter which represents the characters of the system. In a CAC problem, the state is often represented by the number of accepted users. A state is feasible if some specific QoS requirements can be satisfied under this state.
State space	State space is the set of all feasible states.
Decision epochs	At each decision epoch, the network makes a decision that may occur in the time interval $(t_k, t_{k+1}]$.
Action	Action is a decision which indicates if a user can be accepted or not. An action is admissible if this action ensures that, after taking this action, the next transition state is still in the state space.
Admissible action space	Admissible action space is the set of admissible actions.
Expected holding time	Expected holding time is the expected time until the next decision epoch after an action is chosen in the present state.
Transition probability	Transition probability is the probability that the state transits from one state to another for a specific action at the current state.
Policy	Policy is a mapping rule from the state space to the action space, which is chosen according to some criterion.

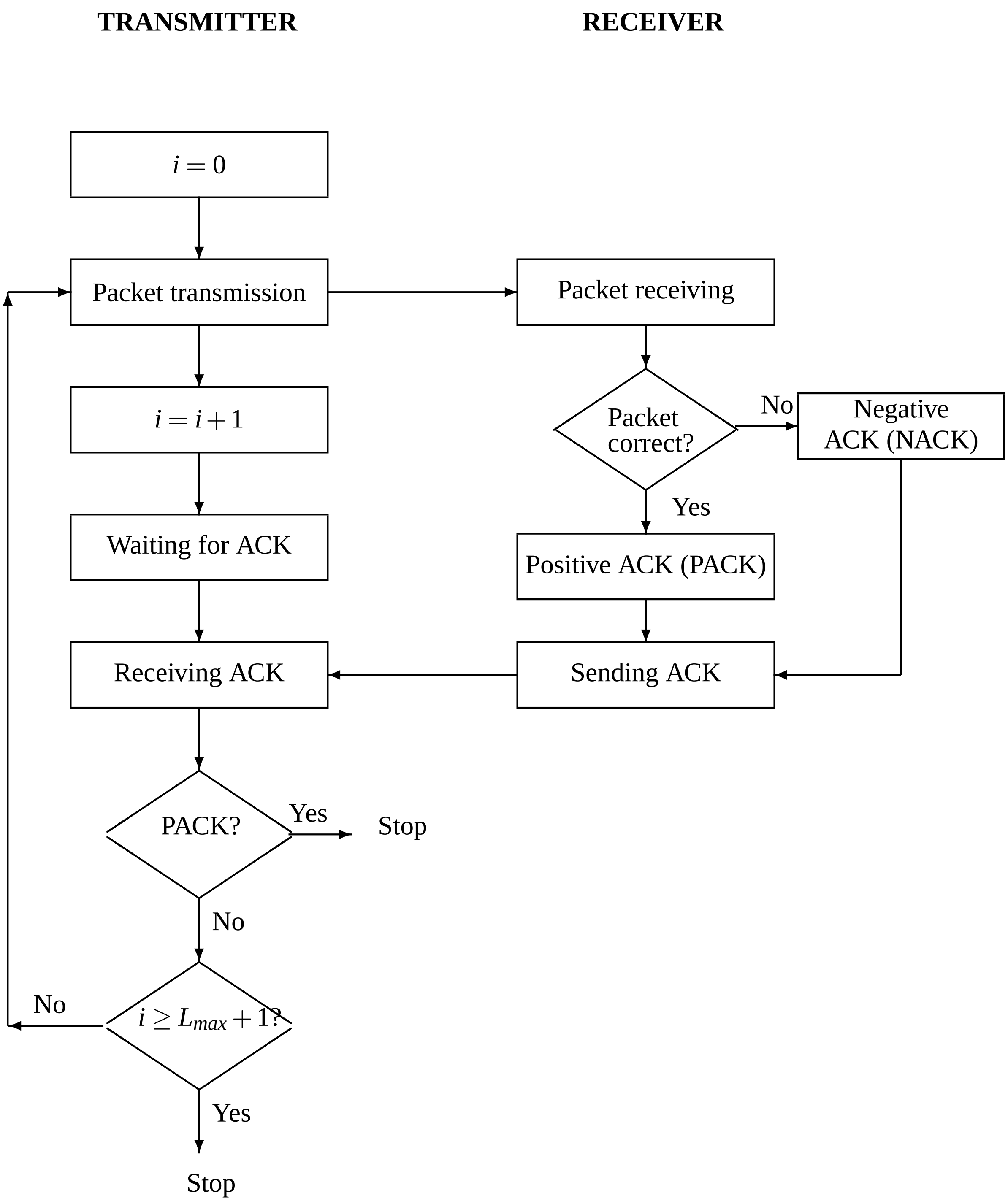

Figure 2.3. Truncated ARQ schemes.

widely adopted to limit the maximum number of retransmissions [54]. For a truncated ARQ scheme, a packet transmits until it is correctly received or the maximum number of retransmissions is reached. The procedure for a truncated ARQ scheme is presented in Figure 2.3.

We define two kinds of packet-error-rates (PERs): overall PER and instantaneous PER. Overall PER, denoted by $PER^j_{overall}$, is defined as the probability that a class j packet is incorrectly received after its maximum number of retransmissions is reached, i.e., an error occurs in each of the L_j+1 transmission rounds, where L_j denotes the maximum number of retransmissions. Instantaneous PER, denoted as $PER^j_{in}(l)$, is defined as the probability that an error occurs in a single transmission round l of a class j packet.

Under the assumption that each retransmission round is independent from the others, the achieved overall PER can be expressed as [54]

$$\begin{aligned} PER^j_{overall} &= \prod_{l=1}^{L_j+1} PER^j_{in}(l) \\ &\le \rho_j \end{aligned} \tag{2.14}$$

where ρ_j denotes the target PER for class j.

To ensure the above inequality, we require

$$PER^j_{in}(l) \le (\rho_j)^{\frac{1}{L_j+1}}. \tag{2.15}$$

Obviously, with a given instantaneous PER, the overall PER can be reduced when the number of retransmissions is increased. However, with an increased number of retransmissions, delay is also increased due to retransmissions. In general, with an average instantaneous PER, $p_{e,j}$, the average number of transmissions for a class j packet, denoted by N^j_{ARQ}, can be approximated by

$$N^j_{ARQ} = (1-p_{e,j}) + 2p_{e,j}(1-p_{e,j}) + \ldots + L_j(p_{e,j})^{L_j-1}(1-p_{e,j}) + (L_j+1)(p_{e,j})^{L_j}$$

$$= \quad 1 + p_{e,j} + (p_{e,j})^2 + ... + (p_{e,j})^{L_j - 1} + (p_{e,j})^{L_j}. \tag{2.16}$$

For a given instantaneous PER, the number of retransmissions, L_j, represents the trade-off between the overall PER and delay.

In the above, we have discussed a standard ARQ protocol, i.e., if a packet cannot be received correctly, it is discarded and retransmitted [69]. Recently, hybrid ARQ schemes are attracting a lot of attention [43], in which the receiver combines all transmissions of a packet to improve the likelihood of decoding success [43]. Hybrid ARQ schemes, e.g., [7] [16] [53] [82], are capable of improving the spectral efficiency. However, these schemes may incur high complexity when a cross-layer design is of interest. In this thesis, a standard ARQ scheme is employed for simplicity, while the extension to hybrid ARQ schemes is left for future work.

2.8 Summary

In this chapter, we briefly reviewed some related background, in which multiple access, CDMA, base-station beamforming, layered architecture, call admission control, semi-Markov decision process and automatic retransmission request are introduced.

Chapter 3

Maximum-Throughput Optimal Call Admission Control

3.1 Introduction

Current wireless systems, such as those based on the Universal Mobile Telecommunications System (UMTS), aim to provide high data-rate multimedia services with guaranteed quality-of-service (QoS) constraints in physical and network layers. This requires a cross-layer design of the call admission control (CAC). Previous research on cross-layer CAC focuses on single-antenna systems, which does not capitalize on the significant benefits provided by multiple antenna systems. In this chapter, we investigate the admission control problem for a code-division-multiple-access (CDMA) beamforming system.

For multiple antenna systems, the spatial channel response, parameterized by the angle-of-arrival (AoA) information, is employed at the receiver to suppress interference. The resulting signal-to-interference ratio (SIR) is a random process determined by the realizations of AoAs. The large fluctuations in this spatially filtered SIR can lead to a significant outage probability in the physical layer, defined as the probability that the target SIR cannot be satisfied. Existing methods for cross-layer admission control in the current literature treat the SIR as quasi-static which do not work well for multiple antenna systems. Therefore, designing an optimal CAC policy for multiple antenna systems can be a very challenging problem since the outage probability must be controlled jointly with the network layer operation.

In this chapter, we develop an exact outage probability in the presence of both voice activity and multiple antennas. Based on this exact outage probability, an optimal CAC policy is proposed by formulating a constrained semi-Markov decision process. The proposed CAC policy can maintain arbitrary outage probability constraints as well as other QoS requirements, while simultaneously optimizing the overall system throughput. To the best of our knowledge, the CAC design which maximizes the overall system throughput across different layers has not been addressed in the literature.

The rest of this chapter is organized as follows. The signal model and problem formulation are presented in Sections 3.2 and 3.3, respectively. Section 3.4 discusses the physical layer performance and provides an analytical expression for outage probability. Optimal CAC policies for single-class and multiple-class systems are proposed in Sections 3.5 and 3.6, respectively. Numerical results are presented in Section 3.7.

3.2 Signal Model

In CDMA systems, the uplink is interference limited while the downlink is power limited. Usually, the admission requests for uplink and downlink directions are asymmetric and are treated independently [50]. The model considered in this thesis is for the uplink only. However, with an appropriate physical layer model for power allocation, the methodology can be extended to the downlink CAC problem.

3.2.1 Traffic model

We consider a single-cell CDMA beamforming system which supports J classes of users, characterized by different data rates and QoS requirements. As mentioned in Section 2.5, in this thesis, the term *user* is not distinguished from the term *call*.

Without loss of generality, we assume class 1 is voice traffic. Voice activity indicators $\kappa_i(t)$ can be one or zero corresponding to an active or inactive status. It is assumed that the voice activity indicators have independent and identical distributions, and the average time percentage that a user is at state ON is defined as the voice activity factor, denoted by p_v.

The traffic model is shown in Figure 3.1. In the admission control process considered here, the mobile stations send requests for services to the base station. The base station queues these requests locally and decides which requests can be accepted. Whenever an incoming call arrives, the CAC policy implemented at the BS is employed to decide if this call can be accepted, stored in the buffer, or blocked if the buffer is full. Each class of users shares a common buffer with size B_j for class j. The aggregate arrival process for all user calls is modeled by a Poisson process with arrival rate λ_j for each class j, where $j = 1, .., J$. The number of simultaneously accepted users in class j is denoted by n_s^j, which is a random variable. We assume that the time duration of a call follows an exponential distribution with mean duration $\frac{1}{\mu_j}$, and the initial angle-of-arrivals (AoAs) of mobile users follow a uniform distribution within the service area.

From the above traffic, the total arrival rate is expressed as $\sum_{j=1}^{J} \lambda_j$, and the total departure rate is expressed as $\sum_{j=1}^{J} n_s^j \mu_j$. The overall traffic intensity, denoted by r_t, can then be represented by

$$r_t = \frac{\sum_{j=1}^{J} \lambda_j}{\sum_{j=1}^{J} n_s^j \mu_j}$$

which depends on the number of simultaneously accepted calls in the system.

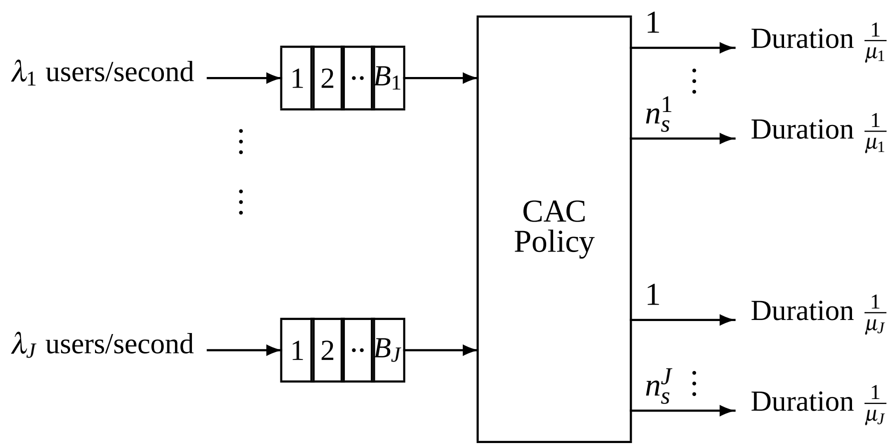

Figure 3.1. Signal model in the network layer.

We remark that if the above assumptions on memoryless Poisson and exponential distributions do not hold, the CAC problem formulated in this chapter is a generalized semi-Markov process (GSMP), instead of a semi-Markov decision process (SMDP). While an optimal solution for this GSMP problem is hard to obtain, the linear programming approach discussed in this chapter provides a sub-optimal solution to an otherwise very complicated problem [73].

For a circuit-switched network considered in this chapter, the network layer QoS is just the call-level QoS, which can be characterized by blocking probability and connection delay. From Little's Theorem [9], connection delay can be equivalently represented by the average number of users waiting in the queue.

3.2.2 Signal model at the physical layer

We consider a CDMA beamforming system which has M antennas at the BS and a single antenna for each user. We assume that there are K accepted users in the system, communicating over a channel with slow fading.

To focus on the CAC design across physical and upper layers, in the following, we neglect the effects due to multi-path. However, the proposed schemes in this paper can be extended straightforwardly to the case where multi-path exists, provided that the multi-path delay profile information can be obtained.

As shown in Chapter 2, with a temporal matched-filter receiver, the received signal-to-interference ratio (SIR) for a desired user k can be written as

$$SIR_k = \frac{W}{R_k} \frac{p_k \phi_{kk}^2}{\sum_{i \neq k} p_i \phi_{ik}^2 + \eta_0 W} \tag{3.1}$$

where W and R_k denote the bandwidth and data rate for user k, respectively, and the ratio $\frac{W}{R_k}$ represents the processing gain; p_i denotes the received power for user i, and η_0 denotes the one-sided power spectral density of background additive white Gaussian noise (AWGN); the parameter ϕ_{ik}^2 captures the effects of beamforming, defined as $\phi_{ik}^2 = \left|\mathbf{w}_k^H \mathbf{a}_i\right|^2$, in which $\mathbf{w}_k$ and $\mathbf{a}_i$ denote the weighting factor for a desired user k and the array response vector for user i, respectively, and $(.)^H$ denotes Hermitian. In the following, we consider a spatially matched filter receiver, i.e., $\mathbf{w}_k = \mathbf{a}_k$.

The communication reliability in the physical layer can be represented by a target bit-error-rate (BER) or packet-error-rate (PER), which can be equivalently mapped to a target SIR requirement. In a wireless communication network, we must allow for outage, defined as the probability that a target SIR, or equivalently, a target bit-error-rate (BER) or target packet-error-rate (PER), cannot be satisfied. The QoS measurement in the physical layer is represented by a target outage probability.

In this chapter, we consider two types of outage probability constraints: worst-state-outage-probability (WSOP) constraint, denoted by ρ_w, and average-outage-probability (AOP) constraint, denoted by ρ_{av}. While WSOP constraint is very conservative which ensures that at any time instance and at any system state this outage probability constraint cannot be violated, the AOP ensures a long-run average outage probability constraint, which may be more practical.

3.3 Problem Formulation

The overall system throughput, defined as the number of QoS guaranteed calls per second, can be evaluated by [54]

$$\text{Throughput} = \sum_j (1 - P_b^j)(1 - P_{out}^{av})\lambda_j \tag{3.2}$$

where P_b^j and P_{out}^{av} denote the blocking probability for class j and average-outage-probability, respectively.

In this chapter, we aim to investigate the cross-layer CAC problem which incorporates the benefits provided by multiple antennas and voice activity. The objective is to maximize the overall system throughput, while simultaneously guaranteeing QoS requirements in terms of average-outage-probability, worst-state-outage-probability, blocking probability and connection delay.

The above optimal CAC problem is a constrained optimization problem, which can be solved by formulating a semi-Markov decision process [81].

3.4 Physical Layer Investigation: Outage Probability

This section investigates physical layer performance. We first provide an exact power control feasibility condition and an exact system state which ensure zero-outage. Then a simplified system state is introduced, which leads to a non-zero outage probability.

3.4.1 Exact power control feasibility condition and system state

With n_s^j users for class j, the total number of users in the system, denoted by K, can be expressed as $\sum_{j=1}^{J} n_s^j$. Letting the SIR for an arbitrary user k, given in (3.1), achieve its target value, we have

$$\gamma_k = \frac{W}{R_k} \frac{p_k \phi_{kk}^2}{\sum_{i \neq k} p_i \phi_{ik}^2 + \eta_0 W} \tag{3.3}$$

in which γ_k denotes the target SIR for user k, where $k = 1, 2, .., K$.

By grouping the above K equations, we have the following matrix equation

$$[I_K - QF]\mathbf{p} = Q\mathbf{u} \tag{3.4}$$

where I_K is a $K-$dimensional identity matrix, power vector $\mathbf{p} = [p_1, .., p_K]^t$, $\mathbf{u} = \eta_0 W[1, .., 1]^t$, $(.)^t$ denotes transpose, Q is a K-dimensional diagonal matrix with the i^{th} non-zero element as $\frac{\frac{\gamma_i R_i}{W}}{1+\frac{\gamma_i R_i}{W}}$, and F is a K by K matrix in which the element at the i^{th} row and the j^{th} column can be expressed as $F_{ij} = \frac{\phi_{ij}^2}{\phi_{ii}^2}$.

To ensure a positive solution for power vector $\mathbf{p_s}$, we require the following feasibility condition [76],

$$\upsilon(QF) < 1 \tag{3.5}$$

where $\upsilon(.)$ denotes the maximum eigenvalue, which is real-valued since the matrices are symmetric.

We remark that (3.5) represents a sufficient and necessary condition in which the target SIRs of all users can be satisfied with zero outage probability.

The exact system state, introduced in Chapter 2, denoted by $\mathbf{s}_{exact}$, which characterizes the physical layer performance, is defined as

$$\mathbf{s}_{exact} = [n_q^1, n_s^1, A_1, .., n_q^j, n_s^J, A_J] \tag{3.6}$$

where n_s^j and n_q^j denote the numbers of users in the system and in the queue, respectively, and A_j denotes a 1-by-n_s^j matrix which represents the AoA realizations for n_s^j users in the system.

Under perfect power control, the outage probability for an exact system state $\mathbf{s}_{exact}$ can be zero or one depending on whether (3.5) holds. By appropriately choosing the system states, zero-outage probability can be guaranteed.

3.4.2 Outage probability for a simplified system state

A system state is feasible if and only if the condition in (3.5) can be satisfied. Due to arbitrarily located users, the AoA realizations in the above system state can take any real value within, say, $[0, 2\pi)$, and the state space, formed from the set of all feasible system states, therefore, has an infinite size. A SMDP approach, which requires a finite-size state space, however, cannot be applied.

By allowing outage in the physical layer, from here on, we simplify the system state in (3.6) to

$$\mathbf{s} = [n_q^1, n_s^1, .., n_q^j, n_s^J]. \tag{3.7}$$

dropping the dependence on the sets of AoAs. This leads to a finite-size feasible space. In the following, system state refers to the above simplified system state (3.7).

With this simplified system state, one cannot tell if (3.5) holds, since no AoA realizations are specified. Instead, for this simplified state, a non-zero outage probability is introduced, which can be expressed as

$$P_{out}(\mathbf{s}) \;=\; \text{Prob}\{\upsilon(QF) \geq 1\}. \tag{3.8}$$

where Prob$\{A\}$ denotes the probability of event A.

With voice activity, (3.8) is modified to

$$P_{out}(\mathbf{s}) = \sum_{m=0}^{n_s^1} p(m)\text{Prob}\{\upsilon(Q_m F_m) \geq 1\} \tag{3.9}$$

where Q_m and F_m are the parameters Q and F with m active voice users in state $\mathbf{s}$, $m = 1,..,n_s^1$, and $p(m)$ denotes the probability that m out of n_s^1 users are active, expressed as

$$p(m) = \binom{n_s^1}{m} p_v^m (1-p_v)^{n_s^1 - m}. \tag{3.10}$$

As shown in Appendix B, for a single antenna system, the outage probability in (3.9) can be written in a simplified closed-form as follows

$$P_{out}(s) \;=\; \begin{cases} 1 & \text{if } n_s^2 > D_2 \cup n_s^3 > D_2 \cup \ldots \cup n_s^J > D_J \\ 0 & \text{if } n_s^1 < D_1 \left(1 - \sum_{j=2}^{J} \frac{n_s^j}{D_j}\right) \\ 1 - I_{1-p_v}\left(n_s^1 - A, A+1\right) & \text{otherwise.} \end{cases} \tag{3.11}$$

where $\cup$ denotes union, $D_j = 1 + \frac{W}{\gamma_j R_j}$, $\llcorner a \lrcorner$ denotes the maximum integer less than a, $A = \llcorner D_1 \left(1 - \sum_{j=2}^{J} \frac{n_s^j}{D_j}\right) \lrcorner$, and $I_p(c,b)$ represents a regularized incomplete beta function with parameters p, c, b.

For a multiple antenna system, the outage probability in (3.9) is very complicated, which can be evaluated numerically as

$$P_{out}(\mathbf{s}) \;=\; \sum_{m=0}^{n_s^1} p(m) \int_0^{2\pi} .. \int_0^{2\pi} f\left(\theta_1^0, .., \theta_N^0\right) \text{Prob}\left\{\upsilon(Q_m F_m) \geq 1 | \theta_1^0 .. \theta_N^0\right\} d\theta_1^0 .. d\theta_N^0 \tag{3.12}$$

in which θ_i^0 is the initial angle of arrival (AoA) for user i, where $i = 1,..,N$ and N denotes the total number of active users, which can be obtained by $N = m + \sum_{j=2}^{J} n_s^j$; $f\left(\theta_1^0,..,\theta_N^0\right)$ denotes the joint probability density function (PDF) of $\theta_1^0,..,\theta_N^0$, $f(\theta_i^0)$ denotes the PDF of θ_i^0, and $\mathrm{Prob}(A|B)$ denotes conditional probability.

The above conditional probability in (3.12) can be obtained by time averaging via

$$\mathrm{Prob}\{\upsilon(Q_m F_m) \geq 1|\theta_1^0,..,\theta_N^0\} = \lim_{T\to\infty} \frac{1}{T} \int_0^T 1 - \delta\left(1 - \upsilon\left(Q_m F_m(t)\right)|\theta_1^0,..,\theta_N^0\right) dt \quad (3.13)$$

where $F_m(t)$ is F_m at time t, and

$$\delta(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.14)$$

Lacking a closed form analytical expression, the exact outage probability for multiple antenna systems, given in (3.12), can be very hard to evaluate. In a practical system, the outage probability can be obtained by time averaging via

$$P_{out}(\mathbf{s}) = \lim_{N_r\to\infty} \frac{1}{N_r} \sum_{i=1}^{N_r} \left\{1 - \delta(1 - \upsilon(Q^i F^i))\right\} \quad (3.15)$$

where N_r is the number of independent AoA realizations, and Q^i and F^i denote Q and F for the i^{th} AoA realization, respectively.

Based on this state outage probability, the worst-state outage probability, denoted by P_{out}^w, and the average outage probability, denoted by P_{out}^{av}, can be expressed, respectively, as

$$P_{out}^w = \max_{\mathbf{s}\in S} P_{out}(\mathbf{s}) \quad (3.16)$$

$$P_{out}^{av} = \sum_{\mathbf{s}\in S} P_{\mathbf{s}} P_{out}(\mathbf{s}) \quad (3.17)$$

where $P_{\mathbf{s}}$ denotes the steady-state probability that the system is at state $\mathbf{s}$ and S represents the set of all feasible system states.

3.5 Optimal CAC Policy for a Single-Class System

Before discussing the optimal CAC problem for multi-class systems, we first derive the optimal admission policy for a single-class system, in which all the users have the same QoS requirements and follow the same transmission rate. Although this single-class assumption seems impractical in current and future multimedia wireless networks, it provides some insight for the more complicated multiple service class wireless networks.

For single-class traffic, the optimal CAC policy is simplified to a threshold-based policy. With a threshold K_{th}, whenever the number of accepted users reaches this threshold, the incoming user would be blocked. Therefore, the problem to derive the optimal CAC policy for a single-class system is to find the optimal threshold, denoted by K_{th}^{opt}, which maximizes the throughput subject to the WSOP, AOP and blocking probability constraints. For simplicity, no buffer is employed and the connection delay constraint is ignored here.

For single-class system, the system state, which is a scalar, denoted by s, is just the number of the total accepted users, and the state space can be expressed as

$$S = \{s; s \leq K_{th}\}. \tag{3.18}$$

In terms of the physical layer QoS, the outage probability for any state s can be obtained by (3.9). As presented in (3.16) and (3.17), the average and worst-state outage probabilities for a given threshold K_{th} can be derived as

$$\begin{aligned} P_{out}^{av} &= \sum_{0 \leq s \leq K_{th}} P_s P_{out}(s) \\ P_{out}^{w} &= P_{out}(K_{th}) \end{aligned} \tag{3.19}$$

where P_s denotes the steady-state probability, given by [12]

$$P_s = \begin{cases} \frac{\frac{(B_v)^s}{s!}}{\sum_{i=0}^{K_{th}} \frac{(B_v)^i}{i!}} & \text{when} \quad 0 \leq s \leq K_{th} \\ 0 & \text{otherwise.} \end{cases} \tag{3.20}$$

where $B_v = \frac{\lambda}{\mu}$, in which λ and μ represent the arrival and departure rates for voice users, respectively. We note that B_v represents the traffic load, and can be interpreted as the arrival-to-departure-rate ratio.

With Poisson arrival and departure, using queuing analysis, the blocking probability for single-class traffic with threshold K_{th} can be derived as [12]

$$P_b = \frac{B_v^{K_{th}}/K_{th}!}{\sum_{j=0}^{K_{th}} B_v^j/j!}. \tag{3.21}$$

As shown in (3.2), the overall system throughput with threshold K_{th} can be expressed as

$$\text{Throughput} = \sum_{s=0}^{K_{th}} P(s)(1-P_b)(1-P_{out}(s)). \tag{3.22}$$

Inserting Equation (3.21) into (3.22), we obtain the analytical system throughput for a given threshold as follows

$$\text{Throughput} = \sum_{s=0}^{K_{th}} P(s)\Big(1-\frac{B_v^{K_{th}}/K_{th}!}{\sum_{j=0}^{K_{th}} B_v^j/j!}\Big)(1-P_{out}(s)) \tag{3.23}$$

where $P_{out}(s)$ is given in (3.9).

Therefore, the optimal threshold can be derived by solving the following optimization problem

$$K^{opt} = \arg\max_{K_{th}\in\mathbf{K_{th}}} \text{Throughput} \tag{3.24}$$

where Throughput is given in (3.23), and $\mathbf{K_{th}}$ is given by

$$\mathbf{K_{th}} = \{K_{th}, \text{ where } \sum_{0\le s\le K_{th}} P_s P_{out}(s) \le \rho_{av}, \\ \max_{0\le s\le K_{th}} P_{out}(s) \le \rho_w \text{ and } \frac{B_v^{K_{th}}/K_{th}!}{\sum_{j=0}^{K_{th}} B_v^j/j!} \le \Psi\}. \tag{3.25}$$

With this optimal threshold K^{opt}, the system throughput can be maximized.

3.6 Optimal Call Admission Control Policy for Multiple-Class Networks

In the above, we have shown that the optimal CAC policy for single-class systems without buffering can be simplified to a threshold-based policy. For multiple-class traffic with buffering, due to the interaction between the multiple class traffic and the lack of the analytical results on blocking probability and steady-state probability, the threshold-based CAC policy cannot be applied.

To derive the optimal CAC policy for multiple-class networks employing buffering, we need to solve a constrained optimization problem as presented in Section 3.3. This constrained optimization problem can be achieved by formulating the CAC problem as a semi-Markov-decision-process (SMDP) if the Markovian property holds, and then solved by linear programming (LP) [24].

3.6.1 SMDP components

We track the arrival and departure processes whose instants are assumed to be Poisson distributed and whose durations are independent and exponentially distributed [65]. In view of these assumptions, the system state, describing the total number of users in progress at any time, has the Markovian property that the future behavior of the process depends only on the present state and is independent of the past history [81]. In this sense, the CAC problem can be formulated as a SMDP. A SMDP includes the following components: state space, decision epoch, action space, dynamic statistics and policy [99]. By considering the signal model and optimization problem discussed in Section 3.3, the components of our formulated SMDP can be obtained as follows.

State space

The state space, denoted by S, includes all the possible state vectors $\mathbf{s}$. The state space together with the SMDP constraints should ensure the QoS requirements. We express the QoS requirements in terms of blocking probability, connection delay, AOP and WSOP. Among the above four QoS requirements, only the WSOP requirement can be guaranteed by restricting the state space, i.e.,

$$S = \left\{\mathbf{s} = [n_q^1, n_s^1, .., n_q^J, n_s^J], \text{ where } P_{out}(\mathbf{s}) \leq \rho_w \text{ and } n_q^j \leq B_j\right\} \tag{3.26}$$

where $P_{out}(\mathbf{s})$ is given in (3.9), ρ_w denotes the WSOP constraint and B_j is the buffer size for class j.

The formulation of the above state space can be summarized as follows:

- Compute the maximum number of accepted users for each class, denoted by M_j^{max}. The search procedure for M_j^{max} is presented in Figure 3.2.
- An enlarged state space, denoted by $\bar{S}$, can be formulated as

$$\bar{S} = \left\{\mathbf{s} = [n_q^1, n_s^1, .., n_q^J, n_s^J] : n_{s,j} \leq M_j^{max}, n_q^j \leq B_j, \text{ for } j = 1, .., J\right\};$$

- The above $\bar{S}$ can be truncated to the desired state space S as follows:
  - Initialize $S = \{\}$;
  - Evaluate $P_{out}(s)$ for $\mathbf{s} \in \bar{S}$ according to (3.15);
  - If $P_{out}(s) \leq \rho_w$, then $S = S + \{s\}$.
- We remark that in the above step, it is unnecessary to evaluate each system state in $\bar{S}$, since if $\mathbf{s} \in S$, then all $\mathbf{s}' \in \bar{S}$ such that $\mathbf{s}' \leq \mathbf{s}$ are also in S. Similarly, if $\mathbf{s}$ is not in S, then all $\mathbf{s}' \in \bar{S}$ such that $\mathbf{s}' \geq \mathbf{s}$ are also not in S.

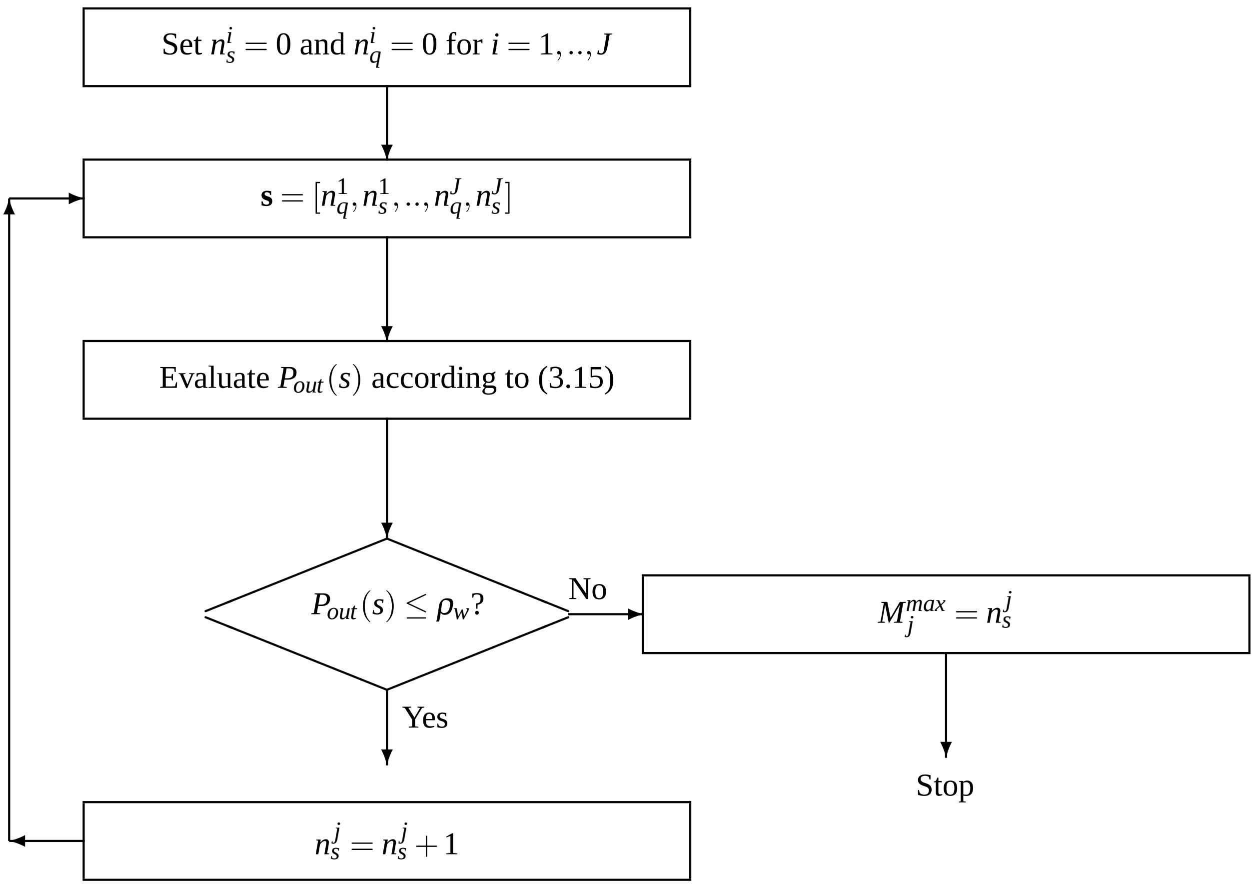

Figure 3.2. Search procedure for M_j^{max}.

For a system without WSOP constraint, i.e., $\rho_w = 1$, the above state space would have a size of infinity. To formulate a finite-size state space, as shown in [99], we can limit the number of users by a large number G,

$$S = \{\mathbf{s} = [n_q^1, n_s^1, .., n_q^J, n_s^J], \sum_j n_s^j < G \text{ and } n_q^j \le B_j\}$$

where G can be decided by the system, which represents a tradeoff between complexity and performance.

Let $\mathbf{s}(t)$ denote the system state at time t, where $\mathbf{s}(t) \in S$. Since the arrivals and departures are random, $\{\mathbf{s}(t)\}_{t \in IR_+}$ is a finite-size stochastic process [99].

Decision epoch and Action space

Decision epochs are chosen to be the set of all arrival and departure instants. At each decision epoch, t_k, $k = 1, 2, ..$, the network takes an action for each possible user arrival or departure that may occur in the time interval $(t_k, t_{k+1}]$. The action, denoted by $\mathbf{a}$, indicates if the incoming user or the user waiting in the queue can be transmitted. An action $\mathbf{a}$ can be expressed as

$$\mathbf{a} = [a_1, d_1 .., a_J, d_J] \tag{3.27}$$

where a_j can be 1 or 0, corresponding to decisions of acceptation or rejection, respectively, and d_j can be one or zero, corresponding to decisions of making a user in the queue active (i.e., transmitted), or maintaining the queue unchanged, respectively.

For any $\mathbf{s} \in S$, the admissible action space $A_{\mathbf{s}}$ is defined as the set of all possible actions, which ensures that after taking one action in the admissible action space, the next transition state is still in state space S,

$$\begin{aligned} A_{\mathbf{s}} = \{\mathbf{a} \in A : a_j = 0, \text{ if } \mathbf{s} + (0, .., \underbrace{0, 1}, ..,) \notin S \\ \text{and } (a_1, .., a_J) \neq (0, .., 0) \text{ if } \mathbf{s} = (0, .., 0)\}. \end{aligned} \tag{3.28}$$

Dynamic statistics

Dynamic statistics can be characterized by expected holding time and transition probability. The expected holding time, denoted by $\tau_{\mathbf{s}}(\mathbf{a})$, is the expected time until the next decision epoch after action $\mathbf{a}$ is chosen in the present state $\mathbf{s}$. As discussed in [24], $\tau_{\mathbf{s}}(\mathbf{a})$ can be expressed by the inverse of the cumulative event rate, while the cumulative event rate is just the sum of the rates for all arrival and departure processes, i.e.,

$$\tau_{\mathbf{s}}(\mathbf{a}) = \left(\sum_{j=1}^{J} \lambda_j a_j + \sum_{j=1}^{J} \lambda_j (1-a_j)\delta(B_j - n_q^j) + \sum_{j=1}^{J} \mu_j n_s^j \right)^{-1} \tag{3.29}$$

where $\delta(x)$ is defined in (3.14).

The transition probability, denoted by $p_{\mathbf{sy}}(\mathbf{a})$, is the probability that the state at the next decision epoch is $\mathbf{y}$ if action $\mathbf{a}$ is selected at the current state $\mathbf{s}$. The decomposition property of a poisson process can be employed to derive the transition probability, which indicates that an event of certain type occurs with a probability equal to the ratio between the rate of that particular type of event and the total cumulative event rate, i.e., $\frac{1}{\tau_{\mathbf{s}}(\mathbf{a})}$ [24].

By using the decomposition property, the transition probability can be represented by

$$p_{\mathbf{sy}}(\mathbf{a}) = \begin{cases} \lambda_j a_j \tau_{\mathbf{s}}(\mathbf{a}) & \text{if } \mathbf{y} = \mathbf{s} + \mathbf{e}_s^j \\ \lambda_j (1-a_j)\delta(B_j - n_q^j)\tau_{\mathbf{s}}(\mathbf{a}) & \text{if } \mathbf{y} = \mathbf{s} + \mathbf{e}_q^j \\ \mu_j n_s^j (1-d_j)\tau_{\mathbf{s}}(\mathbf{a}) + \mu_j n_s^j d_j (1-\delta(n_q^j))\tau_{\mathbf{s}}(\mathbf{a}) & \text{if } \mathbf{y} = \mathbf{s} - \mathbf{e}_s^j \\ \mu_j n_s^j d_j \delta(n_q^j)\tau_{\mathbf{s}}(\mathbf{a}) & \text{if } \mathbf{y} = \mathbf{s} - \mathbf{e}_q^j \end{cases} \tag{3.30}$$

in which $\mathbf{e}_s^j$ represents a vector with a dimension of $2J$, which contains only zeros except for position $2j$ which contains a 1, and $\mathbf{e}_q^j$ represents a vector with a dimension of $2J$, which contains only zeros except for position $2j-1$ which contains a 1.

For each given state $\mathbf{s} \in S$, an action $\mathbf{a} \in A_{\mathbf{s}}$ is chosen according to a policy R. A policy defines a mapping rule from the state space to the action space [99], i.e.,

$$R = \{R_s : S \rightarrow A | R_{\mathbf{s}} \in A_{\mathbf{s}}, \forall \mathbf{s} \in S\}.$$

where A denotes the set of all admissible action space.

A brief description of the above SMDP components is summarized in Table 3.1, and a detailed SMDP formulation can be found in [81].

3.6.2 QoS constraints

As discussed previously, we have QoS constraints in terms of blocking probability, connection delay, average outage probability and worst-state outage probability constraints. Among the QoS requirements, WSOP constraint can be satisfied by restricting the state space, and there are blocking probability, connection delay and AOP constraints, which should be incorporated into the SMDP constraints.

In the following, we take the average cost as the performance criterion. For any policy R with an initial system state $\mathbf{s}_0$, where $\mathbf{s}_0 \in S$, the achieved long-run blocking probability for class j, where $j = 1, .., J$, can be expressed as

$$\begin{aligned} P_b^j &= \lim_{T\to\infty} \frac{1}{T} E\left\{ \int_0^T P_b(\mathbf{s}(t), \mathbf{a}(t))dt \right\} \\ &\leq \Psi_j \end{aligned} \tag{3.31}$$

where $E[.]$ denotes expectation, Ψ_j denotes the blocking probability constraint for class j, and $P_b(\mathbf{s}(t), \mathbf{a}(t))$ denotes the expected cost function in terms of blocking probability for state $\mathbf{s}(t)$ and action $\mathbf{a}(t)$, which can be obtained as

$$P_b(\mathbf{s}(t), \mathbf{a}(t)) = (1 - a_j(t))(1 - \delta(B_j - n_q^j(t))). \tag{3.32}$$

Similarly, the achieved long-run connection delay can be expressed as

$$\begin{aligned} \text{Delay}_j &= \lim_{T\to\infty} \frac{1}{T} E\left\{ \int_0^T n_q^j(t)dt \right\} \\ &\leq D_j, \; j = 1, .., J \end{aligned} \tag{3.33}$$

where D_j denotes the connection delay constraint for class j.

Table 3.1. Formulating the optimal CAC problem as SMDP.

SMDP components	Notation	Expression for optimal CAC
System state	$\mathbf{s}$	$\mathbf{s} = [n_q^1, n_s^1, .., n_q^J, n_s^J]$.
State space	S	$S = \left\{ \mathbf{s},\ \text{where } P_{out}(\mathbf{s}) \leq \rho_w \text{ and } n_q^j \leq B_j \right\}$ where ρ_w denotes the WSOP constraint and B_j is the buffer size for class j
Decision epochs	t_k	The set of all arrival and departure instances.
Action	$\mathbf{a}$	$\mathbf{a} = [a_1, d_1 .., a_J, d_J]$, where $a_j = 1$ represents the decision to accept a class j user, while $a_j = 0$ represents a rejection; $d_j(t)$ can be one or zero, corresponding to decisions of making a user in the queue active.
Admissible action space	$A_\mathbf{s}$	$A_\mathbf{s} = \{\mathbf{a} : a_j = 0,\ \text{if } \mathbf{s} + \mathbf{e}_s^j \notin S,\ \text{and}$ $(a_1, .., a_J) \neq (0, .., 0) \text{ if } \mathbf{s} = (0, .., 0)\}$, in which $\mathbf{e}_s^j$ represents a $2J$- dimensional vector, which contains only zeros except for position $2j$ which contains a 1.
Expected holding time	$\tau_\mathbf{s}(\mathbf{a})$	$\tau_\mathbf{s}(\mathbf{a}) = \left(\sum_{j=1}^J \lambda_j a_j + \sum_{j=1}^J \lambda_j (1 - a_j) \delta(B_j - n_q^j) + \sum_{j=1}^J \mu_j n_s^j \right)^{-1}$.
Transition probability	$p_{\mathbf{sy}}(\mathbf{a})$	Given in (3.30).
Policy	R	$R = \{R_s : S \rightarrow A \mid R_\mathbf{s} \in A_\mathbf{s}, \forall \mathbf{s} \in S\}$ where A denotes the set of all admissible action space.
Constraints		$P_{out}^{av} \leq \rho_{av}$, $\text{Delay}_j \leq D_j$, and $P_b^j \leq \Psi_j$.

The achieved long-run average-outage-probability (AOP) can be represented by

$$\begin{aligned} P_{out}^{av} &= \lim_{T\to\infty} \frac{1}{T} E\left\{ \int_0^T P_{out}(\mathbf{s}(t))dt \right\} \\ &\leq \rho_{av} \end{aligned} \tag{3.34}$$

where ρ_{av} denotes the AOP constraint.

3.6.3 Deriving an optimal policy by solving the SMDP

The policy can be chosen according to a certain performance criterion, such as minimizing-blocking-probability or maximizing-throughput. Here we aim to find an optimal policy R^* which maximizes the throughput for any initial system state.

The overall system throughput, defined as the number of correctly received calls per unit time, can be expressed by

$$\begin{aligned} \text{Throughput} &= \sum_{j=1}^{J} \lambda_j (1 - P_b^j)(1 - P_{out}^{av}) \\ &= \lim_{t\to\infty} \frac{1}{T} E\left\{ \int_0^T \sum_{j=1}^{J} \lambda_j (1 - P_b(\mathbf{s}(t), \mathbf{a}(t)))(1 - P_{out}(\mathbf{s}(t)))dt \right\} \end{aligned} \tag{3.35}$$

where $P_b(\mathbf{s}(t), \mathbf{a}(t))$ is given in (3.32).

Under the assumption that the embedded chain is a unichain [81], which is a common assumption in the CAC problem, an optimal CAC policy exists [99]. For the problem above, in terms of the decision variables $z_{\mathbf{sa}}, \mathbf{s} \in S, \mathbf{a} \in A_{\mathbf{s}}$, the following linear programming (LP) problem [10] can be formulated as:

$$\max_{z_{\mathbf{sa}} \geq 0, \mathbf{s}, \mathbf{a}} \sum_{\mathbf{s}\in S} \sum_{\mathbf{a}\in A_{\mathbf{s}}} \sum_{j=1}^{J} \lambda_j (1 - (1 - a_j)(1 - \delta(B_j - n_q^j)))(1 - P_{out}(\mathbf{s}))\tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} \tag{3.36}$$

subject to the set of constraints

$$\sum_{\mathbf{a}\in A_{\mathbf{m}}} z_{\mathbf{ma}} - \sum_{\mathbf{s}\in S} \sum_{\mathbf{a}\in A_{\mathbf{s}}} p_{\mathbf{sm}}(\mathbf{a}) z_{\mathbf{sa}} = 0, \mathbf{m} \in S \tag{3.37}$$

$$\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} \tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} = 1 \tag{3.38}$$

$$\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} (1-a_j)(1-\delta(B_j - n_q^j))\tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} \leq \Psi_j, \; j=1,..,J \tag{3.39}$$

$$\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} n_q^j \tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} \leq D_j, \; j=1,..,J \tag{3.40}$$

$$\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} P_{out}(\mathbf{s})\tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} \leq \rho_{av}. \tag{3.41}$$

In the above LP formulation, the decision variables $z_{\mathbf{sa}}$ represent the normalized frequency of taking action $\mathbf{a}$ when the system is in state $\mathbf{s}$, i.e., $z_{\mathbf{sa}} = \frac{N_{\mathbf{sa}}}{T}$, where $N_{\mathbf{sa}}$ denotes the number of times that the action $\mathbf{a}$ is taken when the system is in state $\mathbf{s}$ over a long time period T. From the definition of $z_{\mathbf{sa}}$, it is easy to observe that $\tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}}$ represents the steady-state probability that the system is at state $\mathbf{s}$ and action $\mathbf{a}$ is chosen. The objective function in (3.36) is to maximize the system throughput, which is obtained from (3.35). The first constraint is the balance equation, and the second constraint ensures the sum of all the steady-state probabilities to be one. The latter three constraints represent the QoS requirements in terms of blocking probability, connection delay and average-outage-probability, respectively, which are obtained directly from inequalities (3.31)(3.33)(3.34).

Since the sample path constraints are included in the above linear programming approach, the optimal policy resulting from the SMDP is a randomized policy [99]: the optimal action $\mathbf{a}^* \in A_{\mathbf{s}}$ for state $\mathbf{s}$, where $A_{\mathbf{s}}$ is the admissible action space, is chosen probabilistically according to the probabilities $z_{\mathbf{sa}} / \sum_{\mathbf{a}\in A_{\mathbf{s}}} z_{\mathbf{sa}}$.

We remark that the above randomized CAC policy allows for resources to be more flexibly reserved for potential arriving traffic, and as a result, can optimize the long-run performance.

3.7 Numerical Examples

3.7.1 Simulation parameters

In the following examples, a circular antenna array with a uniformly distributed AoA is employed at the BS. The total bandwidth is $B = 3.84MHz$. A two-class system is considered in which the SIR requirements are given by $\gamma_1 = 10$ dB and $\gamma_2 = 7$ dB, and the data rates are set to $R_1 = 48$ kbps and $R_2 = 144$ kbps, respectively. The arrival and departure rates for class 1 and class 2 are denoted by $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\mu_1 = 0.25$, and $\mu_2 = 0.1375$, respectively, unless specified otherwise. For simplicity, it is assumed that there is no user mobility.

We note that compared with beamforming systems, single antenna systems encounter an infeasibility problem more easily, i.e., the QoS requirements may not be satisfied by any CAC policy. Since we aim to quantify the performance difference between single and multiple antenna systems, this infeasibility situation should be avoided. Therefore, unless specified, we employ relatively relaxed blocking probability constraints, which are set to 0.25 and 0.45 for classes 1 and 2, respectively.

In the following examples, we first derive the policy and then apply the policy to a dynamic system with random arrival and departure processes. The simulation implementation is presented in Appendix A.

3.7.2 Performance for single-voice-class systems

In this section, we analytically evaluate the performance for a single-voice-class system, in which the SIR and rate requirements are given by $\gamma = 10$ dB and $R = 48$ kbps, respectively. Voice activity factor is set to $p_v = 3/8$. The arrival rate λ is 1. In the system under investigation, no buffer is employed.

In Figure 3.3, the blocking probability, the outage probability and the system throughput are presented as a function of the threshold. From (3.17), it is known that the average outage probability depends on the traffic model. For a heavy traffic, i.e., high B_v, the achieved AOP is very close to WSOP. To illustrate this point, high B_v is chosen.

We observe that the blocking probability can be reduced dramatically by increasing the threshold, which, however, increases the outage probability. The threshold should be chosen appropriately by considering the above impacts.

From Figure 3.3, it is also found that the worst-state-outage-probability remains unchanged no matter how heavy the traffic load becomes, while the average-outage-probability strongly depends on the traffic load. For any chosen threshold, the achieved average-outage-probability is lower than worst-state-outage-probability. However, with an increased traffic load B_v, the average-outage-probability approaches the worst-state-outage-probability. When the traffic load exceeds a certain level, e.g., $B_v = 60$ in the system we investigate, the average-outage-probability can be approximated by the worst-state-outage-probability, since under this high traffic load, the system would stay in the worst-state with a probability of almost 1.

From Figure 3.3, we also observe that when the threshold exceeds a certain cutoff level, denoted as L_{cutoff}, the achieved average outage probability is not monotonically degraded by the increased threshold. Instead, for a certain traffic load, all states beyond the cutoff level have very little impact on the achieved-outage-probability. The cutoff level is decided by the traffic load. For example, when $B_v = 10$, the cutoff level is around 20, while this cutoff level is increased to 60 when $B_v = 60$.

The overall system throughput is also presented in Figure 3.3, which represents the tradeoff between physical layer and network layer performance. It is observed that there exists an optimal threshold which can achieve the maximum throughput. For example, when traffic load is 30, with outage probability constraint $\rho_w = 0.5$, $\rho_{av} = 0.1$, blocking probability constraint $\Psi = 0.1$, we have the optimal threshold $K_{th}^{opt} = 37$, which achieves a maximum throughput of 0.83.

3.7.3 Performance for two-class systems

In the above, we have studied the performance for single-class networks. Now we investigate the long-run average performance in terms of blocking probability, average outage probability and overall system throughput for two-class networks.

Consider a system in which WSOP is relatively relaxed, e.g., 0.5. This relaxed WSOP constraint is used to limit state space, and the link reliability is ensured by average-outage-probability constraint (AOP). We investigate a system with AOP constraint over $[10^{-4}, 10^{-2}]$. Figure 3.4 shows the analytical and simulated performance for a two-antenna system, in which P_b^j and P_{out}^{av} denote the achieved blocking probability and average-outage-probability, respectively. The analytical results are derived from linear programming, while the simulation results are obtained by Monte-Carlo simulation. It is observed that the simulation results are very close to the analytical results.

From Figure 3.4, we also observe that the blocking probability for class 1 is not monotonically reduced with ρ_{av}. Instead, with a relaxed ρ_{av} from $\rho_{av} = 10^{-2}$ to 5×10^{-2}, the blocking probability is increased for class 1. This can be explained by the fact that although relaxing ρ_{av} allows more users accepted in the system, to achieve a maximum throughput, the connection requests from class 1 may be blocked more frequently.

,

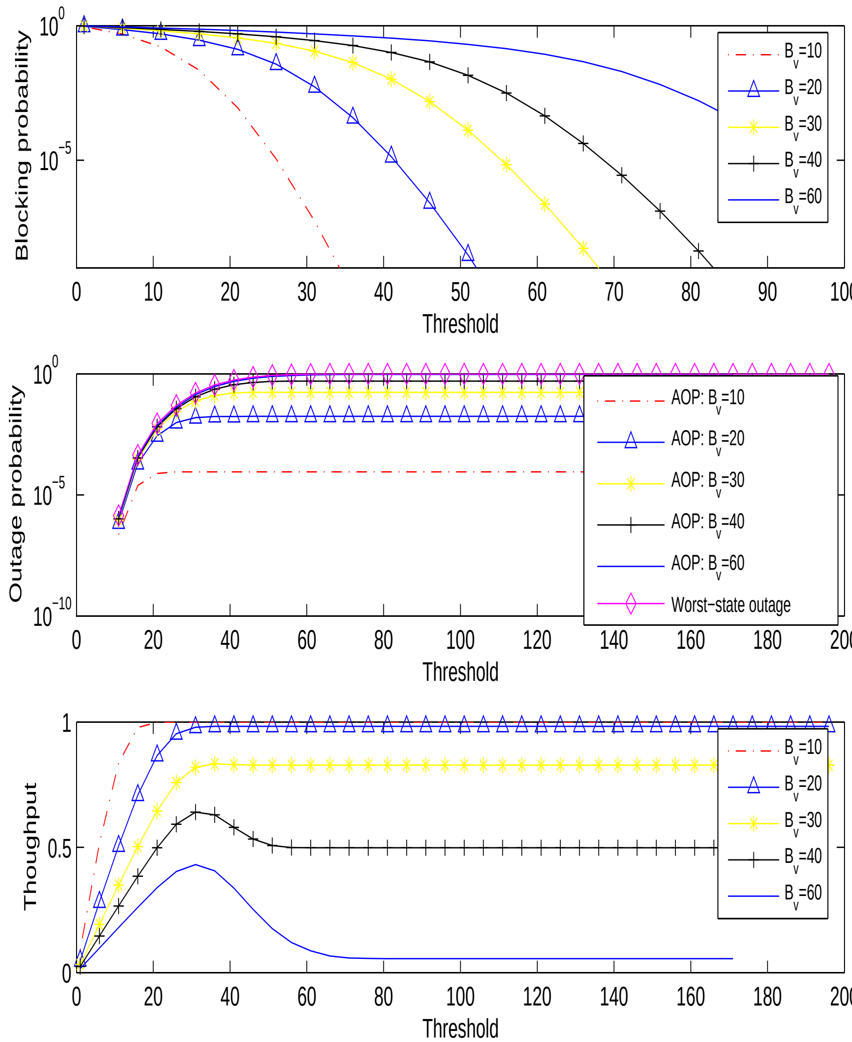

Figure 3.3. Single-voice-class: performance as a function of the threshold.

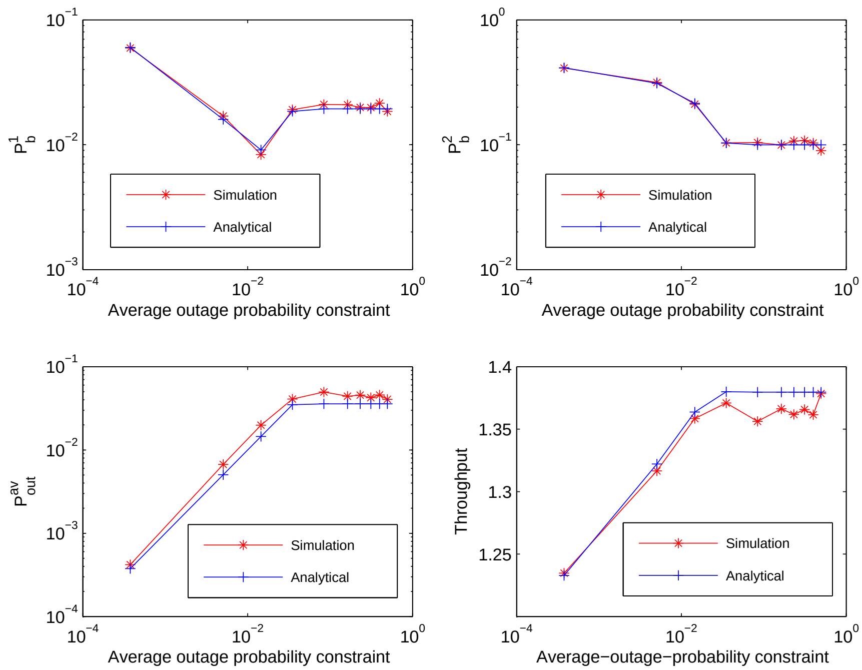

Figure 3.4. Performance comparison between simulation and analytical results with $p_v = 1$.

3.7.4 Comparison between multiple antenna and single antenna systems

Figure 3.5 compares the analytical performance for single antenna and two-antenna systems, obtained through linear programming (LP) approach, in which P_b is obtained by $(P_b^1 + P_b^2)/2$. It is observed that the system performance can be dramatically improved by employing antennas at the BS.

As mentioned before, allowing for outage probability in the physical layer can reduce the overall blocking probability and improve the throughput. For single antenna systems, the outage is introduced by employing voice activity, while for beamforming systems, outage is introduced by both voice activity and randomly distributed AoAs. From Figure 3.5, it is also observed that allowing outage for multiple antenna systems provides a more flexible

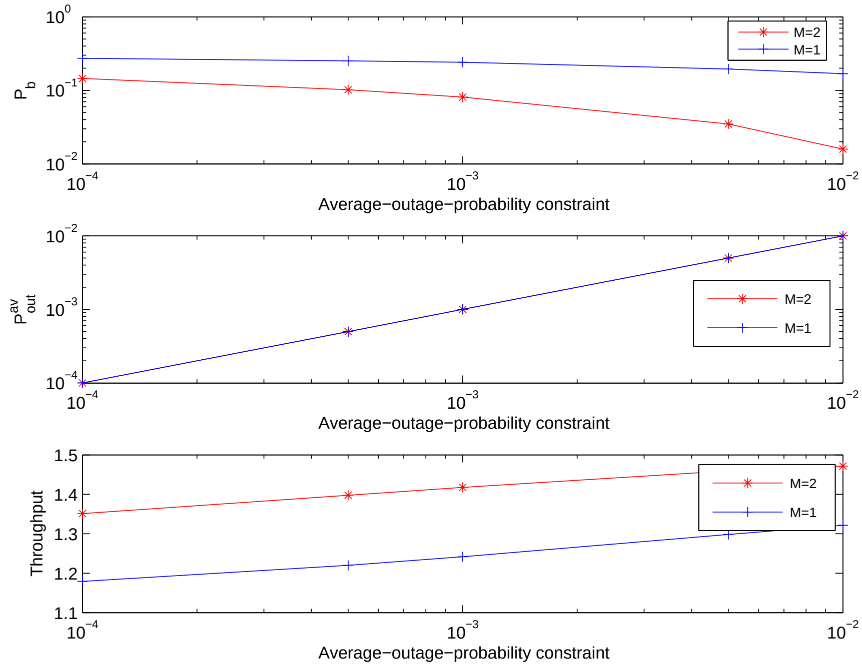

Figure 3.5. Performance comparison between single antenna and two-antenna systems with $p_v = 3/8$.

way to handle QoS requirements. For example, when average-outage-probability (AOP) constraint is relaxed from 10^{-4} to 10^{-2}, the overall blocking probability for single antenna system can be reduced from 0.27 to 0.17, i.e., reduced by 37%, while for a two-antenna sysstem, the blocking probability can be reduced from 0.14 to 0.016, i.e., reduced by 88%.

3.7.5 Comparison between proposed and existing CAC policies

Table 3.2 presents the simulation results for the proposed SMDP-based CAC policy and an existing complete-sharing (CS)-based policy, in which $\Psi_1 = 0.02$, $\Psi_2 = 0.15$, $\rho_{av} = 0.05$, $M = 2$, $B_1 = B_2 = 0$ and $p_v = 1$. For a CS-based CAC policy, an incoming user is accepted if and only if the exact PCFC can be satisfied.

From the simulation results, it is observed that a SMDP-based CAC policy can guarantee the QoS requirements in terms of blocking probability and outage probability, while for a CS-based CAC policy, the achieved blocking probability for class 1 is 0.04, which exceeds its constraint of 0.02. This can be explained by the fact that a SMDP-based CAC policy has the capability of flexibly allocating resources according to QoS requirements and may block some users to reserve spaces for other users, while for a CS-based CAC policy, the incoming user is accepted whenever there is enough space in the physical layer and no heterogeneous traffic is distinguished.

In Table 3.2, it is also observed that the achieved system throughput for a CS-based policy is even larger than the maximum-throughput SMDP-based policy. This is because: 1) The worst-outage-probability constraint is guaranteed for SMDP-based CAC policy, while no such constraint is imposed for CS-based policy, which relaxes the state space and increases the acceptance probability; 2) For a SMDP-based CAC policy, the information for AoA realizations is not required at the decision epoch, while for a CS-based CAC, this information must be exploited to design the policy, which improves the system throughput, while at the same time also increases the latency.

We remark that for a CS-based CAC policy, the outage probability can be reduced to zero for a system without user mobility. For a practical system with user mobility, non-zero outage probability is introduced, and a CS-based CAC policy cannot guarantee the outage probability constraint, while for a SMDP-based policy, the outage probability evaluation in (3.9) can be easily extended to include user mobility and QoS in terms of blocking probability and outage probability can both be satisfied.

Table 3.2. Comparison between SMDP-based CAC and CS-based CAC policies.

CAC policies	P_b^1	P_b^2	P_{out}^{av}	Throughput
CS-based policy	0.0417	0.0742	0	1.4212
SMDP-based policy	0.0197	0.0991	0.0481	1.3619

3.7.6 Numerical example in a practical UMTS system

In the following, we evaluate the performance of the proposed CAC policies using actual UMTS system [42] parameters. Wideband CDMA (WCDMA) radio access is employed with a chip rate of 3.84 Mcps and we consider two service classes: real-time voice and video classes. Voice and video sessions can be modeled as independent Poisson arrival processes and their durations are assumed to be exponentially distributed. These services are represented by a constant-bit-rate flow with 37.5% voice activity (VA) in circuit-switched mode over dedicated channels. The maximum data transmission rate for these two classes are 32kbps and 128kbps, respectively. As determined by experiment, both voice and video services have a mean duration of 3 minutes [25], corresponding to a departure rate of 0.0056. The arrival rates for class 1 and class 2 users are assumed to be both 0.02, respectively.

In Figure 3.6, we compare the performances for single and multiple antenna systems in which optimal CAC policy is employed, $\rho_w = 0.5$ and $B_1 = B_2 = 0$. The performances with and without voice activity are also studied. It is observed that employing voice activity and multiple antennas can improve the system performance. For example, with a single antenna at the BS, i.e., $M = 1$, employing VA can improve the throughput 10%, from 0.034 to 0.037. By employing two antennas at the BS, the system throughput can be further improved by 14%, from 0.037 to 0.042.

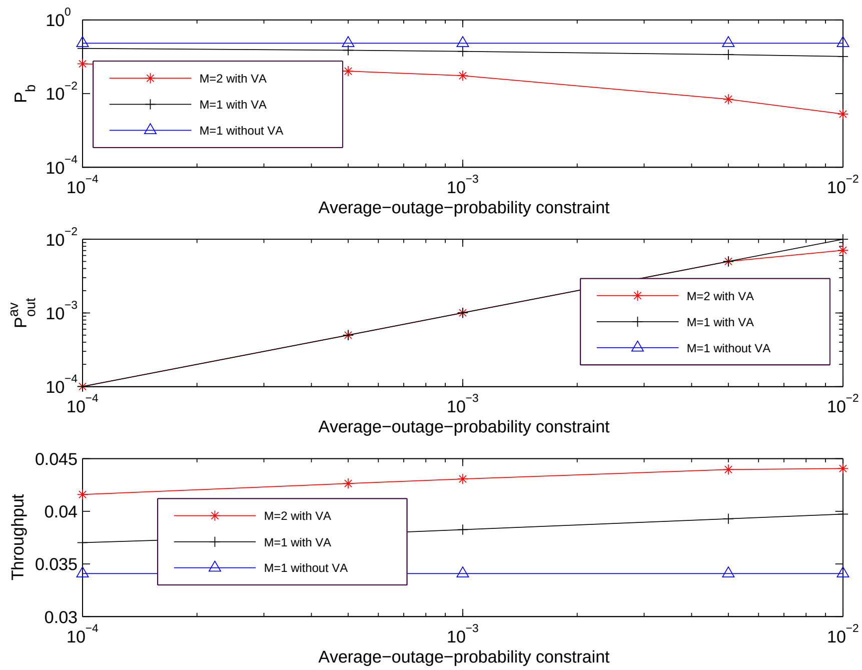

Figure 3.6. Blocking probabilities, outage probability and system throughput for an optimal CAC policy.

From Figure 3.6, it is also observed that with voice activity, when the average-outage-probability (AOP) constraint is relaxed from 10^{-4} to 10^{-2}, the overall blocking probability for a single antenna system can be reduced by 39%, from 0.167 to 0.101, while for a two-antenna system, the blocking probability can be reduced by 96%, from 0.064 to 0.0028. Therefore, as we mentioned previously, relaxing outage constraints in multiple antenna systems provides a more flexible way to handle QoS requirements.

From this example, we conclude that the proposed optimal CAC policy may perform well in a practical system.

Throughout the thesis, we employ MATLAB optimization toolbox to solve the linear programming problem. The time required to solve the problem strongly depends on the size of the feasible state space. Table 3.3 presents the sizes of the state space as well as the

Table 3.3. Size of feasible state space and CPU time required to solve the LP problem

	Size of state space	CPU time in seconds
Single antenna system without voice activity	49	1.5
Two-antenna system without voice activity	143	13
Two-antenna system with voice activity $p_v = 3/8$	367	87

time required to solve the LP problem. It is observed that employing multiple antennas and voice activity can increase the size of the feasible state space which leads to an increased CPU processing time. This represents a tradeoff between the improved system performance and the computational complexity.

3.8 Conclusions

We investigate the cross-layer admission control problem for a CDMA beamforming system. A maximum-throughput optimal CAC policy is proposed, which optimizes the long-term system performance while simultaneously guaranteeing all the QoS requirements. The proposed optimal CAC policy is capable of achieving a significant performance gain in terms of blocking probability, outage probability and system throughput, compared with the case of single antenna systems. The multiple QoS requirements can be flexibly handled by employing the tradeoff between network and physical layer performance.

Chapter 4

Low-Complexity Suboptimal Call Admission Control

4.1 Introduction

In the previous chapter, we have proposed an optimal CAC policy, which can achieve a maximum system throughput while simultaneously guaranteeing QoS. This optimal CAC policy requires information on outage probability for each system state. For most systems of interest, outage probability must be evaluated numerically, which increases the computational complexity.

In this chapter, we propose a low-complexity suboptimal CAC policy. In contrast to the optimal CAC policy in which AOP is ensured by evaluating the outage probability for each system state and adding a SMDP constraint, the proposed suboptimal CAC policy guarantees the average outage probability constraint by deriving an approximate power control feasibility condition (PCFC), which limits the number of users that a system can accommodate, and then employing a separate reduced-outage-probability (ROP) algorithm.

The rest of this chapter is organized as follows. Signal model and problem formulation are discussed in Sections 4.2 and 4.3, respectively. In Sections 4.4 and 4.5, the outage control is designed which includes the derivation of an approximate PCFC and ROP algorithms. The cross-layer CAC policy is then derived in Section 4.6, and the numerical results are presented in Section 4.7.

4.2 Signal Model

4.2.1 Traffic model

In this chapter, we consider the same signal model as the one in Chapter 3, i.e., a single-cell CDMA beamforming system which supports J classes of users. Requests for connections are assumed to be Poisson distributed, with rates $\lambda_j,\ j = 1, .., J$. The call durations are assumed to have an exponential distribution with mean duration $\frac{1}{\mu_j}, j = 1, .., J$. Without loss of generality, class 1 is assumed to be voice class with voice activity factor p_v.

Whenever an incoming user arrives, the CAC policy decides if this user can be accepted, stored in the buffer, or blocked if the buffer is full. Each class of users shares a common buffer with size B_j for class j. Network layer QoS is characterized by blocking probability and connection delay constraints.

4.2.2 Signal model at the physical layer

We consider a CDMA beamforming system which has M antennas at the BS and a single antenna for each user. Suppose there are K active users in the system, and a channel with slow fading is assumed.

As shown in Chapter 3, the received signal-to-interference ratio (SIR) for a desired user k can be written as

$$SIR_k = \frac{W}{R_k} \frac{p_k \phi_{kk}^2}{\sum_{i \neq k} p_i \phi_{ik}^2 + \eta_0 W} \tag{4.1}$$

where W and R_k denote the bandwidth and data rate for user k, respectively, and the ratio $\frac{W}{R_k}$ represents the processing gain; $p_i = P_i G_i^2$ denotes the received power for user i, and η_0 denotes the one-sided power spectral density of background additive white Gaussian noise (AWGN); the parameters ϕ_{ii}^2 and ϕ_{ik}^2 are defined as

$$\phi_{ik}^2 = \left|\mathbf{w}_k^H \mathbf{a}_i\right|^2 \tag{4.2}$$

which capture the effects of beamforming. With a spatially matched filter receiver, we have $\mathbf{w}_k = \mathbf{a}_k$.

Physical layer QoS can be represented by a target average outage probability, denoted by ρ_{av}^j, where $j = 1, .., J$. The achieved outage probability for class j is denoted by P_{out}^j.

4.3 Problem Formulation

In this chapter, we aim to derive a low-complexity CAC policy which can guarantee the QoS requirements in terms of average outage probability, blocking probability and connection delay, while simultaneously minimizing the blocking probability. We note that for the suboptimal CAC policy, the system throughput cannot be directly maximized due to the unavailable information on outage probability. Alternatively, a minimum-blocking-probability criterion is employed in this chapter, which is equivalent to the maximum-throughput criterion for a fixed outage probability.

4.4 Power Control Feasibility Condition

For an optimal CAC policy discussed in the previous chapter, to formulate the state space, the outage probability based on an exact PCFC should be evaluated for each possible state, which leads to high computation complexity.

In order to reduce the complexity, we now derive an approximate PCFC to formulate the state space. This approximate PCFC can be represented by a linear function in terms of the maximum number of users that the system accommodates, and therefore can achieve a dramatically reduced complexity.

4.4.1 Approximate PCFC

A system state is defined as $\mathbf{s} = [n_q^1, n_s^1.., n_q^J, n_s^J]$, where n_q^j and n_s^j denote the number of accepted users in the queue and in the server, respectively. Denote K_j, where $j = 1,..,J$, as the number of active users in class j.

Under the assumption that class 1 is the voice class, as discussed previously, K_1 has a Binomial distribution with parameter $[n_s^1, p_v]$. The number of active users in class j, where $j = 2,..,J$, is just the number of accepted users in that class, i.e. $K_j = n_s^j$.

Without loss of generality, we consider an arbitrary user i in class 1, where $i = 1,..,K_1$, the SIR requirements of user i can be written as

$$SIR_i \geq \gamma_i \tag{4.3}$$

where γ_i denotes the target SIR for user i.

By considering specific traffic classes and letting SIR achieve its target value, the expression in (4.1) can be written as follows,

$$\gamma_i = \frac{p_i \phi_{ii}^2 \frac{W}{R_1}}{\sum_{l=1, l\neq i}^{K_1} p_l \phi_{il}^2 + \sum_{l=1}^{K_2} p_l \phi_{il}^2 + .. \sum_{l=1}^{K_J} p_l \phi_{il}^2 + \sigma^2}$$

where $\sigma^2 \triangleq \eta_0 W$ denotes the noise variance, and p_i represents the received power for user i.

Since users in the same class have the same target SIR requirements, it is reasonable to assume that the same class users have the same received power. By denoting the received power in class j as p_j, the above expression can be written as

$$\gamma_i = \frac{p_1 \phi_{ii}^2 \frac{W}{R_1}}{p_1(K_1-1)\beta_1 + \sum_{j=2}^{J} p_j K_j \beta_j + \sigma^2} \tag{4.4}$$

where $\beta_1 = \frac{1}{K_1-1}\sum_{l=1,l\neq i}^{K_1} \phi_{il}^2$ and $\beta_j = \frac{1}{K_j}\sum_{l=1}^{K_j} \phi_{il}^2$ for $j = 2,..,J$.

By exchanging the numerator and denominator, Equation (4.4) is equivalent to

$$\frac{p_1(K_1-1)\beta_1 + \sum_{j=2}^{J} p_j K_j \beta_j + \sigma^2}{p_1 \frac{W}{\gamma_1 R_1}} = \phi_{ii}^2 \tag{4.5}$$

where $i = 1,..,K_1$.

Summing the above K_1 equations, and calculating the sample average, we obtain

$$\frac{p_1(K_1-1)\alpha_1 + \sum_{j=2}^{J} K_j p_j \alpha_j + \sigma^2}{p_1 \frac{W}{\gamma_1 R_1}} = \frac{1}{K_1}\sum_{i=1}^{K_1} \phi_{ii}^2 \tag{4.6}$$

where $\alpha_1 = \frac{1}{K_1}\sum_{i=1}^{K_1} \beta_1$ and $\alpha_j = \frac{1}{K_1}\sum_{i=1}^{K_1} \beta_j$.

When the number of users is large enough, the above $\alpha_1,..,\alpha_J$ can be approximated by their mean values, and (4.6) can be further simplified as

$$\begin{aligned} & E_1[\phi_{des}] \\ &= \frac{p_1(K_1-1)E_{11}[\phi_{int}] + \sum_{j=2}^{J} K_j p_j E_{1j}[\phi_{int}] + \sigma^2}{p_1 \frac{W}{\gamma_1 R_1}} \end{aligned} \tag{4.7}$$

in which $E_{cn}[\phi_{int}]$ is the expected fraction of an interferer user in class n passed by a beamforming weight vector for a desired user in class c, where $c,n = 1,..,J$, while $E_j[\phi_{des}]$ is the expected fraction of a desired user in class j passed by its beamforming weight vector, where $j = 1,..,J$.

The AoAs of active users in the system are assumed to have identically independent distributions, which are independent of a user's specific class. Therefore, it is reasonable to assume that $E_{cn}[\phi_{int}]$ is also independent of specific classes c and n, which can be denoted by $E[\phi_{int}]$. Similarly, $E_j[\phi_{des}]$ is independent of class j, and can be denoted by $E[\phi_{des}]$.

$E[\phi_{des}]$ and $E[\phi_{int}]$ represent the expected fractions of the desired user's power and interference, respectively.

From the above discussion, (4.7) can be written as

$$\frac{p_1(K_1-1)E[\phi_{int}]+\sum_{j=2}^{J}K_j p_j E[\phi_{int}]+\sigma^2}{p_1\frac{W}{\gamma_1 R_1}}=E[\phi_{des}].$$

By exchanging the numerator and denominator of the above equation, we have

$$\frac{p_1\frac{W}{\gamma_1 R_1}}{p_1(K_1-1)\frac{E[\phi_{int}]}{E[\phi_{des}]}+\sum_{j=2}^{J}K_j p_j\frac{E[\phi_{int}]}{E[\phi_{des}]}+\frac{\sigma^2}{E[\phi_{des}]}} = 1. \tag{4.8}$$

The QoS requirement for class 1 in (4.8) can be extended to any class j,

$$\frac{p_j\frac{W}{\gamma_j R_j}}{p_j(K_j-1)\frac{E[\phi_{int}]}{E[\phi_{des}]}+\sum_{c=1,c\neq j}^{J}K_c p_c\frac{E[\phi_{int}]}{E[\phi_{des}]}+\frac{\sigma^2}{E[\phi_{des}]}}=1 \tag{4.9}$$

where $j=1,..,J$.

The power solution can be obtained by solving the above J equations [68]

$$p_j=\frac{\frac{\sigma^2}{E[\phi_{int}]}}{(1+\frac{W}{\gamma_j R_j\frac{E[\phi_{int}]}{E[\phi_{des}]}})[1-\sum_{j=1}^{J}\frac{K_j}{1+\frac{W}{\gamma_j R_j\frac{E[\phi_{int}]}{E[\phi_{des}]}}}]} \tag{4.10}$$

where $j=1,..,J$.

Positivity of power solution implies the following power control feasibility condition

$$\sum_{j=1}^{J}\frac{K_j}{1+\frac{W}{\gamma_j R_j\frac{E[\phi_{int}]}{E[\phi_{des}]}}}<1. \tag{4.11}$$

As shown in [93], $E[\phi_{int}]$ and $E[\phi_{des}]$ can be determined numerically for a beamforming system.

With this condition, the transmitted power for each user can be adjusted to meet the target SIR requirement. From (4.10), the transmitted power of user i in class j can be represented by

$$p_j^i = \frac{\frac{\sigma^2}{E[\phi_{int}]}}{G_i(1+\frac{W}{\gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}})[1-\sum_{j=1}^{J}\frac{K_j}{1+\frac{W}{\gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}}}]} \tag{4.12}$$

where G_i denotes the channel gain for user i in class j, where $i = 1,..,K_j$, and $j = 1,..,J$.

We have derived the power control feasibility condition in terms of K_j, which is the number of active users for class j, where $j = 1,..,J$. In an admission problem, we need to derive a PCFC in terms of n_s^j, which limits the maximum number of accepted users for class j, where $j = 1,..,J$. In this case, the approximate PCFC in (4.11) can be modified as follows

$$\frac{n_s^1 p_v}{1+\frac{W}{\gamma_1 R_1 \frac{E[\phi_{int}]}{E[\phi_{des}]}}} + \sum_{j=2}^{J} \frac{n_s^j}{1+\frac{W}{\gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}}} < 1. \tag{4.13}$$

We note that the above approximated power control feasibility condition is independent of the angle of arrivals, and thus can provide a less-complicated offline-CAC policy, which does not require the estimation of the current AoA realizations of each user.

4.4.2 Accuracy of the approximate PCFC

In the derivation of the above PCFC, it is assumed that the users in the same class have the same powers, and α_j in (4.6) can be estimated by its mean value. It is obvious that the above assumptions are reasonable for a large number of users. However, for a randomly arriving and departing process, the number of users varies with time, and as a result may not be always large, which impacts the accuracy of the approximate PCFC. With this approximate PCFC, the instantaneous SIR fluctuates around the target SIR which leads to an outage probability in the physical layer. Next we discuss how to mitigate the outage probability.

4.5 ROP Algorithms

In the previous section, we derive an approximate PCFC which limits the maximum number of users that the system can accommodate. This approximation, however, increases outage probability. Reduced-outage-probability (ROP) algorithms are then employed to ensure a target outage probability constraint.

In this section, we propose two simple ROP algorithms, denoted by ROP-I and ROP-II, respectively, which can reduce the outage probability to a tolerably small level by adjusting ROP parameters.

4.5.1 ROP-I

The proposed ROP-I algorithm aims to reduce the outage probability by reducing the target SIR.

For a given transmission scheme with a target bit-error-rate (BER) or packet-error-rate (PER) requirement, an equivalent target SIR, $\gamma_j,\ j = 1,..,J$, can be obtained. The power control feasibility condition based on this target SIR, γ_j, is shown as follows

$$\frac{n_s^1 p_v}{1+\frac{W}{\gamma_1 R_1 \frac{E[\phi_{int}]}{E[\phi_{des}]}}} + \sum_{j=2}^{J} \frac{n_s^j}{1+\frac{W}{\gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}}} < 1. \tag{4.14}$$

The target transmitted power can be derived according to (4.12). At the transmitter, instead of using the original transmission scheme with target SIR γ_j, the transmitter adjusts its modulation and coding scheme to reduce the target SIR by a decrease-factor, denoted by α_{dec}, where $\alpha_{dec} < 1$. Without loss of generality, we assume the same decrease-factor α_{dec} for all users. With an appropriately chosen α_{dec}, the outage probability constraints can be guaranteed.

Figure 4.1 illustrates the ROP-I algorithm as well as how to choose an appropriate ROP parameter α_{dec}, in which α_0 denotes the initial ROP parameter, P_{out}^j and ρ_{av}^j denote the achieved AOP and AOP constraint for class j, respectively, $[\rho_{av}^j - \bar{\rho}_j, \rho_{av}^j + \bar{\rho}_j]$ denotes the allowed interval on the achieved AOP, and Δ represents the adjustment step for the ROP parameters. By following the search procedure in Figure 4.1, an appropriate α_{dec} can be obtained, which can reduce the outage probability to a tolerable level.

We remark that in the case of ROP-I, the network-layer performance remains same with the decrease of α_{dec}. Therefore, the outage probability can be reduced to a very small level without affecting network-layer performance. However, there is a cost in spectral efficiency due to the enhanced modulation and coding schemes. The tradeoff in ROP-I is between the power efficiency and spectral efficiency.

4.5.2 ROP-II

The proposed ROP-II algorithm aims to reduce the outage probability by imposing a more restrictive PCFC, which is based on an increase target SIR. We use $\alpha_{inc} \geq 1$ to denote the increase-factor. For simplicity, α_{inc} is chosen to be the same for all the users. By increasing the target SIR for class j to $\alpha_{inc}\gamma_j$, the power control feasibility in (4.13) is revised to

$$\frac{n_s^1 p_v}{1 + \frac{W}{\alpha_{inc}\gamma_1 R_1 \frac{E[\phi_{int}]}{E[\phi_{des}]}}} + \sum_{j=2}^{J} \frac{n_s^j}{1 + \frac{W}{\alpha_{inc}\gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}}} < 1 \tag{4.15}$$

and the target transmitted power in (4.12) is modified to

$$p_j^i = \frac{\frac{\sigma^2}{E[\phi_{int}]}}{G_i(1 + \frac{W}{\alpha_{inc}\gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}})[1 - \sum_{j=1}^{J} \frac{K_j}{1+\frac{W}{\alpha_{inc}\gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}}}]}. \tag{4.16}$$

With an appropriately chosen α_{inc}, the outage probability constraint can be satisfied. The ROP-II algorithm as well as the search procedure for α_{inc} are illustrated in Figure 4.2.

The shortcoming of the proposed ROP-II algorithm is that the network layer performance, in terms of blocking probability and connection delay, degrades with α_{inc}. The

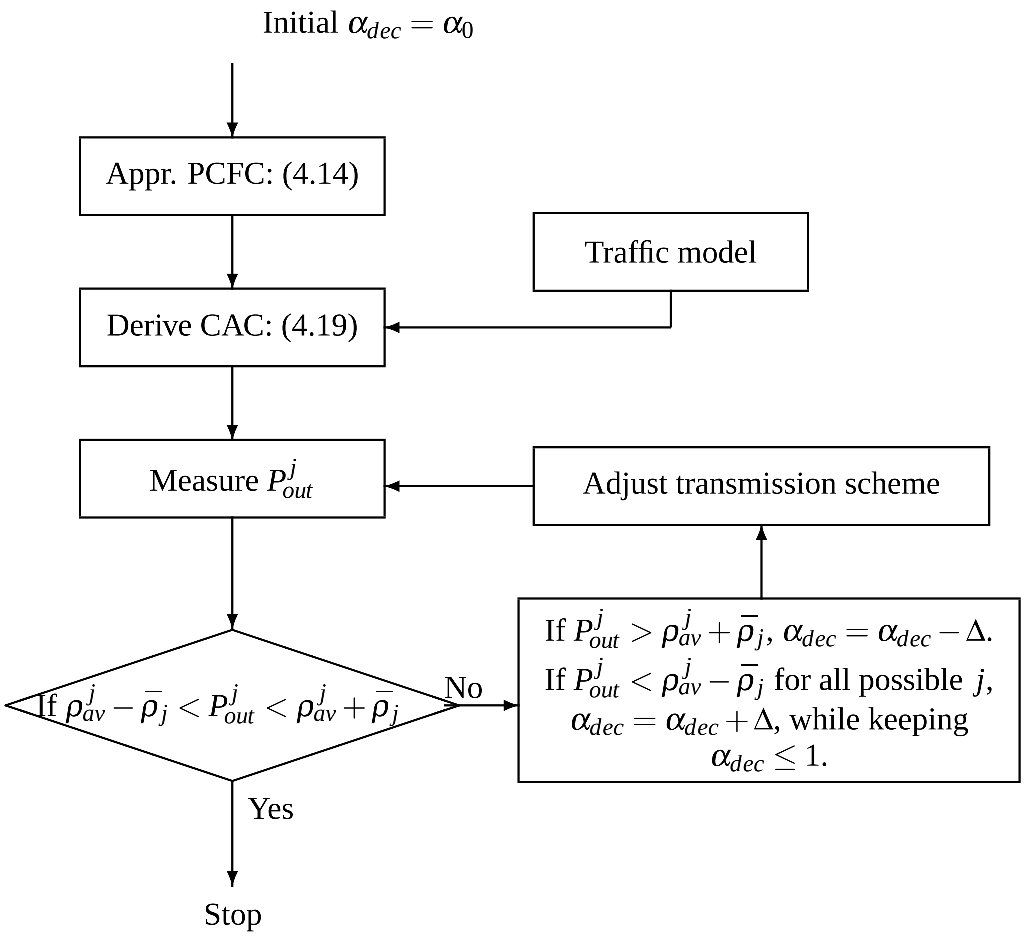

Figure 4.1. Suboptimal CAC policy based on ROP-I.

parameter α_{inc} represents the tradeoff between network-layer and physical-layer performances.

4.6 Suboptimal CAC Policy based on the Approximate PCFC and ROP

The CAC policy is performed at the BS, and the following information is necessary to derive an admission control policy: traffic model in the system, such as arrival and departure

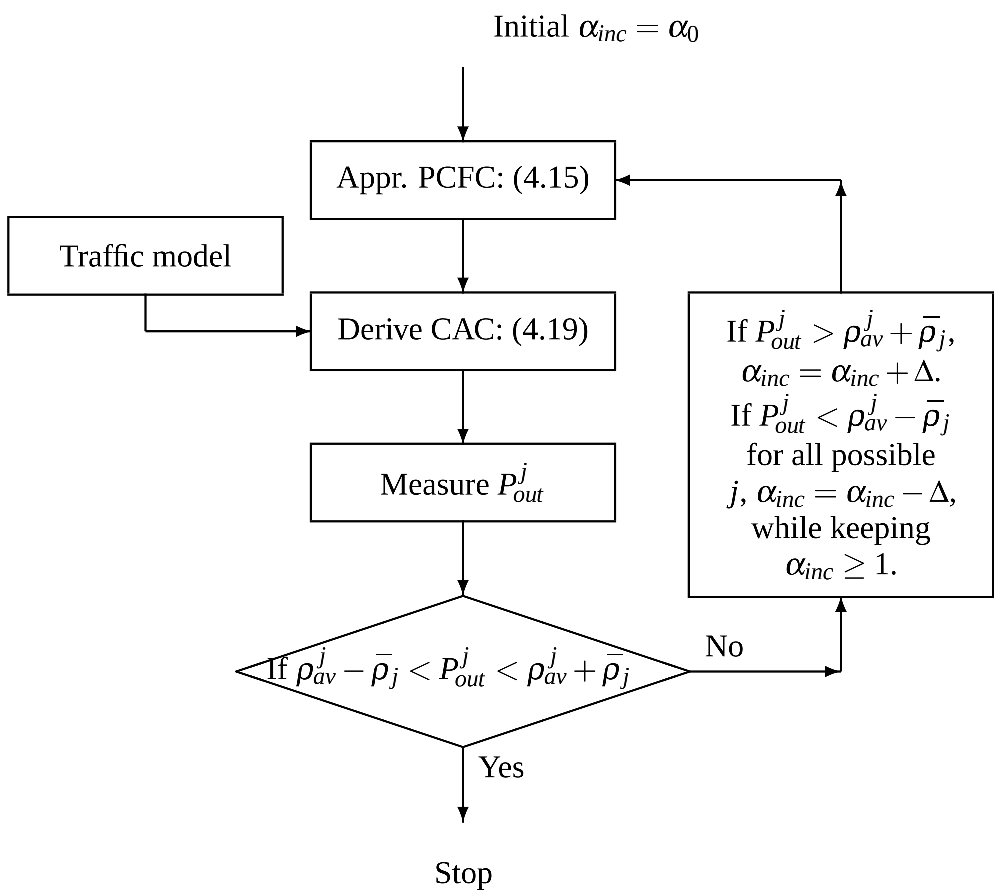

Figure 4.2. Suboptimal CAC policy based on ROP-II.

rate, and QoS requirements in both physical and network layers.

Similar to the optimal CAC policy, the suboptimal CAC policy can be obtained by formulating and solving a SMDP.

With an appropriately chosen ROP-I parameter, α_{dec}, which is presented in Figure 4.1, the state space can be rewritten as

$$S = \left\{ \mathbf{s} : n_q^j \leq B_j, j = 1, .., J; \frac{n_s^1 p_v}{1 + \frac{W}{\gamma_1 R_1 \frac{E[\phi_{int}]}{E[\phi_{des}]}}} + \sum_{j=2}^{J} \frac{n_s^j}{1 + \frac{W}{\gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}}} < 1 \right\} \tag{4.17}$$

where system state $\mathbf{s}$ is defined in (3.7).

With an appropriately chosen ROP-II parameter, α_{inc}, which is presented in Figure 4.2, the state space can be rewritten as

$$S = \left\{ \mathbf{s} : n_q^j \leq B_j, j = 1, .., J; \frac{n_s^1 p_v}{1 + \frac{W}{\alpha_{inc} \gamma_1 R_1 \frac{E[\phi_{int}]}{E[\phi_{des}]}}} + \sum_{j=2}^{J} \frac{n_s^j}{1 + \frac{W}{\alpha_{inc} \gamma_j R_j \frac{E[\phi_{int}]}{E[\phi_{des}]}}} < 1 \right\}. \tag{4.18}$$

The other components of a SMDP remain the same as the SMDP formulation for optimal CAC policy, which are summarized in Table 4.1.

After formulating the CAC problem as a SMDP, the suboptimal CAC policy can be obtained by a linear programming approach. We remark that the maximum-throughput criterion employed in the optimal CAC policy requires the evaluation of outage probability for each state, and as a result is inappropriate for the suboptimal CAC policy. We instead aim to find a policy R^* which minimizes the blocking probability for any initial system state, i.e.,

$$R^* = \arg\min_{R \in \mathbf{R}} \lim_{t \to \infty} \frac{1}{T} E \left\{ \int_0^T \sum_{j=1}^{J} \eta_j (1 - a_j(t))(1 - \delta(B_j - n_q^j(t)))dt \right\}$$

while simultaneously guaranteeing QoS.

The suboptimal CAC policy can be obtained by solving the following LP problem:

$$\max_{z_{\mathbf{sa}} \geq 0, \mathbf{s}, \mathbf{a}} \sum_{\mathbf{s} \in S} \sum_{\mathbf{a} \in A_{\mathbf{s}}} \sum_{j=1}^{J} \eta_j (1 - a_j)(1 - \delta(B_j - n_q^j)) \tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} \tag{4.19}$$

Table 4.1. Components of the SMDP which represents the suboptimal CAC problem.

SMDP components	Notation	Expression
System state	$\mathbf{s}$	$\mathbf{s} = [n_q^1, n_s^1, .., n_q^J, n_s^J]$
State space	S	Given in (4.17) or (4.18)
Decision epochs	t_k	The set of all arrival and departure instances.
Action	$\mathbf{a}$	Given in (3.27)
Admissible action space	$A_{\mathbf{s}}$	Given in (3.28)
Expected holding time	$\tau_{\mathbf{s}}(\mathbf{a})$	Given in (3.29)
Transition probability	$p_{\mathbf{sy}}(\mathbf{a})$	Given in (3.30)
Policy	R	$R = \{R_{\mathbf{s}} : S \rightarrow A \| R_{\mathbf{s}} \in A_{\mathbf{s}}, \forall \mathbf{s} \in S\}$
Constraints		$P_b^j \leq \Psi_j$ and $\text{Delay}_j \leq D_j$.

subject to the constraints (3.37)-(3.40), where η_j denotes the weighting factor for class j.

In the above optimization problem, there is no explicit outage probability constraint. Instead, the outage probability constraint is guaranteed by adjusting the ROP parameter. The suboptimal CAC policies based on ROP-I and ROP-II as well as the search procedures for the ROP parameters are illustrated in Figures 4.1 and 4.2, respectively.

In comparing the optimal and suboptimal CAC policies, some remarks are in order:

- *Complexity*: as shown in (3.15), to derive the outage probability for an optimal CAC policy, multiplication of two K-by-K matrices as well as N_r eigenvalue computations are required, where N_r can be large, e.g., $N_r = 1000$, to ensure accuracy. This results in very high computational complexity, while for a suboptimal CAC policy, a linear PCFC is employed to decide if a system state is feasible, which only requires J summations and J multiplications, where $J \leq 10$, a significant reduction.

- *QoS guarantee*: both optimal and suboptimal CAC policies can ensure network-layer QoS requirements. However, regarding the physical layer QoS, the optimal CAC policy can ensure the more stringent WSOP constraint as well as an AOP constraint, while the suboptimal CAC policy can only ensure an AOP constraint.

- *Performance gain*: the suboptimal CAC policy relies on a separate ROP algorithm to mitigate the outage probability, which may degrade the network layer performance or reduce the spectral efficiency.

- *Measurement*: for the suboptimal CAC policy, outage measurements are necessary to monitor the ROP parameter, which introduces a delay. For the optimal CAC policy, no such procedure is required.

4.7 Numerical Examples

In this section, we first evaluate the performance of the proposed suboptimal CAC policy, which is then compared with single antenna systems. After that, the performance comparisons between suboptimal and optimal CAC policies are presented. Finally, we compare the proposed suboptimal policy with the existing CS-based policy.

4.7.1 Simulation parameters

In the following examples, we consider a circular antenna array with a uniformly distributed AoA. We remark that the proposed CAC policies can be applied to any other array geometry and AoA distribution. For simplicity, it is assumed that there is no user mobility.

In our examples, a two-class system is considered. The SIR requirements for each class are given as $\gamma_1 = 10$ dB and $\gamma_2 = 7$ dB, and the rate for each class is set to $R_1 = 48$ kbps

Table 4.2. Simulation parameters.

W	3.84 MHz	η_0	10^{-6}
R_1	48 kbps	R_2	144 kbps
λ_1	1	λ_2	0.5
μ_1	0.25	μ_2	0.1375
Ψ_1	0.1	Ψ_2	0.1
D_1	1.5030	D_2	1.5750
η_1	0.5	η_2	0.5

Table 4.3. Numerical values of $E[\phi_{des}]$ and $E[\phi_{int}]$ for a beamforming system.

M	1	2	3	4	5	6
$E[\phi_{des}]$	1.0	1.0	1.0	1.0	1.0	1.0
$E[\phi_{int}]$	1.0	0.5463	0.3950	0.3241	0.2460	0.2058

and $R_2 = 144$ kbps. The total bandwidth is $W = 3.84MHz$, and the AWGN noise can be characterized by spectral density $\eta_0 = 10^{-6}$.

The arrival and departure rates for class 1 and class 2 users are denoted by $\lambda_1 = 1$, $\lambda_2 = 0.5$, $\mu_1 = 0.25$ and $\mu_2 = 0.1375$, respectively. In the network layer, the blocking probability constraints for the two classes are set to $\Psi_1 = \Psi_2 = 0.1$, unless specified otherwise. The constraints on connection delay are set to 1.67 and 3.5 seconds, respectively, which can be equivalently represented by the average queue length, $D_1 = 1.5030$ and $D_2 = 1.5750$. Numerical values of parameters $E[\phi_{des}]$ and $E[\phi_{int}]$, derived in [93], are shown in Table 4.3. Simulation parameters are summarized in Table 4.2.

4.7.2 CAC policy based on ROP-I algorithm

For a SMDP-based CAC policy combined with ROP-I, Table 4.4 and Figures 4.3-4.4 present the blocking probability, the outage probability and the system throughput as a function of the decrease-factor, respectively. Two and six antennas are employed at the BS, respectively, and no buffer is employed for both classes. In this example, the case of a single antenna is not included due to the infeasible buffer configuration for a SMDP formulation. It is observed that with a decreased α_{dec}, the outage probability and the system throughput can be improved while simultaneously maintaining the network layer performance in terms of blocking probability. For example, with six antennas at the BS, when the decrease-factor is decreased from 1 to 0.7 by an enhanced modulation and coding scheme, the outage probability for class 1 users can be decreased from 0.4438 to 0.0013, i.e., decreased by 99.7%, and the throughput can be improved from 0.84 to 1.5, i.e., improved by 78.6%, while remaining the blocking probability within a very small level. As observed, the throughput can be improved by either increasing the number of antennas, or decreasing the decrease-factor.

We note that the above performance gain in terms of reduced outage probability is achieved at a necessary loss in spectral efficiency due to enhanced modulation and coding. With an decreased α_{inc}, the spectral loss is increased. From the above simulation results, it is observed that the spectral loss can be reduced by increasing the number of antennas at the BS. For a given number of antennas, an appropriate chosen parameter α_{dec} can mitigate the outage probability to the desired level while simultaneously minimizing the spectral loss. For example, with six antennas at the BS with an average outage probability constraint $\rho_{av}^1 = \rho_{av}^2 = 0.03$ and adjustment step $\Delta = 0.1$, by following the procedure shown in Figure 4.1, we can obtain $\alpha_{dec} = 0.8$, which introduces the least spectral efficiency loss while guaranteeing the outage probability constraints.

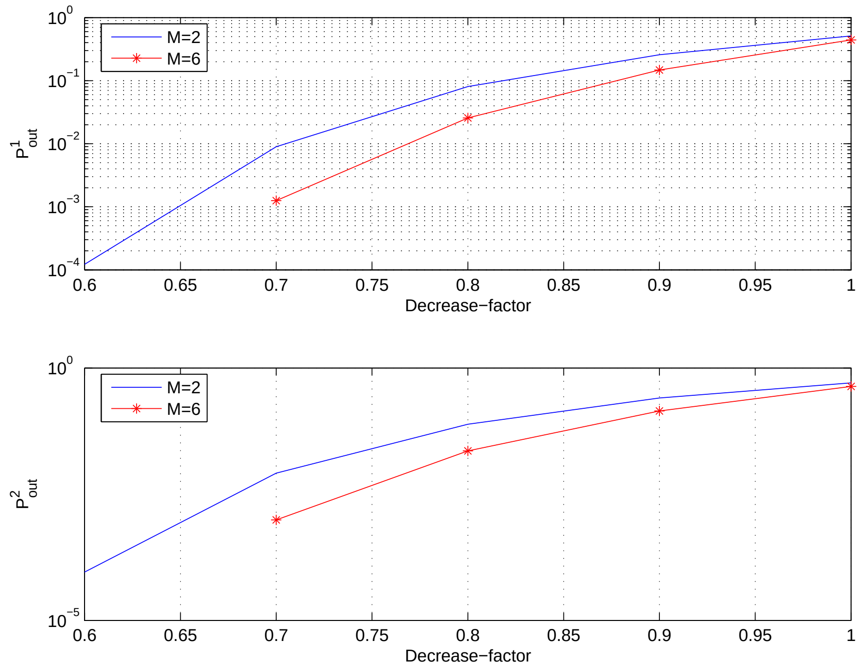

Figure 4.3. Suboptimal CAC policy based on ROP-I: outage probability.

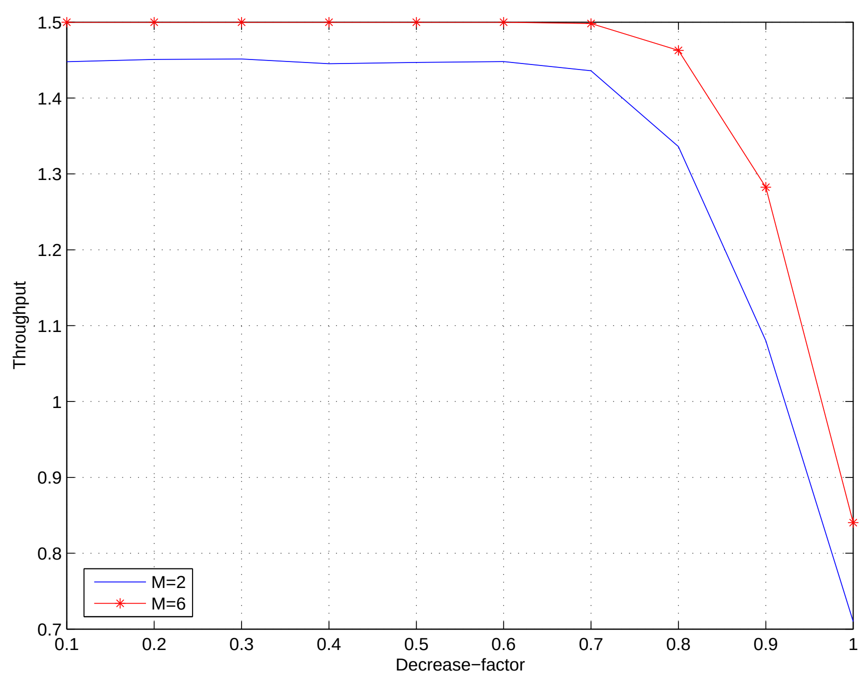

Figure 4.4. Suboptimal CAC policy based on ROP-I: system throughput.

Table 4.4. Suboptimal CAC policy based on ROP-I: blocking probability.

Decrease-factor α_{dec}	0.5	0.6	0.7	0.8	0.9	1.0
P_b^1 $(M=2)$	0.0302	0.0304	0.0292	0.0273	0.0266	0.0262
P_b^2 $(M=2)$	0.0459	0.0428	0.0446	0.0424	0.0427	0.0409
P_b^1 $(M=6)$	0	0	0	1.8×10^{-4}	0	0
P_b^2 $(M=6)$	0	0	0	0	0	0

4.7.3 CAC policy based on ROP-II algorithm

We now investigate the performance for suboptimal CAC policy based on ROP-II. With ROP-II, the power control feasibility condition is derived based on the increased target SIRs, $\alpha_{inc}\gamma_j$, where $j = 1,2$, and as a result, both the network layer performance and the outage probability are affected by the increase-factor α_{inc}.

The blocking probability, the average outage probability and the system throughput are numerically presented in Table 4.5 and Figures 4.5-4.6, respectively, as a function of α_{inc}, in which buffer sizes are set to $B_1 = B_2 = 0$. It is observed that as α_{inc} is increased, outage probability decreases significantly. Although there is a degradation in blocking probability, the overall system throughput can be improved. For example, with $\alpha_{inc} = 1.2$ and $M = 2$, the outage probability can be reduced from 0.5023 to 0.1728, i.e., reduced by 65.6%, and the system throughput can be improved from 0.71 to 1.11, i.e., increased by 56.3%. Although the blocking probability is increased from 0.0389 to 0.0957, this degradation can be reduced by increasing the number of antennas. For example, with the same ROP parameter $\alpha_{inc} = 1.2$, increasing the number of antennas to $M = 6$, the blocking probability is only increased from 0 to 6.1×10^{-5}, which can be neglected.

For a given number of antennas, an appropriately chosen α_{inc} can guarantee a desired

Table 4.5. Suboptimal CAC policy based on ROP-II: blocking probability.

Increase-factor α_{inc}	1	1.05	1.1	1.15	1.2
P_b $(M=2)$	3.9×10^{-2}	5.1×10^{-2}	5.1×10^{-2}	7.6×10^{-2}	9.6×10^{-2}
P_b $(M=4)$	0	0	6×10^{-4}	1.6×10^{-3}	1.9×10^{-3}
P_b $(M=6)$	0	2.7×10^{-4}	9.0×10^{-5}	0	6.1×10^{-5}

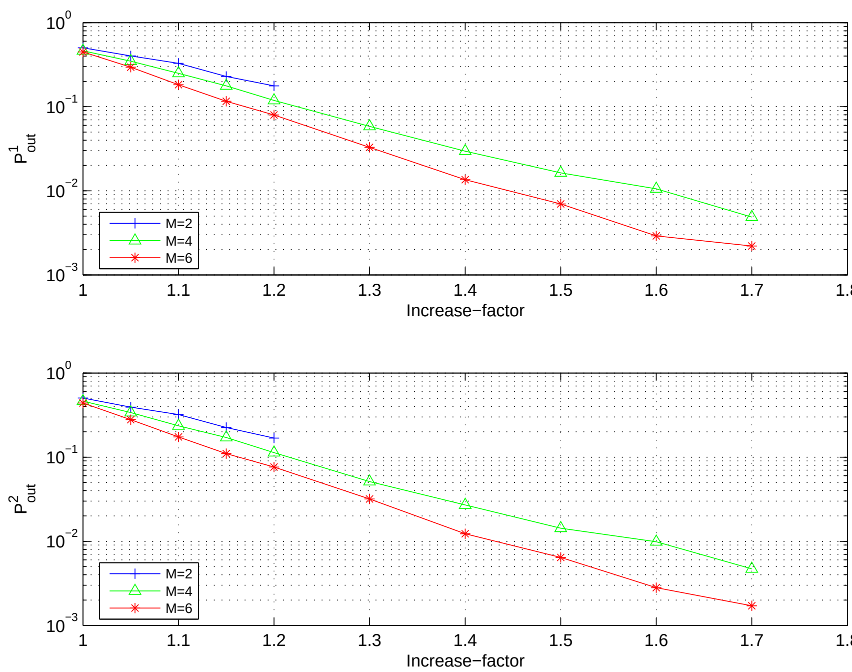

Figure 4.5. Suboptimal CAC policy based on ROP-II: outage probability.

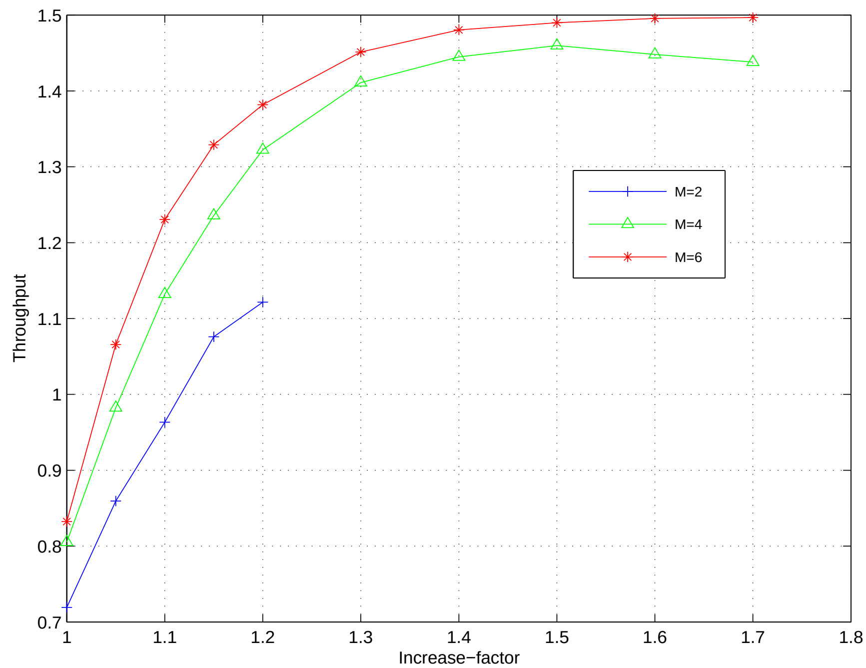

Figure 4.6. Suboptimal CAC policy based on ROP-II: system throughput.

outage probability constraint while minimizing the degradation in the network layer. For example, with $M = 4$, $\rho_{av}^1 = \rho_{av}^2 = 0.05$, desired interval $\bar{\rho}_1 = \bar{\rho}_2 = 0.01$, and adjustment step $\Delta = 0.1$, by following the procedure shown in Figure 4.2, we can obtain the ROP parameter $\alpha_{inc} = 1.3$ which introduces least performance degradation in the network layer while satisfying the outage probability constraints.

4.7.4 Comparison between single and multiple antenna systems

In the following, we demonstrate that the proposed suboptimal CAC policy for beamforming systems can achieve a significant performance gain compared with single antenna systems.

As shown in [24], single antenna systems encounter infeasible buffer problem easily, i.e., for some buffer configurations, the system parameters cannot meet all the constraints of

Table 4.6. Single antenna system: analytical and simulation blocking probabilities and connection delays when SMDP-based CAC is employed.

$[B_1, B_2]$	P_b-LP	P_b-sim	n_q-LP	n_q-sim	P_{out}-sim	$Throughput$-sim
[0,0]	0.2846	0.2828	0	0	0	1.1023
[0,1]	0.2562	0.2507	0.1807	0.1828	0	1.1152
[1,0]	0.2571	0.2602	0.0903	0.0989	0	1.1717
[1,1]	0.2242	0.2289	0.2834	0.3007	0	1.1774

Table 4.7. Two antenna system: analytical and simulation blocking probabilities and connection delays when SMDP-based CAC is employed.

$[B_1, B_2]$	P_b-LP	P_b-sim	n_q-LP	n_q-sim	P_{out}-sim	$Throughput$-sim
[0,0]	0.0364	0.0413	0	0	0.0862	1.3165
[0,1]	0.0243	0.0262	0.0380	0.0471	0.0863	1.3329
[1,0]	0.0280	0.0307	0.0320	0.0408	0.0877	1.3296
[1,1]	0.0173	0.0148	0.0416	0.0474	0.0824	1.3558

the SMDP. To give a quantitative comparison between single and multiple antenna systems, we next relax the blocking probability constraints to $[0.4, 0.4]$. A beamforming system with ROP-I algorithm and decrease-factor 0.8 is employed.

The analytical and simulation results are depicted in Tables 4.6 and 4.7 for single-antenna and two-antenna systems, respectively, where n_q-LP and n_q-sim denote the analytical and simulation results for the average queue length, respectively, P_b-LP and P_b-sim denote the analytical and simulation results for the blocking probability, respectively, and P_{out}-sim denotes the simulation results for the average outage probability. It is observed that the simulation results are very close to the analytical results, and compared with single antenna systems, employing beamforming with only two antennas at the BS can dramatically reduce the blocking probability and connection delay. For example, when no buffering is employed, the blocking probability is 0.2846 for a single antenna system, while this value is decreased to 0.0364 for the case of two antennas.

It is important to note that if the blocking probability constraints are set to $[0.1, 0.1]$, the buffer configurations in Table 4.6 are all infeasible, i.e., for these buffer configurations, the system parameters cannot meet all the constraints of the SMDP. Therefore, extra computation and time are needed to search for a feasible buffer configuration. When two antennas are employed at the BS, a buffer of very small size, or even no buffering at all, leads to the satisfaction of all QoS requirements. Therefore, employing antennas at the BS simplifies the search procedure for a feasible buffer configuration, i.e., reduces the complexity of a SMDP-based policy.

4.7.5 Comparison between suboptimal and optimal CAC policies

We now compare the proposed suboptimal CAC policy with the optimal CAC policy discussed in Chapter 3. Consider a system in which $\rho_w = 0.5$, $B_1 = B_2 = 0$ and $p_v = 1$.

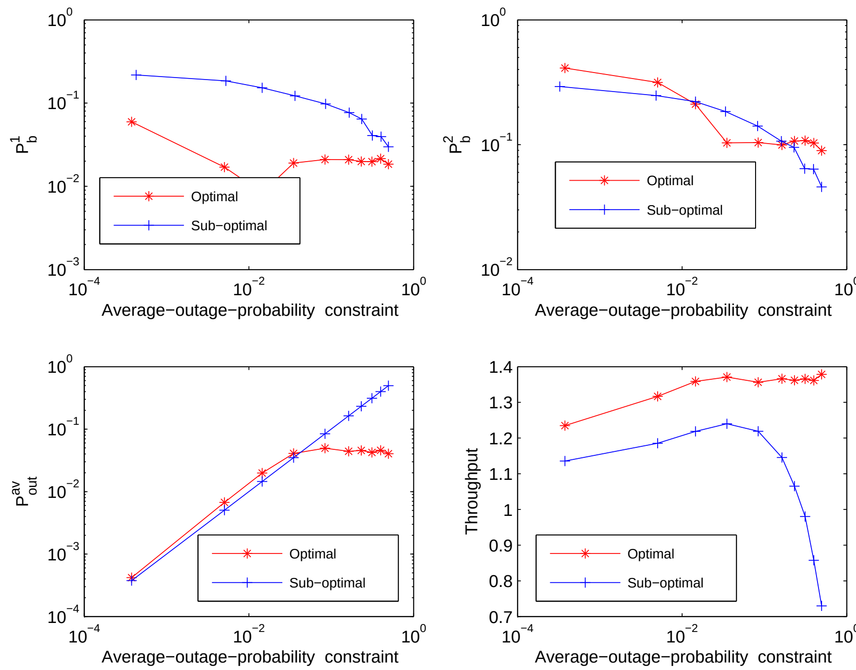

Figure 4.7. Comparison between the optimal and suboptimal CAC policies with $p_v = 1$.

The relaxed WSOP constraint is used to limit the state space, and the link reliability is ensured by the AOP constraint. Figure 4.7 compares the blocking probability, the average-outage-probability and the system throughput for suboptimal and optimal CAC policies for a two-antenna system, in which P_b^j and P_{out}^{av} denote the achieved blocking probability for class j and the average-outage-probability, respectively.

It is observed that for a given average-outage-probability constraint, the proposed optimal CAC policy can achieve a performance gain in terms of lowered blocking probability and improved system throughput. For example, with an AOP constraint of 0.035, compared with the suboptimal CAC policy, the proposed optimal CAC policy can reduce the blocking probability 60%, from 0.15 to 0.06, and increase the throughput 10%, from 1.24 to 1.37 calls/second. With an increased AOP constraint, the throughput gain becomes even larger.

From Figure 4.7, we note that since the objective is to maximize the overall system throughput subject to QoS constraints, it is possible that blocking probability for one class may not decrease monotonically with the average outage probability constraint.

4.7.6 Comparison between proposed and existing CAC policies

As discussed in Chapter 3, complete-sharing (CS)-based CAC policy is employed widely due to its simplicity. For a simple CS-based CAC policy, when a call arrives, the power control feasibility condition in (4.14) or (4.15) is evaluated by incorporating information of this newly arrived call. If this feasibility condition holds, the call is accepted. Otherwise, the call is stored in a buffer, or blocked if the buffer is full. The shortcoming of this CAC policy is that QoS requirements in the network layer are ignored. In this section, we illustrate the difference between our proposed CAC policy and a CS-based CAC policy.

Without loss of generality, we now restrict the blocking probability constraints for both classes to 0.02 , and all the other parameters remain unchanged. The results for a sub-optimal SMDP-based CAC policy and a CS-based CAC policy are shown in Table 4.8, in which two antennas are employed at the BS, and ROP-I algorithm with decrease-factor of 0.8 is employed. In Table 4.8, P_b^j denotes the blocking probability for class j packets, where $j = 1, 2$, and P_b denotes the overall blocking probability. It is observed that for a CS-based CAC policy, the blocking probability constraint cannot be guaranteed. For example, when the buffer size is $[1, 2]$, the blocking probability for class 1 packets is 0.0213, which exceeds its constraint 0.02. When the buffer size is $[2, 1]$, the blocking probability for class 2 packets is 0.0229, which exceeds its blocking probability constraint 0.02. However, for the same buffer sizes, SMDP-based CAC policy can always guarantee blocking probability constraints for both classes. We also observe that compared with the CS-based policy, the proposed SMDP-based CAC policy achieves a lower blocking-probability.

Table 4.8. Comparison between SMDP-based CAC and CS-based CAC policies.

$[B_1, B_2]$	CS: P_b^1	CS: P_b^2	CS: P_b	SMDP: P_b^1	SMDP: P_b^2	SMDP: P_b
[1,2]	0.0213	0.0080	0.0147	0.0133	0.0104	0.0118
[2,1]	0.0108	0.0229	0.0169	0.0118	0.0157	0.0137

4.8 Conclusions

In this chapter, we propose a low-complexity suboptimal call admission control (CAC) policy, which is based on an approximated power control feasibility condition, represented by a linear function. This approximation, however, increases outage probability in the physical layer. We propose two simple ROP algorithms to mitigate the outage probability. A cross-layer CAC policy based on the approximate PCFC and ROP algorithms is then proposed which can guarantee both physical and network layer QoS requirements. Compared with the optimal CAC policy, the suboptimal CAC policy can dramatically reduce the complexity, at a cost of degraded performance.

In this chapter and the previous chapter, we have proposed optimal and suboptimal CAC policies for beamforming systems. Compared with the case of single antenna systems, the proposed CAC policies can achieve a significant performance gain in terms of blocking probability, connection delay and system throughput, and as a result provides a solution to the capacity limitation problem for future wireless networks.

Chapter 5

Packet Admission Control Policies for Packetized Systems with ARQ

5.1 Introduction

The previous chapters focus on a circuit-switched network in which a user occupies its allocated channel during the whole connection. In addition to the circuit-switched services that have dominated earlier generation networks (e.g., voice telephony), there are now a significant number of packet-switched services [77]. Future wireless networks are expected to provide high-rate multimedia services based on packet-switched technology. In this chapter, we extend the previous research to a packet-switched network, and investigate packet-level admission control (AC) policies. To further improve the capacity, a truncated ARQ scheme is employed. Truncated ARQ is an error-control protocol which retransmits an error packet until correctly received or a maximum number of retransmissions is reached.

In the current third generation (3G) system, the application of more efficient methods for packet data transmission such as high speed uplink packet access (HSUPA) has become more important [14]. In HSUPA, a threshold-based call admission control (CAC) policy is employed, which admits a user request if the load reported is below the CAC threshold. Although the CAC decision can be improved upon by taking advantage of resource allocation information [14], and it is simple to implement, it is well known that the threshold-based CAC policy cannot satisfy QoS requirements in the upper layer [73]. In this chapter, we propose a packet-level AC policy with guaranteed QoS requirements in both physical and packet levels.

Similar to call level admission control, for a packet-switched connectionless service, a packet admission control policy decides if an incoming packet can access to the network while still maintaining the quality-of-service (QoS) requirements. In a practical network, blocking a packet instead of blocking the whole connection can be more spectrally efficient, especially for bursty traffic. The proposed AC policy can be derived offline and then stored in a lookup table. Whenever an arrival or departure occurs, an optimal action can be obtained by table lookup, resulting in low enough complexity for admission control at the packet level.

As discussed in the previous two chapters, we have proposed two approaches to ensure the physical layer outage probability constraints: an exact approach and an approximate approach. The exact approach, which is discussed in Chapter 3, depends on the outage probability evaluation for each possible system state, while the approximate approach, which is discussed in Chapter 4, is based on an approximate power control feasibility condition and a ROP algorithm. In this chapter, we use the exact approach to derive the packet-level AC policy. The AC policy in this chapter can be straightforwardly extended to an approximate approach analogous to Chapter 4. Details can be found in [72].

The rest of this chapter is organized as follows. The signal model and problem formulation are presented in Sections 5.2 and 5.3, respectively. Section 5.4 investigates the physical layer performance and provides an analytical expression for outage probability.

optimal and suboptimal AC policies are proposed in Section 5.5. Numerical results are presented in Section 5.6.

5.2 Signal Model

5.2.1 Traffic model

We consider a single-cell CDMA system which supports J classes of packets, characterized by different QoS requirements, where $j = 1, .., J$. Requests for packet access of class j are assumed to be Poisson distributed, with arrival rates λ_j, $j = 1, .., J$.

The admission control (AC) is performed at the BS. An AC policy is derived offline, and stored in a lookup table. When a packet is generated at the mobile station (MS), the MS sends an access request to the BS. In this request, the class of this packet is indicated. After receiving the request, the BS makes a decision, which is then sent back to the MS, on whether the incoming packet should be either accepted, be queued in the buffer, or blocked. Similarly, whenever a packet departs, the BS decides whether the packet in the queue can be served (transmitted).

Once a packet is accepted by the AC policy, it means that this packet is allowed to be transmitted to the BS immediately. Its first transmission round will be performed, and then the receiver will send back an acknowledgement (ACK) signal to the transmitter. A positive ACK indicates that the packet is correctly received while a negative ACK indicates an incorrect transmission. A packet departure means that the packet has finished transmission and is not allowed to be transmitted.

If a positive ACK is received or the maximum number of retransmissions, denoted by L_j, is reached, the packet releases the server and departs. Otherwise the packet will be retransmitted. Therefore, the service time of a packet can comprise at most $L_j + 1$

transmission rounds. Each transmission round includes the actual transmission time of the packet and the waiting time of an ACK signal (positive or negative). The duration of a transmission round for a packet in class j is assumed to have an exponential distribution with mean duration $\frac{1}{\mu_j}, j = 1, .., J$. However, in this chapter, a sub-optimal solution is also provided for a generally distributed duration.

If the packet is not accepted by the AC policy, it will be stored in a queueing buffer provided that the buffer is not full. If the buffer is full, the packet will be dropped, leading to a packet loss. Each class of packets shares a common queueing buffer, and B_j denotes the buffer size of class j.

The QoS requirements in the packet level can be represented by the target packet loss probability and packet access delay, denoted by Ψ_j and D_j for class j, respectively. Packet loss probability is defined as the probability that a incoming packet is dropped, while packet access delay is just the waiting time in the queue, which can be equivalently represented by the average queue length. For each class j, where $j = 1, .., J$, there are n_s^j packets physically present in the system, which have the same target packet-error-rate (PER), packet loss probability, and packet access delay constraints.

We note that there are two types of buffers in the system: queue buffers and server buffers. The queue buffer accommodates queued incoming packets, while the server buffer accommodates transmitted packets in the server in case any packet in the server requires retransmission. For simplicity, we assume that the size of the server buffer is large enough such that all the packets in the server can be stored. In the following, the generic term *buffer* refers to the queueing buffer.

5.2.2 Signal model at the physical layer

We consider an uplink CDMA beamforming system, in which M antennas are employed at the BS and a single antenna is employed for each mobile station. There are K accepted packets in the system, and a channel with slow fading is assumed.

As shown in the previous chapters, the received signal-to-interference ratio (SIR) for a desired packet k can be written as

$$SIR_k = \frac{W}{R_k} \frac{p_k \phi_{kk}^2}{\sum_{i \neq k} p_i \phi_{ik}^2 + \eta_0 W}. \tag{5.1}$$

In a wireless communication network, we must allow for outage, defined as the probability that a target SIR, or equivalently, a target packet-error-rate (PER), cannot be satisfied. The QoS requirement in the physical layer can be represented by a target outage probability.

5.3 Problem Formulation

The average system throughput, defined as the number of correctly received packets per second, can be evaluated by [54]

$$\text{Throughput} = \sum_j (1 - P_L^j)(1 - P_{out}^{av})(1 - PER_j)\lambda_j \tag{5.2}$$

where P_L^j, P_{out}^{av} and PER_j denote the packet loss probability, the average outage probability, and the packet-error-rate for class j, respectively.

In this chapter, we aim to derive an optimal AC policy which incorporates the benefits provided by multiple antennas and ARQ schemes. The objective is to maximize the overall system throughput given in (5.2), while simultaneously guaranteeing QoS requirements in terms of outage probability, packet loss probability and packet access delay.

With ARQ, the retransmissions improve the physical layer performance, while at the same time it may also degrade the network layer performance due to the increased duration of each packet. The cross-layer AC design should consider both positive and negative impacts of ARQ. Furthermore, with retransmissions, the duration of a packet is not exponentially distributed, which violates the Markovian property of the previously formulated SMDP. In order to exploit the benefits of ARQ while still maintaining the Markov property, it is necessary to formulate a novel SMDP.

5.4 Outage Probability in the Presence of ARQ

The system state **s** can be represented by

$$\mathbf{s} = [\underbrace{n_q^1, k^{1,1}, .., k^{1,L_1+1}}, ..., \underbrace{n_q^J, k^{J,1}, ..., k^{J,L_J+1}}] \tag{5.3}$$

where $k^{j,i}$ denotes the number of active packets in class j which is under the i^{th} transmission round, or equivalently, under the $(i-1)^{th}$ retransmission; n_q^j denotes the queue length, i.e., the number of packets in the queue buffer of class j. The number of class j packets can be obtained as $n_s^j = \sum_{i=1}^{L_j+1} k^{j,i}$, and the total number in the system is obtained as $K = \sum_{j=1}^{J} n_s^j$.

As mentioned before, the QoS requirement in the physical layer can be represented by target outage probability, defined as the probability that a target packet-error-rate (PER), or equivalently a target SIR, cannot be satisfied. In the following, we first derive the target SIR corresponding to a target PER. Then the outage probability is discussed.

5.4.1 Derivation of target SIR

We define two kinds of PERs: overall PER and instantaneous PER. Overall PER, denoted by $PER_{overall}^j$, is defined as the probability that a class j packet is incorrectly received after its maximum number of retransmissions is reached, i.e., an error occurs in each of the

$L_j + 1$ transmission rounds, where L_j denotes the maximum number of retransmissions. Instantaneous PER, denoted as $PER_{in}^{j}(l)$, is defined as the probability that an error occurs in a single transmission round l of a class j packet.

Under the assumption that each retransmission round is independent from the others, the achieved overall PER can be expressed as [54]

$$\begin{aligned} PER_{overall}^{j} &= \prod_{l=1}^{L_j+1} PER_{in}^{j}(l) \\ &\leq \rho_j \end{aligned} \tag{5.4}$$

where ρ_j denotes the target PER for class j.

To ensure the above inequality, we require

$$PER_{in}^{j}(l) \leq (\rho_j)^{\frac{1}{L_j+1}}. \tag{5.5}$$

In general, given the above target instantaneous PER, it is not an easy task to derive the target signal-to-interference ratio (SIR). Fortunately, there are approximations in the literature. The instantaneous PER for packet length N_p can be approximately expressed in terms of instantaneous SIR as [54]

$$PER_{in}^{j}(l) = a \exp(-g \times SIR_j) \tag{5.6}$$

for $SIR_j \geq \gamma_0$ dB, where SIR_j is the achieved SIR, given in (5.1); a, g, and γ_0 are constants depending on the chosen modulation and coding scheme. In the above expression, the interference is assumed to be additive white Gaussian noise, which is reasonable in a system with a large number of interferers.

Combining (5.5) and (5.6), we have

$$SIR_j \geq \frac{1}{g}[\ln a - \ln((\rho_j)^{\frac{1}{L_j+1}})] \tag{5.7}$$

where $\ln(.)$ denotes natural logarithm, and the right hand side of the above inequality is the target SIR, i.e.,

$$\gamma_j = \frac{1}{g}[\ln a - \ln((\rho_j)^{\frac{1}{L_j+1}})]. \tag{5.8}$$

5.4.2 Outage probability

As shown in (3.8), given a target SIR in (5.8), the outage probability for a system state **s** by including the impact of ARQ can be represented by

$$P_{out}(\mathbf{s}) \quad = \quad \text{Prob}\{\upsilon(QF) \geq 1\} \tag{5.9}$$

where Prob$\{A\}$ denotes the probability of event A, $\upsilon(.)$ denotes the maximum eigenvalue, Q is a K-dimensional diagonal matrix with the i^{th} non-zero element as $\frac{\gamma_i R_i}{W}$, $i = 1, .., K$, and F is a K by K matrix in which the element at the i^{th} row and the j^{th} column can be expressed as $F_{ij} = \frac{\phi_{ij}^2}{\phi_{ii}^2}$ for $i \neq j$, and $F_{ij} = 0$ for $i = j$.

With this state outage probability, the worst-state outage probability, denoted by P_{out}^w, and the average outage probability, denoted by P_{out}^{av}, can be expressed as follows

$$\begin{aligned} P_{out}^w &= \max_{\mathbf{s} \in S} P_{out}(\mathbf{s}) \qquad (5.10) \\ &\leq \rho_w \\ P_{out}^{av} &= \sum_{\mathbf{s} \in S} P_{\mathbf{s}} P_{out}(\mathbf{s}) \qquad (5.11) \\ &\leq \rho_{av} \end{aligned}$$

where ρ_w and ρ_{av} denote the WSOP and AOP constraints, respectively; $P_{\mathbf{s}}$ denotes the steady-state probability that the system is in state **s** and S represents the set of all feasible system states. The above outage probability constraints can be guaranteed by formulating the state space S and adding SMDP constraints, which will be discussed in Section 5.5.

5.5 Cross-layer AC Policies

In the previous section, we have analyzed the outage probability in the presence of ARQ. In the following, we discuss how to derive AC policies in the packet level. An optimal semi-Markov decision process (SMDP)-based AC policy, as well as a low-complexity generalized semi-Markov process (GSMP)-based AC policy are discussed.

5.5.1 SMDP-based AC policy

Traditionally, the decision epoches are chosen as the time instants that a session arrives or departs. For a packet level AC policy, the decision epoches can be chosen to be the time instants that a packet arrives or departs and system state is represented by the number of simultaneously transmitted packets and the number of packets waiting in the queueing buffer. However, with this decision epoch and system state, the duration of each packet may include several transmission rounds due to ARQ retransmissions, and the time duration until next system state may not be exponentially distributed, which violates the Markov properties required by a SMDP.

In the following, we choose a different decision epoch and system state, which represents the retransmission information. Then we formulate a novel SMDP. The decision epoch is chosen as the arrival and departure of each transmission round. Based on these decision epoches, the time duration until the next state remains exponentially distributed. The components of a Markov decision process, such as state space, action space and dynamic statistics, are modified accordingly to represent the characteristics of different transmission rounds. The formulation of this SMDP as well as its LP solution are now described.

State space

Class j packets are divided into L_j+1 subclasses, in which the state of the i^{th} subclass can be represented by the number of packets which are under the i^{th} round transmission, i.e., the $(i-1)^{th}$ retransmission, where $i = 1,..,L_j+1$.

In admission problems, the discrete-value (finite) state at time t, $\mathbf{s}(t)$, can be written as

$$\mathbf{s}(t) = [\underbrace{n_q^1(t), k^{1,1}(t), .., k^{1,L_1+1}(t)}, ..., \underbrace{n_q^J(t), k^{J,1}(t), ..., k^{J,L_J+1}(t)}]$$

where $k^{j,i}(t)$ represents the number of packets in class j and subclass i served in the system, and $n_q^j(t)$ denotes the number of packets in the queueing buffer of class j. Since the arrival and departure of packets are random, $\{\mathbf{s}(t), t > 0\}$ represents a finite state stochastic process [24]. From here on, we will drop the time index.

The state space S is comprised of any state vector $\mathbf{s}$, in which worst-state outage probability requirements can be satisfied,

$$S = \{\mathbf{s}, \text{ where } P_{out}(\mathbf{s}) \le \delta_w, \text{and } n_q^j \le B_j\}.$$

where $P_{out}(\mathbf{s})$ is given in (5.9), and δ_w, n_q^j and B_j denote the target WSOP, the queue length and the buffer size, respectively.

The formulation of the above state space can be summarized as follows:

- Compute the upper bound of the number of accepted packets for each class and each transmission round. We note that this maximum number only depends on the corresponding class, and is same for each transmission round. Without loss of generality, we derive this number for transmission round 1 in class j, denoted by M_j^{max}. The search procedure for M_j^{max} is presented in Figure 5.1;

- An enlarged state space, denoted by $\bar{S}$, can be formulated as

  $\bar{S} =$

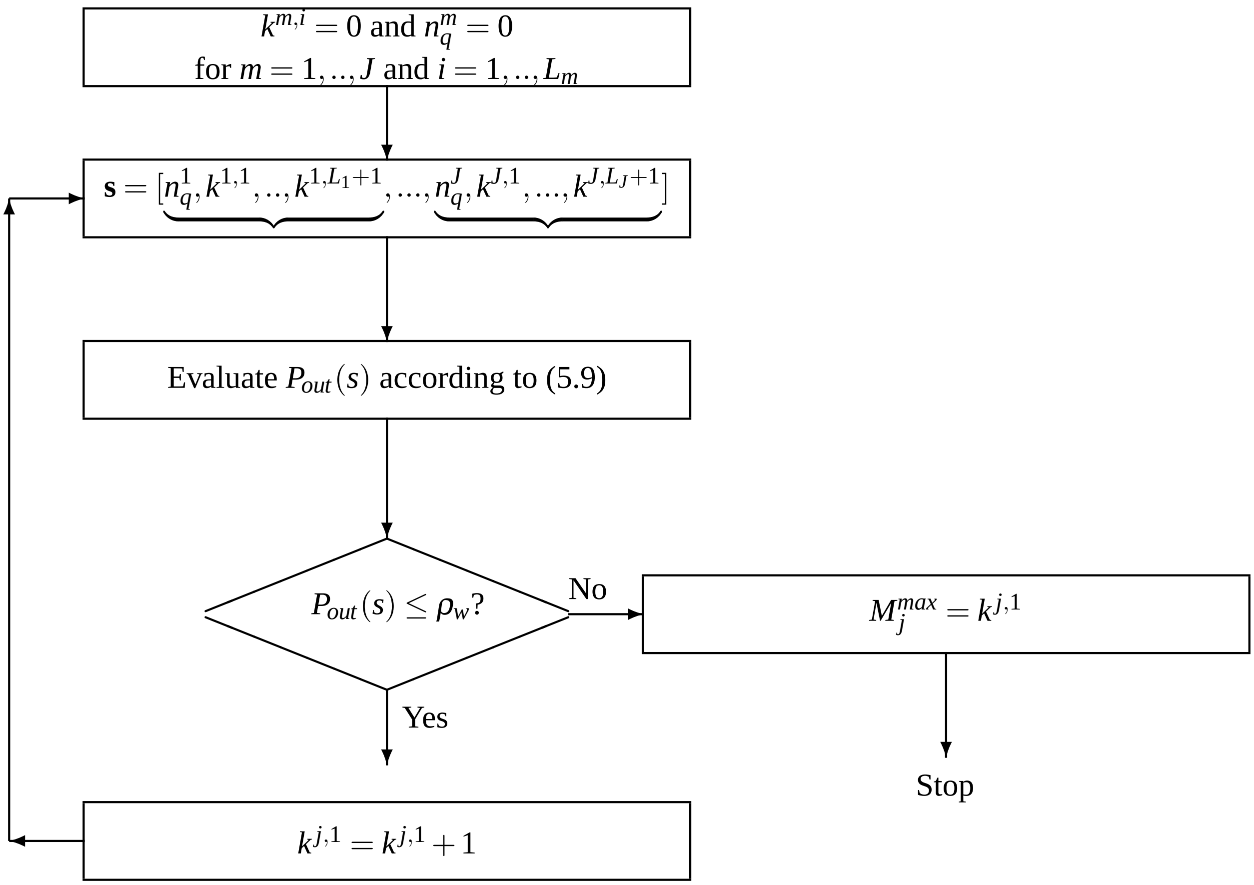

Figure 5.1. Search procedure for M_j^{max}.

$$\left\{ \mathbf{s} = [\underbrace{n_q^1, k^{1,1}, .., k^{1,L_1+1}}, ..., \underbrace{n_q^J, k^{J,1}, ..., k^{J,L_J+1}}] : k_{j,i} \leq M_j^{max}, n_q^j \leq B_j,\ j = 1, .., J \right\};$$

- The above $\bar{S}$ can be truncated to the desired state space S as follows:
  - Initialize $S = \{\}$;
  - Evaluate $P_{out}(s)$ for $\mathbf{s} \in \bar{S}$ according to (5.9);
  - If $P_{out}(s) \leq \rho_w$, then $S = S + \{s\}$.
- We remark that in the above step, it is unnecessary to evaluate each system state in $\bar{S}$, since if $\mathbf{s} \in S$, then all $\mathbf{s}' \in \bar{S}$ such that $\mathbf{s}' \leq \mathbf{s}$ are also in S. Similarly, if $\mathbf{s}$ is not in S, then all $\mathbf{s}' \in \bar{S}$ such that $\mathbf{s}' \geq \mathbf{s}$ are also not in S.

Action space

At each state $\mathbf{s}$, an action is chosen that determines how the admission control will perform at the next decision moment [24]. In general, an action, denoted as $\mathbf{a}$, can be defined as a vector of dimension $\sum_{j=1}^{J} L_j + 2J$

$$\mathbf{a} = [\underbrace{a_1, d_1^1, ..., d_1^{L_1+1}}.....,\underbrace{a_J, d_J^1, ..., d_J^{L_J+1}}]$$

where a_j denotes the action for class j if an arrival occurs, $j = 1, .., J$. If $a_j = 0$, the new arrival is placed in the buffer provided that the buffer is not full, or is blocked if the buffer is full; if $a_j = 1$, the arrival is admitted as an active packet, and the number of servers of class j is incremented by one.

The quantity d_j^i, where $1 \leq i \leq L_j$, denotes the action for class j packet if the i^{th} transmission round is finished, and is received correctly. If $d_j^i = 0$, where $1 \leq i \leq L_j$, $k^{j,i}$ is decremented by one, and no packets that are queued in the buffer are made active; if $d_j^i = 1$, the number of servers is maintained by admitting a packet at the buffer as an active packet.

We note that whenever a class j packet is not received correctly and the maximum number of retransmissions is not reached, it is automatically retransmitted.

The quantity $d_j^{L_j+1}$ denotes the action for class j packet if a connection has finished its $(L_j+1)^{th}$ transmission round. If $d_j^{L_j+1}=0$, no packets that are queued in the buffer are made active, and k^{j,L_j+1} is decremented by one; if $d_j^{L_j+1}=1$, the number of servers is maintained by admitting a packet at the buffer as an active packet.

The admissible action space for state $\mathbf{s}$, denoted by $A_{\mathbf{s}}$, can be defined the set of all feasible actions. A feasible action ensures that after taking this action, the next transition state is still in space S [24].

State dynamics $p_{\mathbf{sy}}(\mathbf{a})$ and $\tau_{\mathbf{s}}(\mathbf{a})$

The state dynamics of a SMDP are completely specified by stating the transition probabilities of the embedded chain $p_{\mathbf{sy}}(\mathbf{a})$ and the expected holding time $\tau_{\mathbf{s}}(\mathbf{a})$: $p_{\mathbf{sy}}(\mathbf{a})$ is defined as the probability that the state at the next decision epoch is $\mathbf{y}$ if action $\mathbf{a}$ is selected at the current state $\mathbf{s}$, while $\tau_{\mathbf{s}}(\mathbf{a})$ is the expected time until the next decision epoch after action $\mathbf{a}$ is chosen in the present state $\mathbf{s}$ [24].

Derivations of $\tau_{\mathbf{s}}(\mathbf{a})$ and $p_{\mathbf{sy}}(\mathbf{a})$ rely on the statistical properties of arrival and departure processes [24]. Since the arrival and departure processes are both Poisson distributed and mutually independent, it follows that the cumulative process is also Poisson, and the cumulative event rate is the sum of the rates for all constituent processes [24]. Therefore, the expected sojourn time, $\tau_{\mathbf{s}}(\mathbf{a})$, can be obtained as the inverse of the event rate,

$$\begin{aligned}
&\tau_{\mathbf{s}}(\mathbf{a})^{-1} = \\
&\quad \lambda_1 a_1 + \lambda_1(1-a_1)\delta(B_1 - n_q^1) + \sum_{i=1}^{L_1+1} \mu_1(k^{1,i}) + \ldots\ldots + \\
&\quad \lambda_J a_J + \lambda_J(1-a_J)\delta(B_J - n_q^J) + \sum_{i=1}^{L_J+1} \mu_J(k^{J,i}). \qquad (5.12)
\end{aligned}$$

To derive the transition probabilities, we employ the decomposition property of a Poisson process, which states that an event of a certain type occurs with a probability equal to the ratio between the rate of that particular type of event and the total cumulative event rate $\frac{1}{\tau_{\mathbf{s}}(\mathbf{a})}$ [24]. Transition probability $p_{\mathbf{sy}}(\mathbf{a})$ is shown in Table 5.1, where ρ_{in}^{j} denotes the target instantaneous packet-error-rate for class j packets. The set of vectors $\left\{\mathbf{q}^j, \mathbf{b}^j, \mathbf{c}_i^j, \mathbf{r}_i^j, \mathbf{e}_i^j, \mathbf{f}^j, \mathbf{g}^j\right\}$ represents the possible state transitions from current state $\mathbf{s}$. Each vector in this set has a dimension of $\sum_{j=1}^{J} L_j + 2J$, and contains only zeros except for several positions. The non-zero positions of this set of vectors, as well as the possible state transitions represented by these vectors, are specified in Tables 5.2 and 5.3, respectively, in which we note that $\sum_{t=1}^{j-1} L_t = 0$ for $j = 1$.

Table 5.1. Expression of transition probability $p_{\mathbf{sy}}$.

y	$p_{\mathbf{sy}}(\mathbf{a})$
$y = \mathbf{s} + \mathbf{q}^j$	$\lambda_j a_j \tau_{\mathbf{s}}(\mathbf{a})$
$y = \mathbf{s} + \mathbf{b}^j$	$\lambda_j (1 - a_j) \delta(B_j - n_q^j) \tau_{\mathbf{s}}(\mathbf{a})$
$y = \mathbf{s} + \mathbf{c}_i^j$, where $1 \le i \le L_j$	$(1 - \rho_{in}^j)[\mu_j k^{j,i} (1 - d_j^i) \tau_{\mathbf{s}}(\mathbf{a})] +$ $(1 - \rho_{in}^j)[\mu_j k^{j,i} d_j^i (1 - \delta(n_q^j)) \tau_{\mathbf{s}}(\mathbf{a})]$
$y = \mathbf{s} + \mathbf{r}_i^j$, where $1 \le i \le L_j$	$(1 - \rho_{in}^j) \mu_j k^{j,i} d_j^i \delta(n_q^j) \tau_{\mathbf{s}}(\mathbf{a})$
$y = \mathbf{s} + \mathbf{e}_i^j$, where $1 \le i \le L_j$	$\rho_{in}^j \mu_j k^{j,i} \tau_{\mathbf{s}}(\mathbf{a})$
$y = \mathbf{s} + \mathbf{f}^j$	$\mu_j k^{j,L_j+1} d_j^{L_j+1} \delta(n_q^j) \tau_{\mathbf{s}}(\mathbf{a})$
$y = \mathbf{s} + \mathbf{g}^j$	$\mu_j k^{j,L_j+1} (1 - d_j^{L_j+1}) \tau_{\mathbf{s}}(\mathbf{a}) +$ $\mu_j k^{j,L_j+1} d_j^{L_j+1} (1 - \delta(n_q^j)) \tau_{\mathbf{s}}(\mathbf{a})$
otherwise	0

Table 5.2. Definition of vectors in Table 5.1: each vector defined in this table has a dimension of $\sum_{j=1}^{J} L_j + 2J$, which contains only zeros except for the specified positions.

Vector	Non-zero positions
$\mathbf{q}^j$	Position $2(j-1)+\sum_{t=1}^{j-1} L_t + 2$ contains a 1.
$\mathbf{b}^j$	Position $2(j-1)+\sum_{t=1}^{j-1} L_t + 1$ contains a 1.
$\mathbf{c}_i^j$, $i=1,..,L_j$	Position $2(j-1)+\sum_{t=1}^{j-1} L_t + i + 1$ contains a -1.
$\mathbf{r}_i^j$, $i=1,..,L_j$	Position $2(j-1)+\sum_{t=1}^{j-1} L_t + 1$ contains a -1, position $2(j-1)+\sum_{t=1}^{j-1} L_t + i + 1$ contains a -1, and position $2(j-1)+\sum_{t=1}^{j-1} L_t + 2$ contains a 1.
$\mathbf{e}_i^j$, $i=1,..,L_j$	Position $2(j-1)+\sum_{t=1}^{j-1} L_t + i + 1$ contains a -1, and position $2(j-1)+\sum_{t=1}^{j-1} L_t + i + 2$ contains a 1.
$\mathbf{f}^j$	Position $2(j-1)+\sum_{t=1}^{j-1} L_t + L_j + 2$ contains a -1, position $2(j-1)+\sum_{t=1}^{j-1} L_t + 1$ contains a -1, and position $2(j-1)+\sum_{t=1}^{j-1} L_t + 2$ contains a 1.
$\mathbf{g}^j$	Position $2(j-1)+\sum_{t=1}^{j-1} L_t + L_j + 2$ contains a -1.

Policy, performance criterion and expected cost function

A policy R defines a mapping rule from state space S to action space A. In the following, we take the average cost as the performance criterion. For any policy R with an initial state $\mathbf{s}_0$, the average cost can be expressed as [99]

$$J_R(\mathbf{s}_0) = \lim_{t\to\infty} \frac{1}{T} E\left\{\int_0^T c(\mathbf{s}(t), \mathbf{a}(t))dt\right\} \tag{5.13}$$

where $E[.]$ denotes expectation, and $c(\mathbf{s}(t), \mathbf{a}(t))$ is the expected cost function which represents the expected cost until the next decision epoch when $\mathbf{a}(t)$ is chosen at the current system state $\mathbf{s}(t)$.

When the average cost in (5.13) represents packet loss probability or packet access delay, the corresponding expected cost functions, denoted by $c_l^j(\mathbf{s},\mathbf{a})$ and $c_d^j(\mathbf{s},\mathbf{a})$, respectively, can be derived as [24]

$$\begin{aligned} c_l^j(\mathbf{s},\mathbf{a}) &= (1-a_j)(1-\delta(B_j-n_q^j)) \qquad (5.14)\\ c_d^j(\mathbf{s},\mathbf{a}) &= n_q^j \qquad (5.15) \end{aligned}$$

in which the detailed derivation can be found in [24].

If the average cost in (5.13) represents average-outage-probability, the expected cost function, denoted by $c_{out}(\mathbf{s},\mathbf{a})$, can be written as

$$c_{out}(\mathbf{s},\mathbf{a}) = P_{out}(\mathbf{s}) \qquad (5.16)$$

which is given in (5.9).

If the average cost in (5.13) represents throughput, the expected cost function, denoted by $c_{thr}(\mathbf{s},\mathbf{a})$, can be expressed as

$$c_{thr}(\mathbf{s},\mathbf{a}) = \sum_{j=1}^{J} \lambda_j (1-c_l^j(\mathbf{s},\mathbf{a}))(1-c_{out}(\mathbf{s},\mathbf{a}))(1-PER_{overall}^j). \qquad (5.17)$$

The optimal policy can be chosen according to a certain performance criterion. In the following, we aim to find an optimal policy R^* which maximizes the throughput for any initial state, i.e.,

$$R^* = \arg\max_{R\in\mathbf{R}} \lim_{t\to\infty} \frac{1}{T} E\left\{\int_0^T C_{thr}(\mathbf{s}(t),\mathbf{a}(t))dt\right\}.$$

where $\mathbf{R}$ represents the set of all possible policies.

Deriving an optimal policy by solving the SMDP

An optimal AC policy can be obtained by using the decision variables $z_{\mathbf{sa}}, \mathbf{s}\in S, \mathbf{a}\in A_{\mathbf{s}}$, in solving the following linear programming (LP) problem:

$$\max_{z_{\mathbf{sa}}\geq 0,\mathbf{s},\mathbf{a}} \sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}}\sum_{j=1}^{J} c_{thr}(\mathbf{s},\mathbf{a})\tau_{\mathbf{s}}(\mathbf{a})z_{\mathbf{sa}} \qquad (5.18)$$

subject to the set of constraints

$$\begin{aligned}
\sum_{\mathbf{a}\in A_m} z_{ma} - \sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} p_{sm}(\mathbf{a}) z_{\mathbf{sa}} &= 0, m \in S &\quad (5.19)\\
\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} \tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} &= 1 &\quad (5.20)\\
\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} (1-a_j)(1-\zeta(B_j - n_q^j))\tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} &\leq \Psi_j &\quad (5.21)\\
\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} P_{out}(\mathbf{s})\tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} &\leq \rho_{av} &\quad (5.22)\\
\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} n_q^j \tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} &\leq D_j &\quad (5.23)
\end{aligned}$$

where Ψ_j, ρ_{av} and D_j denote the target packet loss probability, target average outage probability and target packet access delay, respectively.

In the above LP formulation, $\tau_{\mathbf{s}}(\mathbf{a})z_{\mathbf{sa}}$ represents the steady-state probability that the system is at state $\mathbf{s}$ and an action $\mathbf{a}$ is chosen. The first constraint is the balance equation, and the second constraint ensures the sum of all the steady-state probabilities to be one. The latter three constraints represent the QoS requirements in terms of packet loss probability, average-outage-probability and packet access delay, respectively. The worst-state-outage-probability constraint, if any, is already included in the state space S.

Since the sample path constraints are included in the above linear programming approach, the optimal policy resulting from the SMDP is a randomized policy [99]: the optimal action $\mathbf{a}^* \in A_{\mathbf{s}}$ for state $\mathbf{s}$, where $A_{\mathbf{s}}$ is the admissible action space, is chosen probabilistically according to the probabilities $z_{\mathbf{sa}}/\sum_{\mathbf{a}\in A_{\mathbf{s}}} z_{\mathbf{sa}}$.

We remark that the above randomized AC policy allows for resources to be more flexibly reserved for potential arriving traffic. The decision variables, $z_{\mathbf{sa}}$, where $\mathbf{s} \in S$ and $\mathbf{a} \in A_x$, act as the long-run fraction of decision epoches at which the system is in state $\mathbf{s}$ and action $\mathbf{a}$. At each state $\mathbf{s}$, there exists a set of feasible actions, and each action induces a different cost $c(\mathbf{s},\mathbf{a})$. The long-run performance can be optimized by appropriately allocating these time fractions, and the allocation leads to a randomized AC policy. When a deterministic policy is desired, a constraint regarding the decision variables $z_{\mathbf{sa}}$ should be imposed into the above optimization problem, in order to ensure that at each state $\mathbf{s}$, there is one and only one non-zero decision variable. It is obvious that the more constraints imposed, the worse the achieved performance becomes. We choose a randomized AC policy in order to achieve long-run optimal performance.

5.5.2 GSMP-based AC policy

In the above, we provide an optimal SMDP formulation. The state space has dimension of $2J+\sum_{j=1}^{J}L_j$ for J classes of traffic. For large J and retransmission number, this leads to a computation problem of excessive size. Also, the above policy is optimal under the assumption that the transmission round has an exponential distribution.

In order to reduce complexity and derive a policy which can be applied to a generally distributed transmission round, we now choose the system state to only represent the number of simultaneously transmitted packets and the number of packets in the queueing buffer for each class, and choose the decision epoch as the time instants that a packet arrives or departs. As we discussed previously, based on the above decision epoch and system state, the time interval until the next state is not exponentially distributed, and the Markovian property is violated. Therefore, we have a generalized semi-Markov process (GSMP). While an optimal solution for this GSMP problem is hard to obtain, a linear programming approach provides a sub-optimal solution [73].

In the formulated GSMP, decision epoches are chosen as the time instants that a packet arrives or departs. The arrival process for class j is assumed to have a Poisson distribution with arrival rate λ_j. The duration of the class j packets may have a general distribution, and the mean duration of each packet can be derived as

$$\begin{aligned} C_j &= \frac{1}{\mu_j} N_{ARQ}^j \\ &= \frac{1}{\mu_j}\left(1+(\rho_j)^{\frac{1}{L_j+1}}+..+(\rho_j)^{\frac{L_j}{L_j+1}}\right) \end{aligned} \tag{5.24}$$

where C_j is the packet duration, N_{ARQ}^j is the average number of transmissions for a class j packet, which is derived in (2.16), and μ_j denotes the departure rate for each transmission round for the class j packets. When deriving (5.24), we approximate the instantaneous PER by its upper bound, given in (5.5).

The state space S is comprised of any state vector $\mathbf{s}$, which satisfies outage requirements,

$$S = \{\mathbf{s} = [n_q^1, n_s^1, ..., n_q^J, n_s^J] :, \text{ where } P_{out}(\mathbf{s}) \leq \delta_w, \text{and } n_q^j \leq B_j\}$$

where n_s^j denotes the number of active packets for class j.

At each decision epoch, an action is chosen as $\mathbf{a} = [a_1, d_1....., a_J, d_J]$, where a_j denotes the action for class j if an arrival occurs, $j = 1,..,J$ and d_j denotes the action for class j packet if a packet in this class departs. The admissible action space for state $\mathbf{s}$, denoted by $A_\mathbf{s}$, can be defined as the set of all feasible actions.

The state dynamics of a SMDP are completely specified by stating the expected holding time $\tau_\mathbf{s}(\mathbf{a})$ and the transition probabilities of the embedded chain $p_{\mathbf{sy}}(\mathbf{a})$, which are given in (3.29) and (3.30).

In summary, the formulation of a GSMP is very similar to the AC problem formulation in the previous chapters and in the literature, e.g., [24] [73], except that the state space and the mean duration of a packet are modified to incorporate the impact of ARQ schemes.

After formulating the AC problem as a GSMP, the AC policy, which minimizes the packet loss probability, can be obtained by using the decision variables $z_{\mathbf{sa}}, \mathbf{s} \in S, \mathbf{a} \in A_\mathbf{s}$ from linear programming which is presented in (5.18).

For the GSMP-based AC policy, the dimension of the state space is reduced from $2J + \sum_{j=1}^{J} L_j$ to $2J$. In a low instantaneous PER region, the GSMP-based solution proposed in the above is very close to the SMDP-based AC policy. Intuitively, when the PER is very low, retransmission occurs only occasionally, and the duration of a packet would be very close to an exponential distribution. In this case, the LP approach would provide a nearly-optimal solution to the above GSMP.

We remark that unlike the SMDP-based AC policy in which the transmission round is assumed to have an exponential distribution, the GSMP-based AC policy discussed in the above can be applied to a system with a generally distributed transmission round.

5.5.3 Complexity

SMDP or GSMP-based AC policies are always calculated offline and stored in a lookup table. Whenever an arrival or departure occurs, an optimal action can be obtained by table lookup using the current system state. This facilitates the implementation of packet-level admission control.

Once system parameters change, an updated policy is required. However, in the system we investigate, the policy only depends on buffer sizes, long-term traffic model and QoS requirements. These parameters are generally constant for the provision of a given profile of offered services. Therefore, a SMDP or GSMP-based policy has a very reasonable computation complexity.

5.6 Numerical Examples

In the following examples, a two-class system is considered. The data rates are set to $R_1 =$ 144 kbps and $R_2 =$ 384 kbps. The arrival and departure rates for class 1 and class 2 packets are denoted by $\lambda_1 = 0.5$, $\lambda_2 = 0.25$, $\mu_1 = 0.3$, and $\mu_2 = 0.2$, respectively. The packet loss probability constraints for both classes are set to $\Psi_1 = \Psi_2 = 0.05$, unless specified otherwise, and the packet access delay constraints for class 1 and class 2 packets are set to 1.67 and 3.5 seconds, respectively. The average outage probability constraint is set to $\rho_{av} = 0.05$ and the worst-state outage probability is $\rho_w = 0.5$. Two antennas are employed at the BS, i.e., $M = 2$. The total bandwidth is 3.84 MHz.

For the exact approach employed in this chapter, it is necessary to numerically evaluate the outage probability for each possible system state, which increases the computational complexity. In the following examples, we employ an outage probability approximation to reduce the complexity,

$$P_{out}(\mathbf{s}) = Q\left[\frac{1 - E[\upsilon]}{\sqrt{Var(\upsilon)}}\right] \tag{5.25}$$

where $Q(\alpha) = \frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\infty} e^{-\frac{x^2}{2}}dx$, $E[\upsilon]$ and $Var(\upsilon)$ denote the expectation and variance of random variable $\upsilon(QF)$, which can be obtained as follows:

$$\begin{aligned} E[\upsilon] &= \sum_{j=1}^{J} \frac{1}{W}\gamma_j R_j n_s^j E[F] \\ Var[\upsilon] &= \frac{1}{K}\sum_{j=1}^{J} n_s^j [\frac{1}{W}\gamma_j R_j]^2 Var[F] \end{aligned} \tag{5.26}$$

where $E[F]$ and $Var[F]$ denote the expectation and variance of F_{ij}, which can be evaluated numerically. Table 5.4 presents these numerical values for a uniform circular array, which are derived in [75]. The derivation of the above approximation can be found in Appendix C. We remark that our proposed AC policies can be applied straightforwardly to other outage probability evaluations.

In the following, we investigate the long-run average performance in terms of packet loss probability, packet access delay, average outage probability and overall system throughput as a function of the overall PER constraint, denoted by PER in the figures.

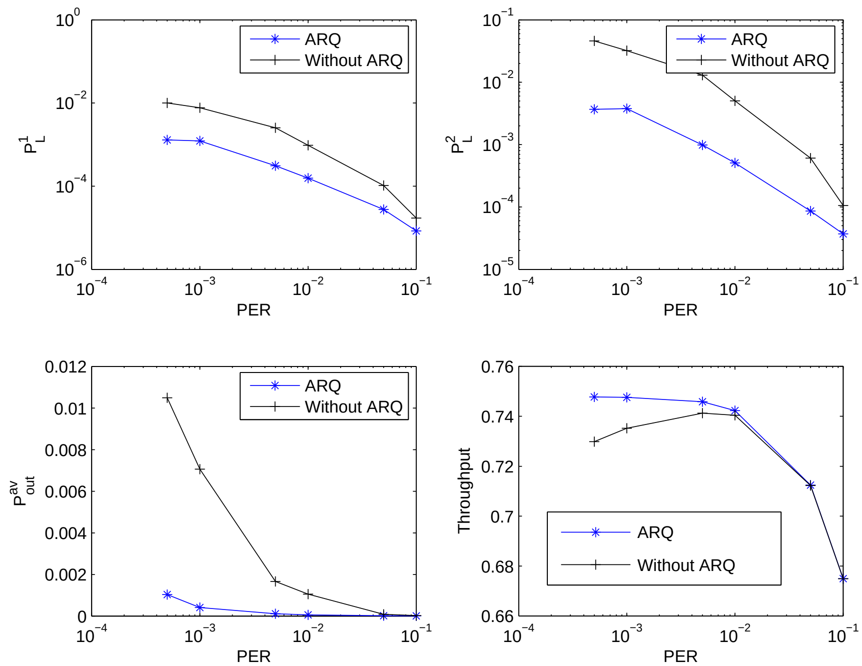

Figure 5.2. Performance of a SMDP-based AC policy.

5.6.1 Performance of the SMDP-based AC policy

We first investigate the performance for a SMDP-based AC policy, in which $L_1 = 0$, $L_2 = 1$, $\rho_1 = \rho_2$, and no buffer is employed. Figure 5.2 compares the performance for the system with ARQ and the system without ARQ when the SMDP formulation is employed and no buffer is allowed. In both cases, the QoS requirements in terms of packet loss probability and average outage probability, denoted by P_L^j and P_{out}^{av}, $j = 1,2$, respectively, can be satisfied. It is observed that the performance can be improved by allowing retransmissions. For example, with a PER constraint of 10^{-3}, employing ARQ can reduce the packet loss probability for class 1 from 0.0077 to 0.0012, i.e., reduced by 84.4%, reduce the outage probability from 0.0071 to 0.0004, i.e., reduced by 94.4%, while improving the system throughput from 0.7352 to 0.7476, i.e., improved by 1.69%.

5.6.2 Performance of the GSMP-based AC policy

Now we study the performance for a GSMP-based AC policy in which $\rho_1 = \rho_2$, $L_1 = L_2$, and $B_1 = B_2 = 1$. We investigate the performance for $L_j = 0, 1$ and 2, respectively. The results for large L_j can be extended straightforwardly.

Table 5.5 compares the overall packet loss probability, denoted by P_L, for difference L_j. We observe that for a small target PER region, which is reasonable in a practical system, the packet loss probability can be dramatically reduced with an increased L_j.

Figure 5.3 shows the delay, the outage probability and the throughput for different L_j. The delay presented here is the overall delay which is the mean of the sum of the packet access delay and the transmission delay. It is observed that ARQ increases the transmission delay and as a result, may increase the overall delay. This delay degradation can be very small in a small PER region due to the only occasionally occurring retransmissions. From Figure 5.3, it is also observed that the performance in terms of the outage probability and the system throughput can be improved by increasing L_j. However, when L_j is increased beyond a certain level, e.g., $L_j = 1$ in the investigated system, the performance improvement is very small. Therefore, in the system we investigated, there is no need to employ an ARQ with large L_j.

5.6.3 Comparison between SMDP and GSMP-based AC policies

Figure 5.4 compares the performance between SMDP and GSMP AC policies, in which $L_1 = 0$, $L_1 = 1$, and no buffer is employed. Figure 5.4 demonstrates that for a small number of retransmissions, SMDP and GSMP-based AC policies have similar performance. Although performance comparison for large L_j is not presented here since a SMDP-based AC policy would involve excessive computation, it is expected that for low PER, these two AC policies would still have similar performance. For a high PER, however, the packet

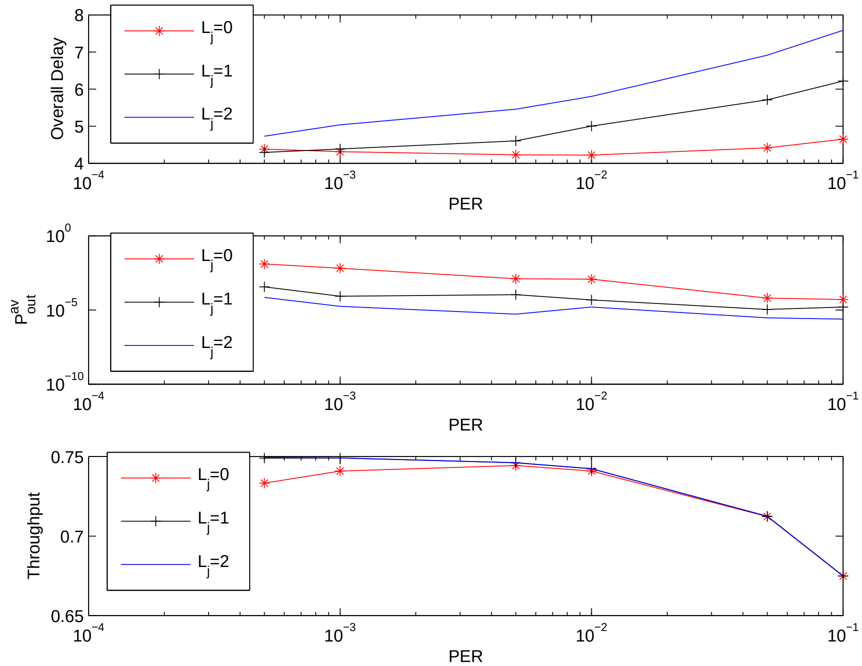

Figure 5.3. Performance of a GSMP-based AC policy.

duration is far from exponentially distributed, and thus linear programming cannot provide an optimal solution to a GSMP and its performance would be inferior to that of SMDP.

5.6.4 Comparison between exact and approximate approaches

As mentioned previously, the packet admission control policy proposed in this chapter uses the exact approach which can guarantee an exact outage probability constraint while maximizing system throughput. An AC policy based on an approximate PCFC and ARQ-based ROP algorithms, termed as approximate approach, is presented in [72].

Figure 5.5 compares the performance between the proposed AC policy with an exact approach and the AC policy employing an approximate approach which is presented in [72]. GSMP formulation is employed with $L_j = 1$ and $B_j = 0$ for $j = 1,2$. Although the average outage probability for the exact approach is inferior to approximate approach,

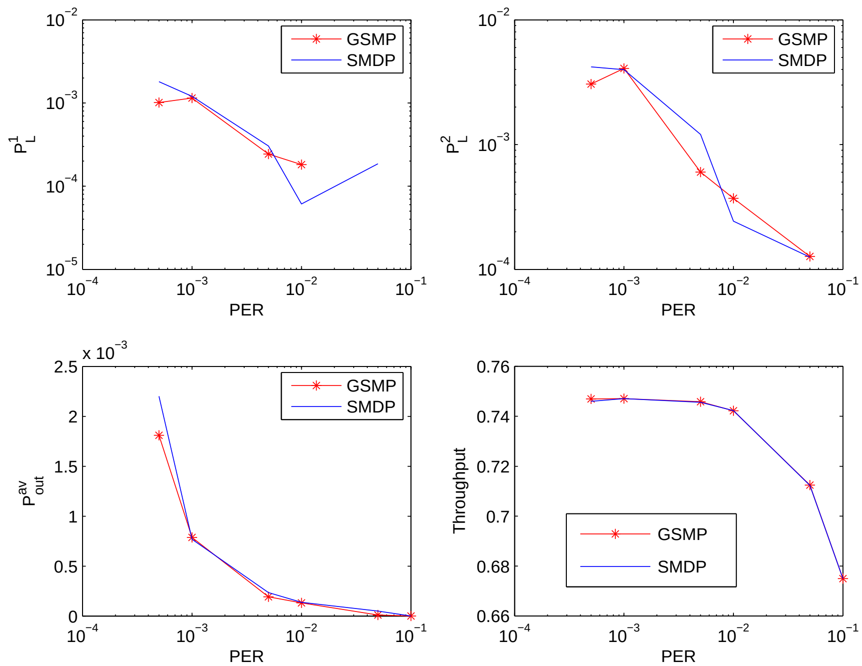

Figure 5.4. Performance comparison between SMDP and GSMP-based AC polices.

it is observed that the exact approach is able to achieve a lower packet loss probability and a higher throughput. This inferior average outage probability is due to the fact that for the exact approach only a required outage probability is ensured and no excessive outage probability reduction is desired, while for an approximate approach, the AC policy allows excess outage probability reduction, which wastes system resources. Therefore, for an exact approach, the resources can be more efficiently utilized to maximize the overall system throughput. Another advantage of the exact approach is that we can guarantee all QoS requirements for an arbitrary choice of ARQ parameters, while the approximate approach in [72] can only ensure the QoS requirements under certain ARQ parameters.

We remark that although the optimal policy proposed in this chapter outperforms the suboptimal policy in [72], it requires high computational complexity due to the evaluation of the outage probability for each possible system state.

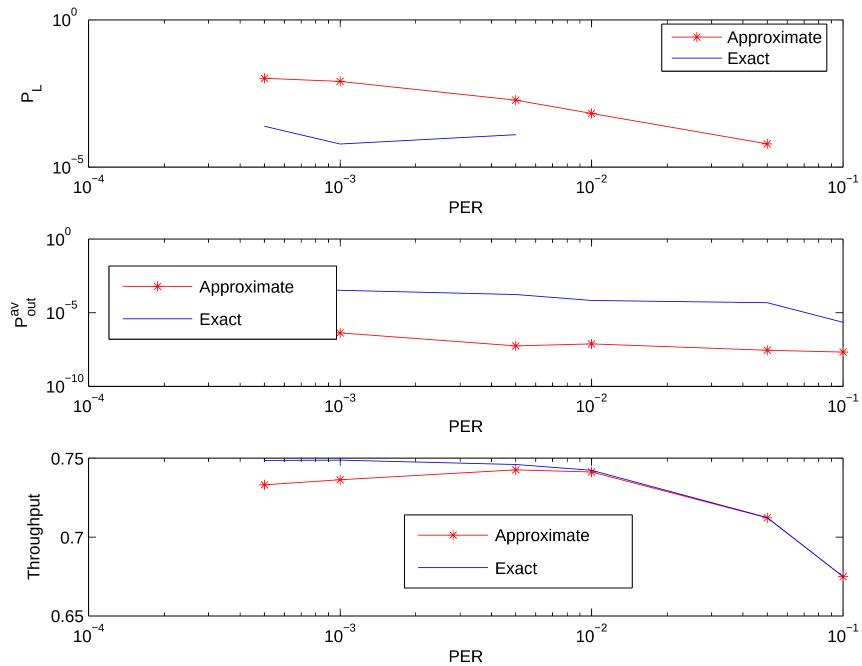

Figure 5.5. Comparison between proposed and existing PAC policies.

5.7 Conclusions

In this chapter, we investigate the cross-layer AC problem for packetized networks which incorporate a truncated ARQ scheme. With retransmissions, the Markovian property is violated and as a result the formerly formulated SMDP cannot be applied. We formulate a novel semi-Markov decision process and propose an optimal AC policy as well as a reduced-complexity GSMP-based AC policy. It can be shown that the proposed AC policies are capable of satisfying QoS requirements in both physical and packet levels, and within a reasonably small PER region, system performance in terms of outage probability, packet loss probability and throughput can be improved by employing ARQ.

Table 5.3. Representation of vectors in Table 5.1: each defined vector represents a possible state transition from current state **s**.

Notation	Event	State transition
$\mathbf{s}+\mathbf{q}^j$	New arrival for class j which is accepted.	An increase in subclass 1 of class j by 1.
$\mathbf{s}+\mathbf{b}^j$	New arrival for class j which is queued.	An increase in queue j by 1.
$\mathbf{s}+\mathbf{c}_i^j$, $i=1,..,L_j$	Departure of a class j and subclass i packet, correctly received and no queue is made active.	A decrease in subclass i for class j by 1.
$\mathbf{s}+\mathbf{r}_i^j$, $i=1,..,L_j$	Same as above, except that a packet in class j queue is made active.	A decrease in queue j by 1, a decrease in subclass i for class j by 1, and an increase in subclass 1 for class j by 1.
$\mathbf{s}+\mathbf{e}_i^j$, $i=1,..,L_j$	Departure of a class j subclass i packet, incorrectly received.	An increase of subclass $i+1$ for class j by 1, and a decrease in subclass i of class j by 1.
$\mathbf{s}+\mathbf{f}^j$	Departure of a class j subclass L_j+1 packet; a packet in class j queue is made active.	A decrease in subclass L_j+1 for class j by 1, a decrease in queue j by 1, and an increase in subclass 1 for class j by 1.
$\mathbf{s}+\mathbf{g}^j$	Same as above except that no packet in class j queue is made active.	A decrease in subclass L_j+1 for class j by 1.

Table 5.4. Numerical values of $E[F]$ and $Var[F]$ for a beamforming system.

M	1	2	3	4
$E[F]$	1.0000	0.54628	0.39504	0.32405
$Var[F]$	1.0000	0.42735	0.24374	0.21897

Table 5.5. Packet loss probability for a GSMP-based AC policy.

Target PER	5×10^{-4}	5×10^{-3}	1×10^{-2}
P_L with $L_j = 0$	1.15×10^{-2}	1.19×10^{-3}	1.1×10^{-3}
P_L with $L_j = 1$	5.43×10^{-4}	6.19×10^{-5}	1.86×10^{-4}
P_L with $L_j = 2$	1.24×10^{-4}	0	6.79×10^{-5}

Chapter 6

Connection Admission Control Policy for Packetized Systems with ARQ

6.1 Introduction

The proposed AC policy in Chapter 5, while dramatically improving system performance by employing both multiple antennas and ARQ, is designed at the packet level, in which connection level QoS, such as blocking probability and connection delay, is ignored. Therefore, this packet level AC policy cannot work well for a connection oriented packet based network. Moreover, AC policies performed at the packet level, instead of at the connection level, may incur implementation difficulties. This fact motivates an investigation into a connection level admission control policy for packet-switched networks with guaranteed QoS constraints at physical, connection and packet levels.

To the best of our knowledge, design of optimal connection admission control policy for a packetized CDMA beamforming system with ARQ has not been addressed previously. For example, in [52] [99], packet traffic is studied in a single-antenna CDMA system, in which the proposed connection admission control policies treat the SIR as quasi-static and do not adequately incorporate multiple antenna systems. In [88], the proposed connection admission control policy considers QoS requirements in different layers. However, there is no automatic retransmission request (ARQ) incorporated in the connection admission control design.

In this chapter, we propose an optimal connection admission control policy for a packet-switched network, in which both multiple antennas and ARQ schemes are employed. The connection admission control policy decides whether an incoming connection can be accepted. Each accepted connection generates a sequence of packets, which are then transmitted over the channel. The erroneously received packets are retransmitted until they are correctly received or a prescribed number of maximum allowed retransmissions is reached. There exists a performance tradeoff across different layers. For example, improving connection level performance allows more accepted connectons, which leads to an increased aggregate packet generation rate. When the packet generation rate exceeds the packet departure rate, extra packets should be dropped, degrading packet level performance. Although packet level performance can be improved by increasing the number of allocated channels, the physical layer performance degrades with an increased number of channels due to multi-access interference. The proposed cross-layer connection admission control policy is designed to determine these tradeoffs across different layers.

In Chapter 5, the ARQ and admission control schemes are both performed at the packet level. In this chapter, the admission control is performed at the connection level, while retransmissions are still performed at packet level, as is widely adopted in practical systems.

The rest of this chapter is organized as follows: The signal model and problem formulation are presented in Sections 6.2 and 6.3, respectively. In Sections 6.4 and 6.5, packet-level and physical-layer QoS requirements in terms of packet loss probability and outage probability are analyzed, respectively. An optimal connection admission control policy is derived in Section 6.6. Numerical results are presented in Section 6.7.

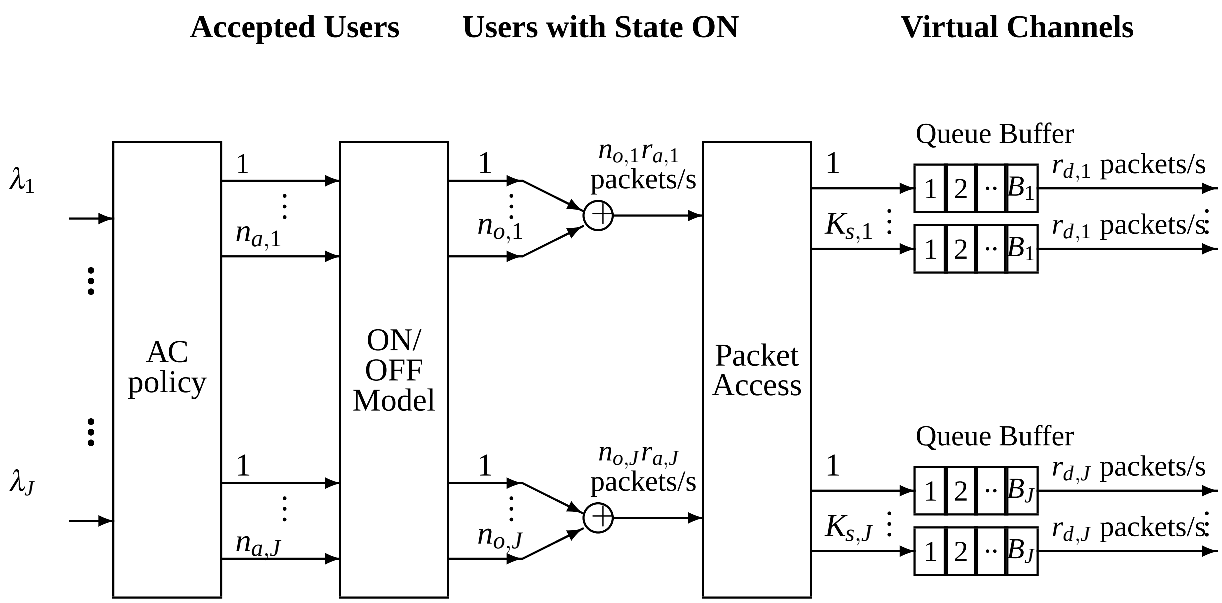

Figure 6.1. Signal model for packet-switched networks.

6.2 Signal Model

6.2.1 Traffic model

The signal model is illustrated in Figure 6.1. We consider a single-cell uplink CDMA beamforming system with M antennas at the BS. Assume that there are J classes of statistically independent traffic in the network. Each class of traffic is distinguished from others by its arrival rate, departure rate, transmission data rate and required QoS. The arrival process of the aggregate connectons is modeled by a Poisson process with rate λ_j for each class j, where $j = 1,..,J$. The duration for each connection is assumed to have exponential distribution with mean $\frac{1}{\mu_j}$.

Whenever a connection arrives, the connection admission control (AC) policy, derived offline and implemented as a lookup table, decides whether the incoming connection can be accepted. Denote $n_{a,j}$ as the number of accepted users for class j, where $j = 1,..,J$.

The system state, representing the number of accepted users for each class, is defined as $\mathbf{s} = [n_{a,1}, .., n_{a,J}]$. To reduce the size of the state space, no queue buffer is implemented at the connection level, which implies that if the incoming connection is not accepted immediately, it is blocked.

6.2.2 Signal model at the packet level

For each accepted connection, the modelling of packet generating traffic is still an open problem. As shown in [52], the user traffic may be modelled as an ON/OFF Markov process. Under this model, the user transits between the ON and OFF states during the whole connection. When a user is in an ON state, packets are generated with a rate $r_{a,j}$ packets per second and when the user is at OFF state, no packets are generated.

For a class j connection, the transition probabilities from ON state to OFF state, or from OFF state to ON state, are denoted by α_j and β_j, respectively. Denote p_{ON}^j as the probability that a class j user is in the ON state, which can be obtained by $p_{ON}^j = \frac{\beta_j}{\alpha_j + \beta_j}$. Given $n_{a,j}$ accepted users, the number of users in the ON state, denoted by $n_{o,j}$, is a Binomial distributed random variable. With $n_{o,j}$ users in the ON state, the overall arrival rate for class j is given by $n_{o,j} r_{a,j}$.

In contrast to a circuit-switched network, in which each user is allocated a dedicated channel with a fixed transmission data rate, for a packet switched network, no dedicated channels are allocated. Instead, all the generated packets from class j users share a given number of channels, which are termed *virtual channels*. The number of virtual channels, denoted by $K_{s,j}$, are decided by the number of accepted users, the traffic model as well as the QoS requirements.

All the generated packets from class j users are allocated to the $K_{s,j}$ virtual channels by a packet access scheme. The packets in each virtual channel are then transmitted with rate

$r_{d,j}$ in a packet-by-packet fashion. Before transmission, the allocated packets for a class j virtual channel is stored in a packet queue buffer, with buffer size B_j, where $j = 1, .., J$.

In this chapter, we consider a truncated ARQ scheme which retransmits an erroneous packet until it is successfully received or the number of maximum allowed retransmissions, denoted by L_j for class j packets, is reached, where $j = 1, .., J$. Once a packet is received, the receiver sends back an acknowledgement (ACK) signal to the transmitter. A positive ACK indicates that the packet is correctly received while a negative ACK indicates an incorrect transmission. If a positive ACK is received or the maximum number of re-transmissions, denoted by L_j, is reached, the packet releases the virtual channel and a packet in the queue can then be transmitted. Otherwise the packet will be retransmitted.

6.2.3 Signal model at the physical layer

Similar to the previous chapters, we consider a CDMA beamforming system with M antennas at the base station (BS). At the receiver, a spatial-temporal matched-filter receiver is employed. With $K = \sum_{j=1}^{J} K_{s,j}$ virtual channels, there are at most K packets simultaneously transmitted. The received signal-to-interference ratio (SIR) for a desired packet k, where $k = 1, .., K$, can be written as

$$SIR_k = \frac{W}{R_k} \frac{p_k \phi_{kk}^2}{\sum_{i=1, i \neq k}^{K} p_i \phi_{ik}^2 + \eta_0 W} \tag{6.1}$$

where W and R_k denote the bandwidth and the data rate for the virtual channel allocated to packet k, respectively, and the ratio $\frac{W}{R_k}$ represents the processing gain; $p_k = P_k G_k^2$ denotes the received power, in which P_k and G_k denote the transmitted power and link gain, respectively; η_0 denotes the one-sided power spectral density of background additive white Gaussian noise (AWGN); ϕ_{ik}^2 denotes the fraction of packet i's signal passed by the beamforming weights for desired packet k, which can be expressed as $\phi_{ik}^2 = \left| \mathbf{a}_k^H \mathbf{a}_i \right|^2$, in which

$\mathbf{a}_i$ denotes the normalized array response vector for packet i, and $(.)^H$ denotes conjugate transpose.

6.3 Problem Formulation

The connection-level and physical-layer QoS can be characterized by blocking probability and outage probability, respectively, while the packet-level QoS can be represented by packet loss probability, defined as the probability that a packet in an accepted connection cannot be delivered to the receiver. Other packet level QoS constraints, such as packet access delay, can be ensured by packet access control, which is not discussed in this thesis.

To characterize the overall system performance across different system layers, we define the system throughput as the number of correctly received packets per second, which can be expressed by

$$\begin{aligned}&\text{Throughput}\\ &= \sum_j \lambda_j (1-P_b^j)(1-P_{out}^{av}) p_{ON}^j r_{a,j} (1-P_L^j)(1-PER_{overall}^j)\end{aligned} \tag{6.2}$$

where P_b^j, P_{out}^{av}, P_L^j and $PER_{overall}^j$ denote the blocking probability, the average outage probability, the packet loss probability and the packet error rate for class j, respectively.

In this chapter, we aim to derive an optimal connection admission control policy which is capable of maximizing the above system throughput, while simultaneously guaranteeing QoS requirements at physical, packet and connection levels.

6.4 Packet-Level Design

A system state $\mathbf{s}$ is defined as $\mathbf{s} = [n_{a,1}, .., n_{a,J}]$, which represents the number of accepted users. In this section, we discuss how to choose $K_{s,j}$ for a given system state to guarantee the packet level QoS requirements in terms of packet loss probability. For simplicity, we first consider the case of zero buffer, i.e., $B_j = 0$. The results are then extended to nonzero buffer sizes.

6.4.1 Departure rate with retransmissions

Without ARQ, the duration for a packet can be expressed as $\frac{N_p}{R_j}$, where N_p denotes the packet length and R_j denotes the transmission rate. With ARQ, the packet duration, denoted by C_j, is the summation of the original packet duration and the duration for at most L_j retransmissions. The mean duration can be expressed from (5.24),

$$C_j = \frac{N_p}{R_j}(1 + (\rho_j)^{\frac{1}{L_j+1}} + .. + (\rho_j)^{\frac{L_j}{L_j+1}}) \tag{6.3}$$

in seconds.

The packet departure rate for each virtual channel, denoted by $r_{d,j}$, can be obtained by

$$\begin{aligned} r_{d,j} &= \frac{1}{C_j} \\ &= \frac{\frac{R_j}{N_p}}{1 + (\rho_j)^{\frac{1}{L_j+1}} + .. + (\rho_j)^{\frac{L_j}{L_j+1}}} \end{aligned} \tag{6.4}$$

in packets per second.

6.4.2 Packet loss probability

In the following, we assume that $B_j = 0$ and the incoming packets are allocated equally to the $K_{s,j}$ virtual channels, e.g., in a round-robin fashion. For each allocated virtual channel, the packet arrival rate can be expressed as $n_{o,j}r_{a,j}/K_{s,j}$, and the packet departure rate for each virtual channel, $r_{d,j}$, is given in (6.4).

To obtain the packet loss probability for given $n_{a,j}$, we first express the packet loss probability for a given $n_{o,j}$ as

$$\begin{aligned} & P_l^j(n_{o,j}, K_{s,j}) \\ = & \begin{cases} 0 & \text{if } n_{o,j}r_{a,j} \leq K_{s,j}r_{d,j} \\ \frac{n_{o,j}r_{a,j} - K_{s,j}r_{d,j}}{n_{o,j}r_{a,j}} & \text{if } n_{o,j}r_{a,j} > K_{s,j}r_{d,j}. \end{cases} \end{aligned} \tag{6.5}$$

Then the packet loss probability for a given $n_{a,j}$ can be obtained by

$$P_L^j(n_{a,j}, K_{s,j}) = \sum_{i=0}^{n_{a,j}} \text{Prob}\{n_{o,j} = i\} P_l^j(i, K_{s,j}) \tag{6.6}$$

$$\leq \nu_j \tag{6.7}$$

where ν_j denotes the packet loss probability constraint, and $\text{Prob}\{n_{o,j} = i\}$ denotes the probability that i out of $n_{a,j}$ accepted users are in the ON state, which has Binomial distribution

$$\text{Prob}\{n_{o,j} = i\} = (p_{ON}^j)^i (1 - p_{ON}^j)^{n_{a,j} - i} \tag{6.8}$$

for $0 \leq i \leq n_{a,j}$.

6.4.3 Choosing $K_{s,j}$

In the above analysis, we assume that the packet generation traffic is modeled by an ON/OFF Markov process and buffer sizes are all zero. Under these assumptions, with a given number of accepted users $n_{a,j}$ and packet-level QoS constraints, $K_{s,j}$ is chosen to satisfy (6.7).

For a general system, virtual channel can be approximated by a $G/G/1/1+B_j$ queue, where G denotes the general distributed arrival and departure processes. Given a nonzero B_j, Equation in (6.6) should be replaced by a corresponding packet loss probability formula by analyzing the G/G/1/1+B_j queue, and then $K_{s,j}$ can be chosen according to (6.7).

We note that for a given system state $\mathbf{s} = [n_{a,1}, .., n_{a,J}]$, an increase in the chosen $K_{s,j}$ can lead to an improved packet-level performance. However, large $K_{s,j}$ introduces more mutual interference, which degrades the physical layer performance. The choice of $K_{s,j}$ represents the tradeoff between physical-layer and packet-level performances.

In the above, we only consider the packet-level QoS requirement in terms of packet loss probability. As discussed previously, other packet-level QoS requirements, such as packet access delay and delay jitter, can be satisfied by performing packet access control.

6.5 Physical-Layer QoS: Outage Probability

Physical-layer performance is determined by the number of virtual channels, i.e., $K_{s,j}$. In the previous section, a lower bound of $K_{s,j}$ is given in (6.7), and an exact $K_{s,j}$ can then be determined by system resource allocation schemes, e.g., packet access control. In this section, we discuss how to ensure the physical-layer QoS requirements in terms of worst-state outage probability and average outage probability for beamforming systems in which $K_{s,j}$, where $j = 1, 2, .., J$, are known for each possible system state.

As discussed in the previous chapters, the outage probability constraints can be ensured by employing exact or approximate approaches. In the following, we employ the exact approach to guarantee the outage probability.

We first derive the outage probability for a given system state $\mathbf{s} = [n_{a,1},..,n_{a,J}]$, in which totally $\sum_{j=1}^{J} K_{s,j}$ channels are allocated. The outage probability for a given state is defined as the probability that a target PER, or equivalently a target SIR, cannot be satisfied. As shown in (5.8), the target SIR for a given PER constraint ρ_j, can be obtained as follows

$$\gamma_j = \frac{1}{g}[\ln a - \ln((\rho_j)^{\frac{1}{L_j+1}})] \tag{6.9}$$

in which a, g are constants depending on the chosen modulation and coding scheme [54].

Letting each transmitted packet achieve its target SIR, we have the following matrix form

$$[I_K - QF]\mathbf{p} = Q\mathbf{u} \tag{6.10}$$

where I_K is a $K-$dimensional identity matrix, power vector $\mathbf{p} = [p_1,..,p_K]^t$, $\mathbf{u} = \eta_0 B[1,..,1]^t$, $(.)^t$ denotes transpose, Q is a K-dimensional diagonal matrix with the i^{th} non-zero element as $\frac{\frac{\gamma_i R_i}{W}}{1+\frac{\gamma_i R_i}{W}}$, and F is a K by K matrix in which the element at the i^{th} row and the j^{th} column can be expressed as $F_{ij} = \frac{\phi_{ij}^2}{\phi_{ii}^2}$.

To ensure a positive solution for power vector $\mathbf{p}$, we require the following feasibility condition,

$$\upsilon(QF) < 1 \tag{6.11}$$

where $\upsilon(.)$ denotes the maximum eigenvalue, which is real-valued since the matrices are symmetric. Under the above feasibility condition, the power solution can be obtained by

$$\mathbf{p} = [I_K - QF]^{-1} Q\mathbf{u} \tag{6.12}$$

where $(.)^{-1}$ denotes matrix inversion.

Therefore, the outage probability for a given system state $\mathbf{s}$, in which totally $K = \sum_{j=1}^{J} K_{s,j}$ virtual channels are allocated, can be obtained as follows

$$P_{out}(\mathbf{s}) = P_{out}(K_{s,1},..,K_{s,J})$$

$$= \text{Prob}\{\upsilon(QF) \geq 1\} \qquad (6.13)$$

where Prob$\{A\}$ denotes the probability of event A.

Based on this state outage probability, the worst-state outage probability, denoted by P_{out}^{w}, and the average outage probability, denoted by P_{out}^{av}, can be expressed as follows

$$\begin{aligned} P_{out}^{w} &= \max_{\mathbf{s} \in S} P_{out}(\mathbf{s}) \qquad (6.14) \\ &\leq \rho_w \\ P_{out}^{av} &= \sum_{\mathbf{s} \in S} P_{\mathbf{s}} P_{out}(\mathbf{s}) \qquad (6.15) \\ &\leq \rho_{av} \qquad (6.16) \end{aligned}$$

where ρ_w and ρ_{av} denote the WSOP and AOP constraints, respectively; $P_{\mathbf{s}}$ denotes the steady-state probability that the system is in state $\mathbf{s}$ and S represents the set of all feasible system states, which will be discussed in Section 6.6.

6.6 Optimal Connection Admission Control Policy

The QoS requirements in the network layer can be characterized by blocking probability, defined as the probability that an incoming connection is blocked. The network-layer QoS as well as the other QoS should be guaranteed by a cross-layer connection admission control design.

In this chapter, we assume that the arrival process is Poisson distributed, the connection duration is exponentially distributed and the connection arrival and departure processes are independent. The system state is represented by the number of accepted connections. Under these assumptions, the process has the Markovian property that the future behavior of the process depends only on the present state and is independent of the past history [81]. In this sense, the connection admission control problem can be obtained by employing the

SMDP approach.

6.6.1 SMDP components

As discussed previously, a semi-Markov decision process includes the following components: system state, state space, action, action space, decision epoch, holding time, transition probability, policy and constraints.

System state is represented by the number of accepted connections, i.e., $\mathbf{s} = [n_{a,1}, .., n_{a,J}]$. A state is considered feasible if and only if this state can satisfy the worst-state-outage-probability and packet-loss-probability constraints. The state space includes all feasible system states, and can be expressed as

$$S = \{\mathbf{s}; P_{out}(\mathbf{s}) < \rho_w, \text{ and } P_L^j(n_{a,j}, K_{s,j}) \leq \nu_j\}.$$

The formulation of the above state space can be summarized as follows:

- Compute the maximum number of accepted users for each class, denoted by M_j^{max}. The search procedure for M_j^{max} is presented in Figure 6.2;
- An enlarged state space, denoted by $\bar{S}$, can be formulated as

$$\bar{S} = \left\{\mathbf{s} = [n_{a,1}, .., n_{a,J}] : n_{a,j} \leq M_j^{max} \text{ for } j = 1, .., J\right\};$$

- The above $\bar{S}$ can be truncated to the desired state space S as follows:
  - Initialize $S = \{\}$;
  - For each state $s \in \bar{S}$, choose appropriate $K_{s,j}$ for each j based on (6.7);
  - Evaluate $P_{out}(s)$ based on (6.13);
  - If $P_{out}(s) \leq \rho_w$, then $S = S + \{s\}$.

- We remark that in the above step, it is unnecessary to evaluate each system state in $\bar{S}$, since if $\mathbf{s} \in S$, then all $\mathbf{s}^{'} \in \bar{S}$ such that $\mathbf{s}^{'} \leq \mathbf{s}$ are also in S. Similarly, if $\mathbf{s}$ is not in S, then all $\mathbf{s}^{'} \in \bar{S}$ such that $\mathbf{s}^{'} \geq \mathbf{s}$ are also not in S.

After formulating the state space, a virtual-channel-table can then be obtained via (6.7), which assigns a required number of virtual channels to each possible system state.

The other components of a SMDP, including action, action space, decision epoch, the holding time, the transition probability, the policy and the constraints are summarized in Table 6.1.

In the admission control problem discussed in this chapter, we have QoS requirements in terms of blocking probability, packet loss probability, AOP and WSOP. While WSOP and packet loss probability requirements can be guaranteed by formulating the state space as shown in Table 6.1, the other QoS requirements can be guaranteed by SMDP constraints.

6.6.2 Deriving an AC policy by linear programming

The policy can be chosen according to a certain performance criterion, such as minimizing-blocking-probability or maximizing-throughput. Here we aim to find an optimal policy R^* which maximizes the throughput for any initial system state.

As formulating the admission problem as a SMDP, an optimal connection admission control policy can be obtained by using the decision variables $z_{\mathbf{sa}}, \mathbf{s} \in S, \mathbf{a} \in A_{\mathbf{s}}$, in solving the following linear (LP) problem:

$$\max_{z_{\mathbf{sa}} \geq 0, \mathbf{s}, \mathbf{a}} \sum_{\mathbf{s} \in S} \sum_{\mathbf{a} \in A_{\mathbf{s}}} \sum_{j=1}^{J} \lambda_j a_j (1 - P_{out}(\mathbf{s})) P_{ON}^{j} r_{a,j} (1 - P_L^{j})(1 - \rho_j) \tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} \tag{6.17}$$

subject to the set of constraints

$$\sum_{\mathbf{a} \in A_m} z_{\mathbf{ma}} - \sum_{\mathbf{s} \in S} \sum_{\mathbf{a} \in A_{\mathbf{s}}} p_{\mathbf{sm}}(\mathbf{a}) z_{\mathbf{sa}} = 0, m \in S$$

$$\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} \tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} = 1$$

$$\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} (1-a_j)\tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} \leq \Psi_j,\ j=1,..,J$$

$$\sum_{\mathbf{s}\in S}\sum_{\mathbf{a}\in A_{\mathbf{s}}} P_{out}(\mathbf{s})\tau_{\mathbf{s}}(\mathbf{a}) z_{\mathbf{sa}} \leq \rho_{av}$$

where Ψ_j and ρ_{av} denotes the blocking probability and AOP constraints, respectively.

In the above LP formulation, $\tau_{\mathbf{s}}(\mathbf{a})z_{\mathbf{sa}}$ represents the steady-state probability that the system is at state **s** and an action **a** is chosen. The objective function in (6.17) is to maximize the system throughput, the first constraint is the balance equation, and the second constraint ensures the sum of all the steady-state probabilities to be one. The latter two constraints represent the QoS requirements in terms of blocking probability and average outage probability, respectively.

6.6.3 Implementation of the cross-layer connection admission control design

The cross-layer connection admission control design can be implemented as follows:

- Derive the connection admission control policy offline:
  - Formulate the state space according to the procedure in Section 6.6.1. Then derive a virtual channel table based on (6.7), which assigns a required number of virtual channels to each system state;
  - Formulate other SMDP components according to Table 6.1;
  - The policy can then be derived according to (6.17);
  - Implement the connection admission control policy as a lookup table;

  - Whenever parameters change, repeat the above procedure to update the connection admission control lookup table and virtual channel table.

- Connection level implementation: whenever a connection arrives, the lookup table is employed to decide whether this packet can be accepted. The current state information, represented by the number of accepted users, and the virtual channel table, are then passed to the packet level.

- Packet level implementation:

  - The current state information and the virtual channel table are obtained from connection level;

  - For each system state, choose $K_{s,j}$ according to the virtual channel table;

  - For each incoming packet in class j, if the current number of simultaneously transmitted packets is less than $K_{s,j}$, the incoming packet can be transmitted. Otherwise, it is stored in the buffer;

  - The packets in the i^{th} virtual channel, where $i = 1,..,K_{s,j}$, are transmitted over the channel. An erroneous packet is retransmitted until it is correctly received or the maximum number of retransmissions is reached;

  - The chosen $K_{s,j}$ information is passed to the physical layer.

- Physical layer implementation:

  - $K_{s,j}$ is obtained from packet level;

  - Power is adjusted to the desired level, which is given in (6.12).

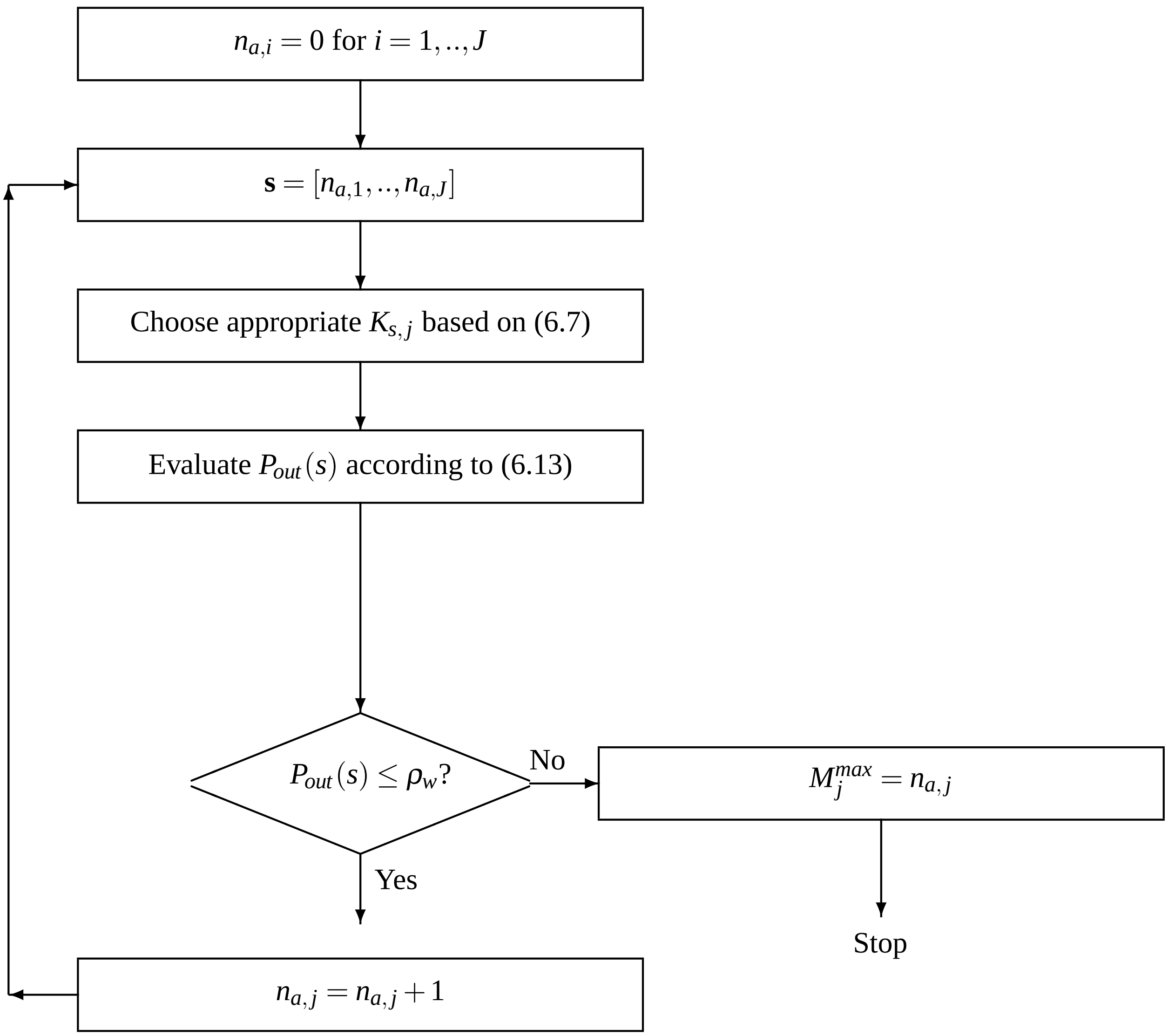

Figure 6.2. Search procedure for M_j^{max}.

Table 6.1. Formulating the optimal connection admission control problem as a SMDP.

SMDP components	Notation	Expression
System state	$\mathbf{s}$	$\mathbf{s} = [n_{a,1}, .., n_{a,J}]$.
State space	S	$S = \{\mathbf{s}; P_{out}(K_{s,1}, .., K_{s,J}) < \rho_w,$ and $P_L^j(n_{a,j}, K_{s,j}) \leq \nu_j\}$.
Decision epochs	t_k	The set of all arrival and departure instances.
Action	$\mathbf{a}$	$\mathbf{a} = [a_1, .., a_J]$, where $a_j = 1$ represents the decision to accept a class j connection, while $a_j = 0$ represents a rejection.
Admissible action space	$A_\mathbf{s}$	$A_\mathbf{s} = \{\mathbf{a} : a_j = 0,$ if $\mathbf{s} + e_s^j \notin S,$ and $\mathbf{a} \neq \mathbf{0}$ if $\mathbf{s} = \mathbf{0}\}$ in which $\mathbf{e}_s^j$ represents a J- dimensional vector, which contains only zeros except for position j which contains a 1.
Expected holding time	$\tau_\mathbf{s}(\mathbf{a})$	$\tau_\mathbf{s}(\mathbf{a}) = \left(\sum_{j=1}^J \lambda_j a_j + \sum_{j=1}^J \mu_j n_s^j\right)^{-1}$.
Transition probability	$p_{\mathbf{sy}}(\mathbf{a})$	$p_{\mathbf{sy}}(\mathbf{a}) = \lambda_j a_j \tau_\mathbf{s}(\mathbf{a}),$ if $y = \mathbf{s} + \mathbf{e}_s^j$; and $p_{\mathbf{sy}}(\mathbf{a}) = \mu_j n_s^j \tau_\mathbf{s}(\mathbf{a}),$ if $y = \mathbf{s} - \mathbf{e}_s^j$.
Policy	R	$R = \{R_s : S \to A \| R_\mathbf{s} \in A_\mathbf{s}, \forall \mathbf{s} \in S\}$ where A denotes the set of all admissible action space.
Constraints		$P_{out}^{av} \leq \rho_{av}$ and $P_b^j \leq \Psi_j$, where Ψ_j denotes the blocking probability constraint for class j.

Table 6.2. Simulation parameters.

W	3.84 MHz	a	90.2514
g	3.4998	γ_0	1.0942 dB
R_1	32 kbps	R_2	128 kbps
λ_1	0.01	λ_2	0.003
μ_1	0.005	μ_2	0.00125
$r_{a,1}$	50	$r_{a,2}$	200
P_{ON}^1	0.4	P_{ON}^2	0.6
ρ_w	0.5	M	2

6.7 Numerical Examples

In the following examples, we consider a packet-switched network with two-class multimedia services. A circular antenna array and a uniformly distributed AoA are assumed. A QPSK and convolutionally coded modulation scheme with rate $\frac{1}{2}$ and packet length $N_p = 1080$ is assumed at the transmitter. Under this scheme, the parameters of a, g and γ_0 in Equation (5.6) can be obtained from [54]. For simplicity, $B_1 = B_2 = 0$ is employed. Simulation parameters are summarized in Table 6.2.

Without loss of generality, we choose $K_{s,j}$ to be the minimum number satisfying (6.7). The chosen $K_{s,j}$ can ensure the packet level QoS requirement while simultaneously minimizing the outage probability in the physical layer.

In the following, we first illustrate the performance for different packet loss probability constraints, in which the proposed policy and the policy for circuit-switched networks, discussed in Chapter 3, are compared. We then present the performance gain for the proposed connection admission control policy with ARQ over the system without ARQ schemes, such as the policies discussed in [52] [88].

6.7.1 Performance for a packet-switched network

In the following, we compare the performance for different packet loss probability constraints, in which no ARQ schemes are employed. Since a strict packet loss probability constraint introduces a large blocking probability, which may lead to infeasibility in (6.18), we now relax the blocking probability constraints to 0.5 for both classes to ensure problem feasibility. The target SIR for class 1 and class 2 users are set to 10 and 7 dB, respectively.

Figures 6.3-6.6 compare the blocking probability, the average outage probability, the average packet loss probability and the system throughput for different packet-loss-probability constraints, respectively. For simplicity, we assume the packet loss probability constraints are the same for both classes, which are denoted by P_{loss} *constraint* in the figures. From these figures, we observe that the performance in one layer strongly depends on the QoS constraints of the other layers. For example, given an average outage probability constraint, relaxing the packet-loss-probability constraint can dramatically reduce the blocking probability in the network layer, while simultaneously improving the overall system throughput. This is because with the same physical layer performance, a large packet loss probability constraint allows more users to access the network. In the system we investigate, with $\rho_{av} = 10^{-2}$, relaxing packet loss probability constraint from 0 to 0.05 can reduce the blocking probability from 10^{-1} to 10^{-3}, i.e., reduced by 99%, while improving the throughput from 0.5 to 0.545, i.e., improved by 9%.

We note that the achieved packet loss probability in Figure 6.5 is obtained by averaging the measurements over a long-term period, while P_{loss} *constraint* denotes the maximum allowed packet loss probability for each system state.

In a circuit-switched network discussed in Chapter 3, a zero packet-loss-probability can

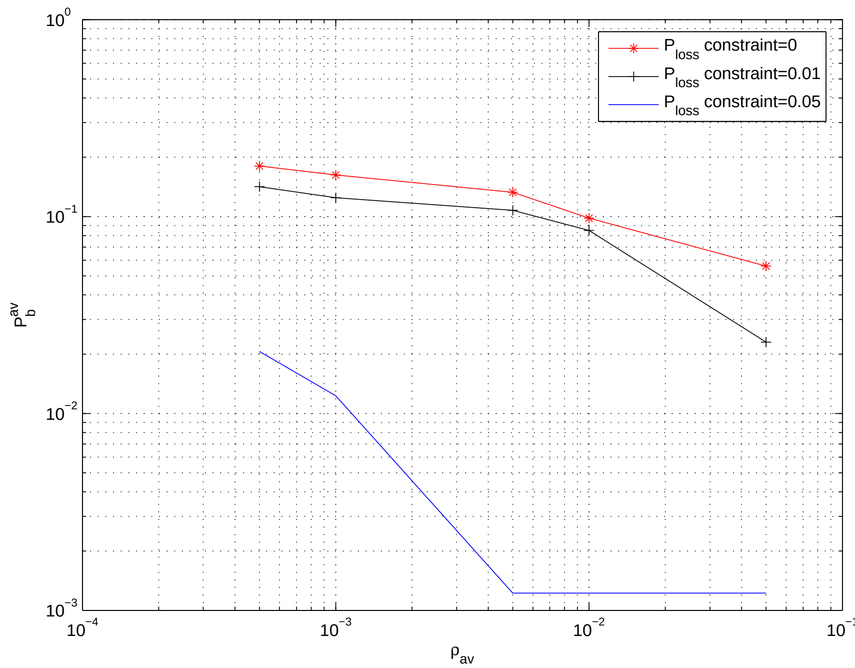

Figure 6.3. Blocking probability as a function of ρ_{av}.

be ensured. As observed in Figures 6.3-6.6, in a packetized system which allows a non-zero packet loss probability, this zero packet loss probability leads to an inefficient utilization of the system resource and as a result degrades the connection level performance as well as the overall system throughput.

6.7.2 Performance by employing packet retransmissions

Figures 6.7-6.9 compare the performance between a system without ARQ and a system with ARQ. In these figure, ARQ $= i$ is equivalent to $L_1 = L_2 = i$. The blocking probability is set to 0.1 for both classes and the target overall PERs are set to $\rho_1 = 10^{-4}$ and $\rho_2 = 10^{-6}$, respectively. The packet loss probability constraints are set to 0.05 for both classes.

From Figure 6.7, it is observed that with ARQ, the blocking probability and outage probability can be reduced. This represents a tradeoff between transmission delay and

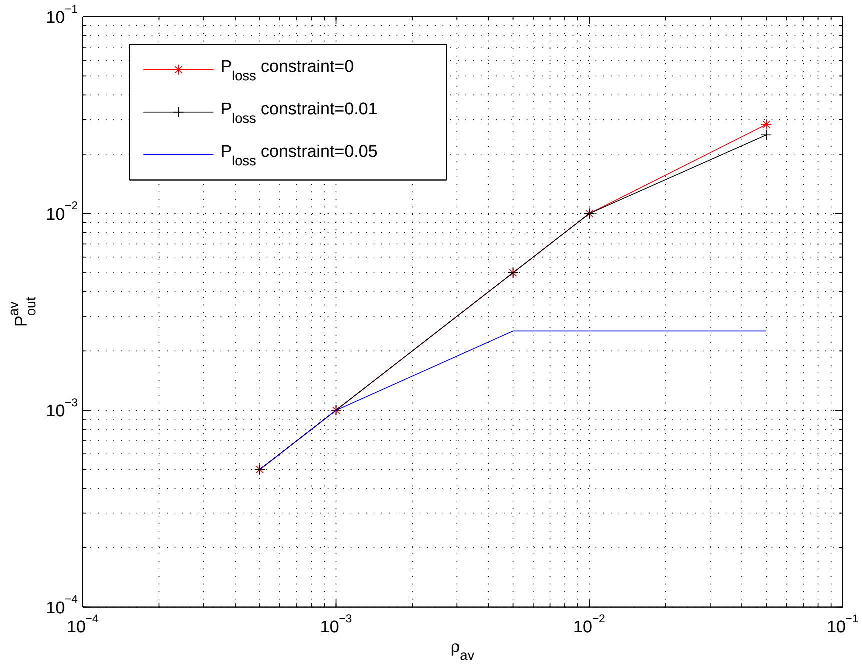

Figure 6.4. Outage probability as a function of ρ_{av}.

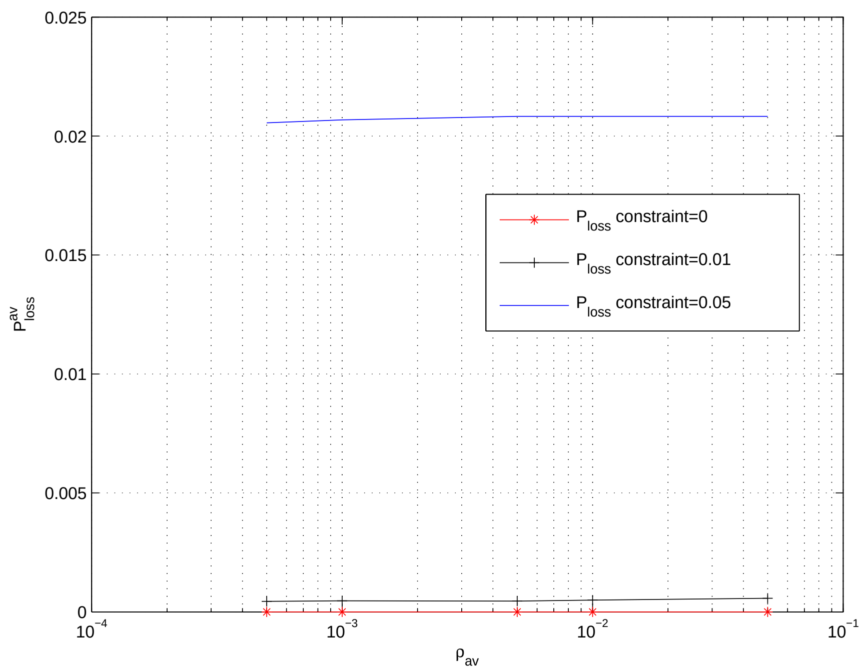

Figure 6.5. Average packet loss probability as a function of ρ_{av}.

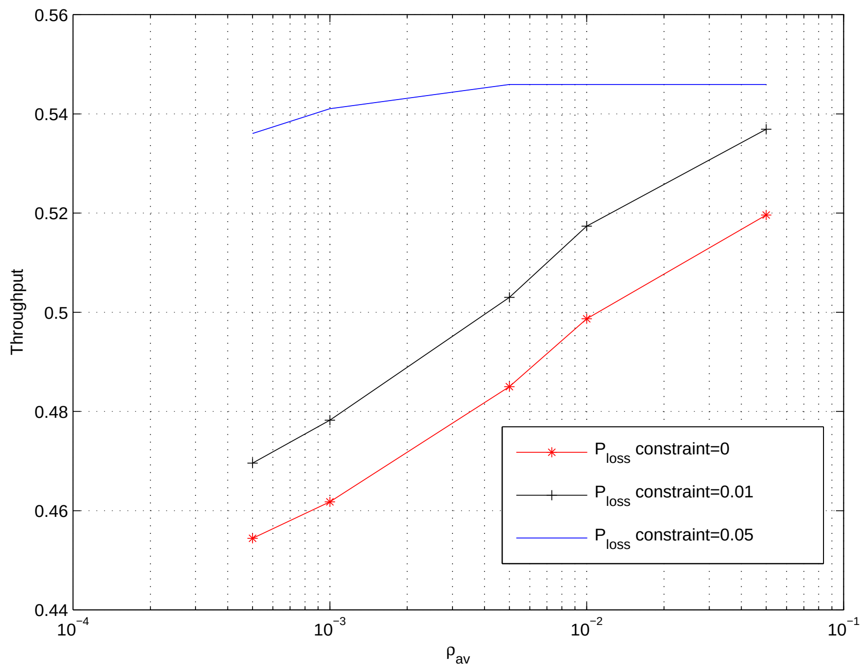

Figure 6.6. Throughput as a function of ρ_{av}.

system performance. For example, with $\rho_{av} = 10^{-3}$, employing an ARQ scheme with $L_j = 1$ can decrease the blocking probability from 10^{-3} to 10^{-4}, i.e., reduced by 90%, while simultaneously reducing the outage probability from 10^{-3} to almost 10^{-6}, i.e., reduced by 99%.

In the above, we have studied the physical and network layer performance by employing ARQ. We now investigate how ARQ schemes affect the packet level performance. As shown in (6.4), with an increased L_j, the departure rate is decreased due to retransmissions, which increases the packet loss probability. However, at the same time, an increased L_j also reduces the transmission error, allowing more virtual channels simultaneously presented in the system, which in turn decreases the packet loss probability. Therefore, the packet loss probability is determined by the above positive and negative impacts of ARQ. If the positive impact dominates, the packet loss probability is reduced by employing ARQ, as shown in

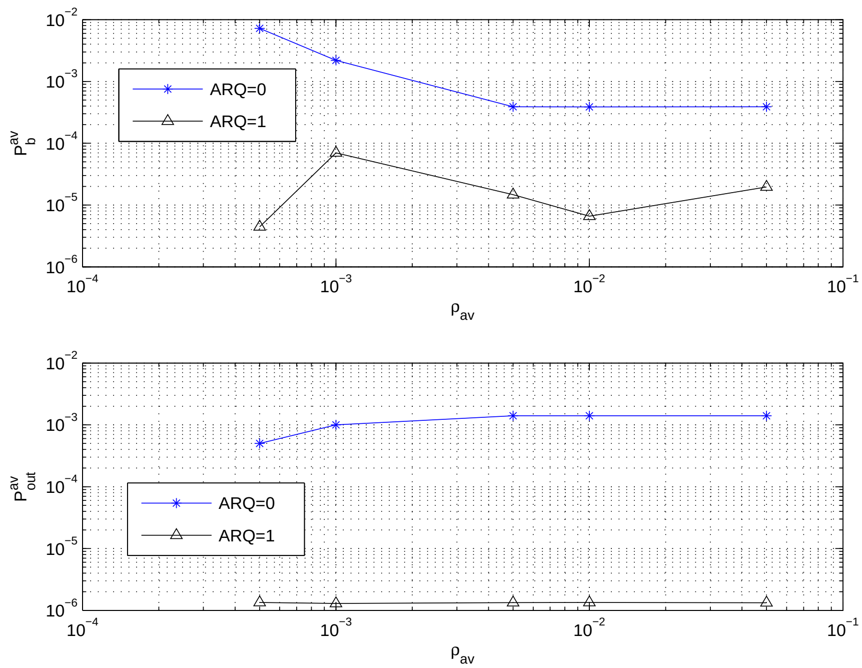

Figure 6.7. Blocking and outage probabilities as a function of ρ_{av}.

the upper figure in Figure 6.8. Otherwise, if the negative impact dominates, the packet loss probability is degraded by employing ARQ, as shown in the lower figure in Figure 6.8. We note that the above degradation is not very significant. As shown in Figure 6.9, with ARQ, the overall system throughput can be improved.

Although increasing L_j may further improve the system performance, it dramatically increases the computational complexity of the SMDP-based connection admission control policy. From Chapter 5, it is observed that when L_j exceeds a certain level, further increasing L_j cannot improve the performance significantly. Therefore, there is no need to choose a large L_j. A detailed discussion on the impact of ARQ and how to choose L_j can be found in [72], in which a packet-level AC is discussed which employs an ARQ-based ROP algorithm. In this chapter, we only discuss the connection admission control policy for a given L_j.

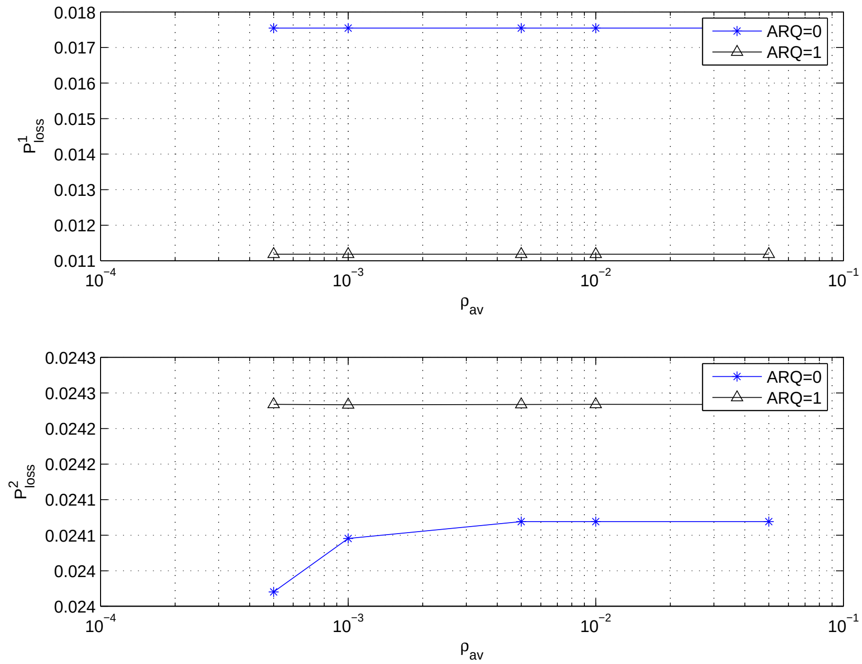

Figure 6.8. Packet loss probability as a function of ρ_{av}.

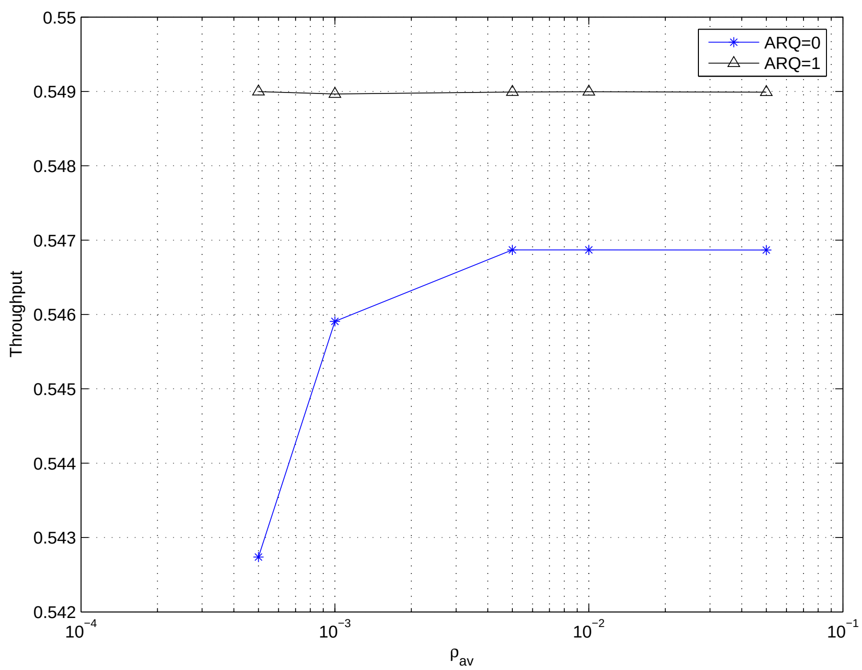

Figure 6.9. Throughput as a function of ρ_{av}.

6.8 Conclusions

In this chapter, an optimal connection admission control policy is proposed for a packet-switched CDMA beamforming system in the presence of ARQ. Compared with the previously proposed packet-level AC policy, in which the connection level QoS is ignored, the proposed connection admission control policy can ensure all the QoS requirements in connection, packet and physical levels. Furthermore, the proposed policy is performed at the connection level, instead of at the packet-level, which is more practical to implement. Compared with the CAC policy for circuit-switched networks, the proposed connection admission control policy allows dynamical allocation of the limited resources, and as a result, is capable of efficiently utilizing the resources and flexibly handling the multiple QoS requirements.

Chapter 7

Summary, Conclusions and Future Work

In this chapter, we summarize the major contributions in this thesis and suggest several topics for future research.

7.1 Summary and Conclusions

We have investigated cross-layer admission control problem in the presence of both multiple antenna and error control schemes.

In Chapter 3, we study an exact approach to ensure physical layer QoS requirements. Based on this exact approach, an optimal CAC policy is proposed, which can maximize the system throughput while simultaneously guaranteeing QoS requirements in both physical and network layers. The proposed cross-layer CAC policy provides a flexible way to trade off physical and network layer performance to optimize the overall system throughput. When average-outage-probability (AOP) constraint is relaxed from 10^{-4} to 10^{-2}, it is possible to reduce the blocking probability 88% for a two-antenna system, and increase the throughput 10%. Therefore, allowing for outage probability in the physical layer within its constraint can reduce the overall blocking probability and may improve the system throughput. In this chapter, it has been established that compared with the case of single antenna systems, employing multiple antennas and voice activity at the BS can dramatically improve the system performance in terms of blocking probability and system throughput. For example, with a single antenna at the BS, employing voice activity can improve the throughput 10%. By employing two antennas at the BS, the system throughput can be further improved by 14%.

In Chapter 4, an approximate approach, which includes an approximate PCFC and a ROP algorithm, is studied to ensure the physical layer QoS requirements. Based on this approximate approach, a low-complexity sub-optimal CAC policy is proposed which can guarantee both physical and network layer QoS requirements. Comparison between optimal and suboptimal CAC policies shows that suboptimal CAC policy has a slightly degraded performance. For example, with an AOP constraint of 0.035, using suboptimal CAC policy degrades the blocking probability from 0.06 to 0.15, and decreases the throughput from 1.37 to 1.24 calls/second, i.e., by 10%. However, by employing suboptimal CAC policy, the complexity can be reduced from N_r eigenvalue computations to J summations and J multiplications, where typically $N_r \geq 1000$ and $J \leq 10$, a significant reduction.

In Chapters 3 and 4, it is also shown that the proposed optimal and suboptimal CAC policies can guarantee QoS requirements in both physical and network layers, while for an existing CS-based CAC policy, QoS constraints in the network layer may be violated.

In Chapter 5, we apply the above approaches to a packet-switched network which employs a truncated ARQ scheme to mitigate transmission errors. We focus on packet-level AC problem, and QoS requirements in the connection level are ignored. By considering the impact of packet-level retransmissions, we formulate a new semi-Markov decision process, and an optimal SMDP-based admission control policy is then proposed. A suboptimal low-complexity GSMP-based AC policy is also discussed in this chapter. Simulation results show that the SMDP and GSMP-based policies can achieve similar performance for a small target packet-error-rate (PER) region. It is also found that within a reasonably small PER region, physical layer and packet level performance can be improved by employing ARQ. For example, compared with the system without ARQ, when the target PER is 10^{-2}, employing ARQ with $L_j = 1$, where L_j denotes the maximum number of retransmissions, can reduce the packet loss probability from 10^{-3} to 10^{-4}. The performance can be further improved by increasing the maximum number of retransmissions, which unfortunately increases the computational complexity. From the simulation results, we observe that when the maximum number of retransmissions is increased beyond a certain level, e.g., $L_j = 1$ in the system investigated, the performance is not improved significantly. Therefore, there is no need to employ ARQ schemes with a large L_j.

In Chapter 6, optimal connection-level AC problem is investigated in a connection-oriented packet-based network. In contrast to the proposed admission control policies in Chapter 5, in which connection-level QoS requirements are ignored, in Chapter 6, the proposed connection admission control policy can guarantee all the QoS requirements in physical, packet and connection levels. It has been shown that compared with the connection admission control policies in circuit-switched networks, allowing a non-zero packet loss probability within its constraint leads to an efficient utilization of system resources, and as a result, improves the connection level performance as well as overall system throughput. The impact of ARQ schemes are also investigated in this chapter. With ARQ, the connection-level and physical-layer performance can be improved. For example, it is possible to reduce the blocking probability from 10^{-3} to 10^{-4} by employing $L_j = 1$, while simultaneously reducing the achieved outage probability from 10^{-3} to almost 10^{-6}. The impact of ARQ schemes on the packet level performance, however, is not straightforward: with ARQ, the packet departure rate is decreased, which degrades packet-level performance, while at the same time, ARQ increases the number of virtual channels, which improves the packet-level performance. The overall packet level performance may be degraded (improved) by employing ARQ if the above negative (positive) impact dominates. In summary, the proposed connection admission control policy in this chapter allows dynamical allocation of limited resources, and as a result, is capable of utilizing resources more efficiently with guaranteed QoS requirements in physical, packet and network layers.

7.2 Future Work

Several possible future directions of investigation are listed as follows:

- In this thesis, we considered a matched-filter receiver in the physical layer. Although in some of our previous work, e.g., [70] [71], we have studied suboptimal policies for a linear-minimum-mean-square-error (LMMSE) receiver, these policies can only be applied to a large system [31]. In the future, one may consider extending this cross-layer CAC research to multiuser receivers with an arbitrary system size.

- For the exact approach discussed in Chapter 3, our proposed optimal policy depends on the outage probability evaluation for each possible system state, which leads to high computational complexity for the system that lacks a closed-form analytical expression for outage probability. In the future, any progress on simplifying this analytical expression for outage probability would help to dramatically reduce the complexity of the proposed optimal algorithms.

- The AC policies proposed in this thesis are based on the assumptions that session arrival and departure processes have independent Poisson distributions. With generally distributed arrival/departure process, we need to formulate a generalized semi-Markov decision process. Although the linear programming approach discussed in this thesis provides a suboptimal solution, a more accurate solution applicable to GSMP would be of great interest.

- In this thesis, we employ an ARQ scheme to mitigate transmission error, which requires a feedback channel from the BS to each mobile user carrying acknowledgement (ACK) information, leading to inefficient resource utilization and latency. Recently, rateless codes [18] are attracting significant attention due to their powerful error control capability while requiring no acknowledgement feedback. One interesting topic in the future is to design a cross-layer admission control policy by employing rateless codes instead of ARQ.

Bibliography

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathemetical Functions*. New York: Dover Publications, 1972.

[2] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, pp. 117–128, Jan. 1977.

[3] ——, *Multiple Access Communications*. New York: IEEE Press, 1993.

[4] M. H. Ahmed, "Call admission control in wireless networks: a comprehensive survey," *IEEE Commun. Surv. & Tutor.*, pp. 50–69, First quarter 2005.

[5] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[6] M. Andersin, Z. Rosberg, and J. Zander, "Soft and safe admission control in cellular networks," *IEEE Trans. Net.*, vol. 5, no. 2, pp. 255–265, Apr. 1997.

[7] A. Banerjee, D. Costello, and T. Fuja, "Diversity combining techniques for bandwidth-efficient turbo ARQ systems," in *Proc. IEEE International Symp. on Inform. Th.*, 2001.

[8] Y. Bao and A. Sethi, "Performance-driven adaptive admission control for multimedia applications," in *Proc. IEEE Int. Conf. Communications*, vol. 1, Vancouver, BC, Canada, 1999, pp. 199–203.

[9] D. Bertsekas and R. Gallager, *Data Networks*. Prentice Hall, 1992.

[10] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*. Belmont, Massachusetts: Athena Scientific, 1997.

[11] S. D. Blostein and H. Leib, “Multiple antenna systems: Their role and impact in future wireless access,” *IEEE Commun. Mag.*, vol. 41, no. 7, pp. 94–101, July 2003.

[12] S. K. Bose, *An Introduction to Queuing Systems*. Kluwer/Plenum Publishers, 2001.

[13] P. T. Brady, “A statistical analysis of on-off patterns in 16 conversations,” *Bell Syst. Tech. J.*, vol. 47, pp. 73–91, Jan. 1968.

[14] S. Brueck, E. Jugl, H. Kettschau, M. Link, J. Mueckenheim, and A. Zaporozhets, “Radio resource management in HSDPA and HSUPA,” *Bell Labs Tech. Journal*, vol. 11, no. 4, pp. 151–167, 2007.

[15] R. M. Buehrer, S. P. Nicoloso, and S. Gollamudi, “Linear versus nonlinear interference cancellation,” *J. Commun. and Networks*, vol. 1, pp. 118–133, June 1999.

[16] G. Caire and D. Tuninetti, “The throughput of hybrid-ARQ protocols for the gaussian collision channel,” *IEEE Trans. on Inform. Th.*, vol. 47, no. 5, pp. 20–25, Jul.-Aug. 2001.

[17] J. I. Capetanakis, “Tree algorithms for packet broadcast channels,” *IEEE Trans. Info. Theory*, vol. IT-25, pp. 505–515, Sep. 1979.

[18] J. Castura and Y. Mao, “Rateless coding for wireless relay channels,” in *Proc. IEEE International Symp. on Inform. Th.*, 2005.

[19] J. Chang, J. Chung, and D. Sung, "Admission control schemes for soft handoff in DS-CDMA cellular systems supporting voice and stream-type data services," *IEEE Trans. Veh. Technol.*, vol. 51, no. 6, pp. 1445–1459, Nov. 2002.

[20] F. Chiti, R. Fantacci, G. Mennuti, and D. Tarchi, "Dynamic sir-based admission control algorithm for 3g wireless networks," in *Proc. IEEE Int. Conf. Communications*, vol. 3, 2003, pp. 1907–1911.

[21] J. Choi, T. Kwon, Y. Choi, and M. Naghshineh, "Call admission control for multimedia services in mobile cellular networks: a Markov decision approach," in *Proc. IEEE Int. Symp. Compt. Commu.*, Antihes, Greece, 2000.

[22] C. Comaniciu, N. B. Mandayam, and H. V. Poor, *Wireless Networks: Multiuser Detection in Cross-Layer Design*. New York: Springer, 2005.

[23] C. Comaniciu, N. B. Mandayan, D. Famolari, and P. Agrawal, "Qos guarantees for third generation (3G) CDMA systems via admission and flow control," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, Boston, MA, USA, 2000, pp. 249–256.

[24] C. Comaniciu and H. V. Poor, "Jointly optimal power and admission control for delay sensitive traffic in CDMA networks with LMMSE receivers," *IEEE Trans. Signal Processing*, vol. 51, no. 8, pp. 2031–2042, Aug. 2003.

[25] L. M. Correia, *Mobile Broadband Multimedia Networks: Techniques, Models and Tools for 4G*. Academic Press, 2006.

[26] R. T. Derrybery, S. D. Gray, D. M. Ionescu, G. Mandyam, and B. Raghothaman, "Transmit diversity in 3G CDMA systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 68–75, Apr. 2002.

[27] Z. Dziong, M. Jia, and P. Mermelstein, "Adaptive traffic admission for integrated services in CDMA wirelessaccess networks," *IEEE J. Select. Areas Commun.*, vol. 14, no. 9, pp. 1737–1747, Dec. 1996.

[28] A. Ephremides and B. Hajek, "Information theory and communication networks: an unconsummated union," *IEEE Trans. on Informaiton Theory*, vol. 44, no. 6, pp. 2416–2434, Oct. 1998.

[29] K. Etemad, *CDMA2000 Evolution*. John Wiley and Sons, 2004.

[30] J. Evans and D. Everitt, "Effective bandwidth-based admission control for multiservice CDMA cellular networks," *IEEE Trans. Veh. Technol.*, vol. 48, no. 1, pp. 36–46, Jan. 1999.

[31] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 2059–2078, Sep. 2000.

[32] F. R. Farrokhi, L. Tassiulas, and K. J. R. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1313–1324, Oct. 1998.

[33] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell labs Tech. Journal*, vol. 1, no. 2, pp. 41–59, Aug. 1996.

[34] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 11, pp. 641–646, Nov. 1993.

[35] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. W. Jr., and C. E. W. III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 303–312, May 1991.

[36] J. P. Gray, "Line control procedures," in *Proc. IEEE*, 1972.

[37] Y. Guo and B. Aazhong, "Call admission control in multiclass traffic CDMA cellular system using multiuser antenna array receiver," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, Tokyo, Japan, 2000, pp. 365–369.

[38] Y. Hara, "Call admission control algorithm for CDMA systems with adaptive antennas," in *Proc. IEEE Veh. Technol. Conf.*, vol. 5, Boston, MA, USA, 2000, pp. 2518–2522.

[39] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Info. Theory*, vol. 48, no. 7, pp. 1804–1824, July 2002.

[40] T. Haynes, "A primer on digital beamforming," *Spectrum Signal Processing*, Mar. 1998.

[41] C. Ho, J. Copeland, C. Lea, and G. Stuber, "On call admission control in DS/CDMA cellular networks," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1328–1343, Nov. 2001.

[42] H. Holma and A. Toskala, *WCDMA for UMTS*. Wiley, 2004.

[43] J. Huang, R. Berry, and M. Honig, "Wireless scheduling with hybrid arq," in *38th Annual Conference on Information Sciences and Systems*, Princeton, NJ, 2004.

[44] S. Y. Hui and K. H. Yeung, "Challenges in the migration to 4G mobile systems," *IEEE Commun. Mag.*, vol. 41, no. 12, pp. 54–59, Dec. 2003.

[45] Y. Ishikawa and N. Umeda, "Capacity design and performance of call admission control in cellular CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 15, no. 8, pp. 1627–1635, Oct. 1997.

[46] V. Kawadia and P. Kumar, "A cautionary perspective on cross layer design," *IEEE Wireless Commun. Mag.*, vol. 12, no. 1, pp. 3–11, Feb. 2005.

[47] K. Kim and Y. Han, "A call admission control with thresholds for multi-rate traffic in cdma systems," in *Proc. IEEE Veh. Technol. Conf.*, vol. 2, Tokyo, Japan, 2000, pp. 830–834.

[48] L. Kleinrock and F. A. Tobagi, "Packet switching in radio channels: Part I - carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Trans. Commun.*, vol. COM-23, pp. 1400–1416, Dec. 1975.

[49] J. Kuri and P. Mermelstein, "Call admission on the uplink of a CDMA system based on total received power communications," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, 1999, pp. 1431–1436.

[50] Y. Kwok and V. K. N. Lau, *Wireless Networks*. Kluwer Academic Publishers, 2002.

[51] J. N. Laneman, "*Cooperative Diversity in Wireless Networks: Algorithms and Architectures*," Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2002.

[52] K. Li and X. Wang, "Cross-layer optimization for LDPC-coded multirate multiuser systems with QoS constraints," *IEEE Trans. Signal Processing*, vol. 54, no. 7, pp. 2567–2578, July 2006.

[53] S. Lin and D. Costello, *Error Control Coding-Fundamentals And Applications*. Prentice-Hall, 1983.

[54] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746–1755, Sep. 2004.

[55] T. Liu and J. Silvester, "Joint admission/congestion control for wireless CDMA systems supporting integrated services," *IEEE J. Select. Areas Commun.*, vol. 16, no. 6, pp. 845–857, Aug. 1998.

[56] Z. Liu and M. E. Zarki, "SIR-based call admission control for DS-CDMA cellular systems," *IEEE J. Select. Areas Commun.*, vol. 12, no. 4, pp. 638–644, May 1994.

[57] A. F. Naguib, A. Paulraj, and T. Kailath, "Capacity improvement with base-station antenna arrays in cellular cdma," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 691–698, Aug. 1994.

[58] T. Ojanpera, *WCDMA: Towards IP Mobility and Mobile Internet*. Artech House Publishers, 2001.

[59] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks*. Prentice Hall, 2002.

[60] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in DS/CDMA system," *IEEE J. Select. Areas Commun.*, vol. 12, no. 5, pp. 796–807, June 1994.

[61] K. Pedersen and P. Mogensen, "Directional power-based admission control for WCDMA systems using beamforming antenna array systems," *IEEE Trans. Veh. Technol.*, vol. 51, no. 6, pp. 1294–1303, Nov. 2002.

[62] H. V. Poor and S. Verdu, "Single-user detectors for multiuser channels," *IEEE Trans. Commun.*, vol. 36, pp. 50–60, Jan. 1988.

[63] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[64] S. V. R. Lupas, "Near-far resistance of multi-user detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 496–508, Apr. 1990.

[65] R. M. Rao, C. Comaniciu, T. Lakshman, and H. V. Poor, "Call admission control in wireless multimedia networks," *IEEE Signal Processing Mag.*, vol. 21, no. 5, pp. 51–58, Sep. 2004.

[66] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, New Jersey: Prentice Hall, 2002.

[67] S. M. Ross, *Stochastic Process*. John Wiley and Sons, 1983.

[68] A. Sampath, P. S. Kumar, and J. M. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, vol. 1, 1995, pp. 21–25.

[69] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," *ACM/Baltzer Wireless Networks Journal*, vol. 8, no. 1, pp. 13–26, Jan. 2002.

[70] W. Sheng and S. D. Blostein, "Cross-layer call admission control policies for CDMA systems with beamforming," in *Proc. 23rd Bienniel Symposium on Communications*, Kingston, Canada, 2006.

[71] ——, "Reduced-outage-probability algorithms for cross-layer call admission control in CDMA beamforming systems," in *Proc. IEEE Global Telecommun. Conf.*, San Francisco, CA, 2006.

[72] ——, "Cross-layer admission control policy for CDMA beamforming systems," *EURASIP Journal on Wireless Commun. and Netw., Special Issue on Smart Antennas*, vol. 2007, July 2007.

[73] S. Singh, V. Krishnamurthy, and H. V. Poor, "Integrated voice/data call admission control for wireless DS-CDMA systems," *IEEE Trans. Signal Processing*, vol. 50, no. 6, pp. 1483–1495, June 2002.

[74] J. Skwirzynski, *New Concepts in Multi-User Communication*. The Netherlands: NATO Advanced Study Institutes, 1981.

[75] G. Song and K. Gong, "Approximate formula for evaluating erlang capacity of CDMA systems with smart antennas," *Electronics Lett.*, vol. 36, no. 12, pp. 1001–1002, June 2000.

[76] ——, "Performance comparison of optimal beamforming and spatial matched filter in power-controlled CDMA systems," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, 2002, pp. 455–459.

[77] B. C. Sowden and K. W. Sowerby, "Variable-bit-rate traffic modelling for dimensioning 3G CDMA systems," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, 2003, pp. 670–674.

[78] A. S. Tanenbaun, *Computer Networks*. Prentice Hall, 2003.

[79] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Technical Report, AT&T Bell Labs*, 1995.

[80] E. Tiirola and J. Ylitalo, "Performance evaluation of fixed-beam beamforming in WCDMA downlink," in *Proc. IEEE Veh. Technol. Conf.*, vol. 2, Tokyo, Japan, 2000, pp. 700–704.

[81] H. C. Tijms, *Stochastic Modelling and Analysis: A Computational Approach*. Wiley Series In Probability And Mathematical Statistics, 1986.

[82] V. Tripathi, E. Visotsky, R. Peterson, and M. Honig, “Reliability-based type II hybrid ARQ schemes,” in *Proc. IEEE Int. Conf. Commun.*, vol. 4, Anchorage, Alaska, USA, 2003, pp. 2899–2903.

[83] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[84] M. K. Varanasi and B. Aazhang, “Multistage detection in asynchronous code-division multiple-accesscommunications,” *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 509–519, Apr. 1990.

[85] B. D. V. Veen and K. M. Buckley, “Beamforming: A versatile approach to spatial filtering,” *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[86] S. Verdu, *Multiuser Detection*. New York: Cambridge University Press, 1998.

[87] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communications*. Reading, MA: Addison-Wesley, 1995.

[88] L. Wang and W. Zhuang, “A call admission control scheme for packet data in CDMA cellular communications,” *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 406–416, Feb. 2006.

[89] X. Wang and H. V. Poor, “Space-time multiuser detection in multipath CDMA channels,” *IEEE Trans. Signal Processing*, vol. 47, no. 9, Sep. 1999.

[90] E. J. Weldon, “An improved selective repeat ARQ strategy,” *IEEE Trans. Commun.*, vol. COM-30, pp. 480–486, 1982.

[91] J. Winters, “Optimum combining in digital mobile radio with cochannel interference,” *IEEE J. Select. Areas Commun.*, vol. 2, no. 4, pp. 528–539, July 1984.

[92] J. Wu, *Wireless Personal Communications: Performance Analysis of QoS-Based Voice/Data CDMA Systems*. Springer Netherlands, 2004.

[93] A. M. Wyglinski and S. D. Blostein, “On uplink CDMA cell capacity: Mutual coupling and scattering effects on beamforming,” *IEEE Trans. Veh. Technol.*, vol. 52, no. 2, pp. 289–304, Mar. 2003.

[94] M. Xiao, N. B. Shroff, and E. K. P. Chong, “Resource management in power-controlled cellular wireless systems,” *Wireless Communications and Mobile Computing*, vol. 1, no. 2, pp. 185–199, 2001.

[95] W. Yang and E. Geraniotis, “Admission policies for integrated voice and data traffic in CDMA packet radio networks,” *IEEE J. Select. Areas Commun.*, vol. 12, no. 4, pp. 654–664, May 1994.

[96] A. Yener, R. D. Yates, and S. Ulukus, “Joint power control, multiuser detection, and beamforming for CDMA systems,” in *Proc. IEEE Veh. Technol. Conf.*, vol. 2, Houston, TX, USA, 1999, pp. 1032–1036.

[97] ——, “Interference management for CDMA systems through power control, multiuser detection, and beamforming,” *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1227–1239, July 2001.

[98] ——, “Combined multiuser detection and beamforming for CDMA dystems: Filter structures,” *IEEE Trans. Veh. Technol.*, vol. 51, no. 5, pp. 1087–1095, Sep. 2002.

[99] F. Yu, V. Krishnamurthy, and V. C. M. Leung, “Cross-layer optimal connection admission control for variable bit rate multimedia trafiic in packet wireless CDMA networks,” *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 542–555, Feb. 2006.

[100] D. Zhao, “*Radio Resource Management in Cellular CDMA Systems Supporting Heterogeneous Services*,” Ph.D. dissertation, Department of Electrical and Computer Engineering, Waterloo University, 2002.

Appendix A

Simulation Implementation

In the thesis, we give the simulation results in terms of blocking probability, connection delay, outage probability, system throughput, etc. In this appendix, we discuss how to obtain these simulation results. We first present the simulation procedure for a dynamic system with random arrival and departure processes, and then briefly discuss how to obtain the simulation results for a dynamic system.

A.1 Dynamic system simulation

For a dynamic system with random arrival and departure processes, state transitions are triggered by call arrivals or departures. The dynamic system with Poisson arrival and exponential duration can be simulated as follows:

1. Initial time index $i = 0$;

2. System state can be represented by the number of accepted calls for class j and the number of calls waiting in the queue for class j, denoted by n_s^j and n_q^j, respectively, where $j = 1, .., J$;

3. Given current time instant t_i and current system state, the time duration until the next

decision epoch can be simulated through a uniformly distributed random variable,

$$\Delta t = -log(u)/(\sum_{j=1}^{J}\lambda_j + \sum_{j=1}^{J}\mu_j n_s^j)$$

where u denotes a uniformly distributed random variable over range $(0,1)$, and log denotes natural logarithm. This is a direct result from the fact that Δt has an exponential distribution with mean $\frac{1}{\sum_{j=1}^{J}\lambda_j+\sum_{j=1}^{J}\mu_j n_s^j}$;

4. At next decision epoch, a class j call arrives with probability $\frac{\lambda_j}{\sum_{j=1}^{J}\lambda_j+\sum_{j=1}^{J}\mu_j n_s^j}$, and a class j call departs with probability $\frac{\mu_j n_s^j}{\sum_{j=1}^{J}\lambda_j+\sum_{j=1}^{J}\mu_j n_s^j}$;

5. Update time index $i = i+1$, and the next decision epoch t_{i+1} can be obtained by $t_{i+1} = t_i + \Delta t$.

6. Repeat step 2 to step 5 until i is equal to the maximum number of decision epoches N.

A.2 Evaluate the performance in a dynamic system

We have discussed how to simulate a dynamic system. Now we give detailed steps on evaluating the system performance in a dynamic system.

The simulation results in terms of blocking probability, denoted by P_b^j, can be expressed as

$$P_b^j = \frac{\text{The number of blocked calls for class } j \text{ over duration } T}{\text{The number of total arriving calls for class } j \text{ over duration } T}. \qquad \text{(A.1)}$$

The connection delay can be equivalently represented by the average queue length, denoted by $n_q^{j,av}$, which can be obtained by

$$n_q^{j,av} = \frac{1}{N}\sum_{i=1}^{N} n_q^j(t_i)$$

where $n_q^j(t_i)$ denotes the queue length at time instant t_i, and N is the maximum number of decision epoches. In this thesis, N is chosen to be $50,000$ to ensure the accuracy of the evaluated performance.

The outage probability, denoted by P_{out}^{av}, can be evaluated by

$$P_{out}^{av} = \frac{1}{N}\sum_{i=1}^{N} P_{out}(\mathbf{s}(t_i))$$

in which $P_{out}(\mathbf{s}(t_i))$ denotes the outage probability at time instant t_i, obtained by

$$P_{out}(s(t_i)) = \begin{cases} 1 & \text{if } \upsilon(QF) \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

where $\upsilon(.)$ denotes maximum eigenvalue, and matrix Q and F are defined in Chapter 3.

With the above expressions, the system performance can be evaluated by following the steps illustrated in Table A.1, which consists two parts: a dynamic system simulator based on the approach discussed in the previous section, and a system performance recorder for $50,000$ decision epoches, i.e., $50,000$ arrival and departure instants. Table A.1 also employs the processing steps for arrival and departure events, which are illustrated in Table A.2 and A.3, respectively.

Table A.1. Evaluate the performance by simulation

System Initialization	Set $N_B^j = 0$, $N_j = 0$, where N_B^j and N_j denote the number of blocked calls and the number of totally arriving calls for class j, respectively, $j = 1, .., J$.
Step 0:	Set time index $i = 0$, and the initial state is denoted by $s(t_0)$. Set $n_s^j(t_0) = 0$ and $n_q^j(t_0) = 0$.
Step 1:	Current system state can be obtained by $s(t_i) = [n_q^1(t_i), n_s^1(t_i), .., n_q^J(t_i), n_s^J(t_i)]$.
Step 2:	From policy lookup table, choose an action **a** for the current state $\mathbf{s}(t_i)$. $\mathbf{a} = [a_1, .., a_J, .., d_1, .., d_J]$, where a_j decides if the incoming call for class j can be accepted, and d_j decides if the call waiting in the class j queue can be transmitted.
Step 3:	Use dynamic system simulator discussed in A.1 to decide the incoming event and the transition duration. Time index is updated to $i = i + 1$.
Step 4:	If an arrival event occurs, follow the arrival processing procedure illustrated in Table A.2; If a departure event occurs, follow the departure processing procedure illustrated in Table A.3.
Step 5:	The system state $s(t_i))$ as well as parameters N_b^j, N_j, $n_q^j(s(t_i))$ and $P_{out}(s(t_i))$ are updated.
Step 6:	Repeat step 1 to step 5 until $i = N$.
Step 7:	Evaluate the performance by time averaging: $P_b^j = \frac{N_b^j}{N_j}$, $P_{out}^{av} = \frac{1}{N}\sum_{i=1}^{N} P_{out}(s(t_i))$, $n_q^{j,av} = \frac{1}{N}\sum_{i=1}^{N} n_q^j(t_i)$.

Table A.2. Arrival processing procedure

Step 1:	Set $N_j = N_j + 1$. The action is given by $\mathbf{a} = [a_1, .., a_J, .., d_1, .., d_J]$.
Step 2:	If $a_j = 1$, go to step 3; Otherwise, go to step 4.
Step 3:	Update the system state $s(t_i)$ with $n_s^j(t_i) = n_s^j(t_{i-1}) + 1$ and $n_q^j(t_i) = n_q^j(t_{i-1})$. Go to Step 5.
Step 4:	If the current queue buffer is full, set $N_b^j = N_b^j + 1$, and the system state remains unchanged, i.e., $s(t_i) = s(t_{i-1})$; Otherwise, update the system state with $n_q^j(t_i) = n_q^j(t_{i-1}) + 1$ and $n_s^j(t_i) = n_s^j(t_{i-1})$.
Step 5:	Evaluate $P_{out}(s(t_i))$ based on (A.2).

Table A.3. Departure processing procedure

Step 1:	The action is given by $\mathbf{a} = [a_1, .., a_J, .., d_1, .., d_J]$.
Step 2:	If $d_j = 1$, go to step 3; Otherwise, go to step 4.
Step 3:	If the current queue buffer is not empty, update the state with $n_q^j(t_i) = n_q^j(t_{i-1}) - 1$ and $n_s^j(t_i) = n_s^j(t_{i-1})$, then go to step 5; otherwise, go to step 4.
Step 4:	Update the system state $s(t_i)$ with $n_s^j(t_i) = n_s^j(t_{i-1}) - 1$, and $n_q^j(t_i) = n_q^j(t_{i-1})$.
Step 5:	Derive $P_{out}(s(t_i))$ based on (A.2).

Appendix B

Exact Outage Probability for Single Antenna Systems Employing Voice Activity

In this appendix, we derive an exact outage probability for single antenna systems by considering voice activity.

As shown in (2.6), power control feasibility condition (PCFC) for single antenna systems, which represents the number of simultaneously transmitted users, can be derived as [68]

$$\sum_{j=1}^{J} \frac{n_s^j}{D_j} \leq 1 \tag{B.1}$$

where $D_j = 1 + \frac{W}{\gamma_j R_j}$.

The PCFC in (B.1) represents the user capacity without employing voice activity, in which the SIR requirements of all users can be satisfied, and as a result, can achieve zero outage probability.

Let κ_i denote a binary random variable representing user i's voice activity indicator, where $i = 1, .., K_1$. It is assumed that κ_i's have independent and identical distributions with success rate p_v, which can be expressed as $p_v = \Pr\{\kappa_i = 1\} = 1 - \Pr\{\kappa_i = 0\}$.

By allowing outage probability and employing voice activity, the above PCFC can be revised as

$$\sum_{i=1}^{n_s^1} \kappa_i \frac{1}{D_1} + \sum_{j=2}^{J} \frac{n_s^j}{D_j} < 1. \tag{B.2}$$

Given a system state $s = [n_q^1, n_s^1, .., n_q^J, n_s^J]$, outage probability, denoted by P_{out}, can be obtained by

$$\begin{aligned} P_{out}(s) &= \Pr\left\{ \sum_{i=1}^{n_s^1} \frac{\kappa_i}{D_1} + \sum_{j=2}^{J} \frac{n_s^j}{D_j} \geq 1 \right\} \\ &= \Pr\left\{ \sum_{i=1}^{n_s^1} \kappa_i \geq D_1 \left(1 - \sum_{j=2}^{J} \frac{n_s^j}{D_j} \right) \right\}. \end{aligned} \tag{B.3}$$

Denote $\mathbf{x} = \sum_{i=1}^{n_s^1} \kappa_i$, and $A = D_1 \left(1 - \sum_{j=2}^{J} \frac{n_s^j}{D_j} \right)$. The above probability can be further expressed as

$$\begin{aligned} P_{out}(s) &= \Pr\{\mathbf{x} > A\} \\ &= \begin{cases} 1 & \text{if } n_s^2 > D_2 \cup n_s^3 > D_2 \cup \ldots \cup n_s^J > D_J \\ 0 & \text{if } n_s^1 < D_1 \left(1 - \sum_{j=2}^{J} \frac{n_s^j}{D_j} \right) \\ 1 - F_{\mathbf{x}}(A) & \text{otherwise} \end{cases} \end{aligned} \tag{B.4}$$

in which $\cup$ denotes union, and $F_{\mathbf{x}}(.)$ denotes the cumulative density function (CDF) of random variable $\mathbf{x}$. Since κ_i has Bernoulli distribution with success rate p_v, random variable $\mathbf{x}$ has a Binomial distribution. The CDF of $\mathbf{x}$ can be further expressed as

$$\begin{aligned} F_{\mathbf{x}}(A) &= \sum_{m=0}^{\llcorner A \lrcorner} f_{\mathbf{x}}(m) \\ &= \sum_{m=0}^{\llcorner A \lrcorner} \binom{n_s^1}{m} p_v^m (1 - p_v)^{n_s^1 - m} \\ &= I_{1-p_v}\left(n_s^1 - \llcorner A \lrcorner, \llcorner A \lrcorner + 1 \right) \end{aligned} \tag{B.5}$$

where $\llcorner a \lrcorner$ denotes the maximum integer less than or equal to a, and $I_p(c, b)$ represents a regularized incomplete beta function with parameters p, c, b. In the above, Equation (B.5) is directly obtained from the definition of a regularized incomplete function [1].

Therefore, the outage probability in (B.4) can be expressed as

$$P_{out}(s) = \begin{cases} 1 & \text{if } n_s^2 > D_2 \cup n_s^3 > D_2 \cup \ldots \cup n_s^J > D_J \\ 0 & \text{if } n_s^1 < D_1 \left(1 - \sum_{j=2}^{J} \frac{n_s^j}{D_j}\right) \\ 1 - I_{1-p_v}\left(n_s^1 - \llcorner A \lrcorner, \llcorner A \lrcorner + 1\right) & \text{otherwise.} \end{cases} \tag{B.6}$$

Appendix C

Derivation of an Approximate Outage Probability

As presented in (5.9), the outage probability for a given system state s by including the impact of ARQ can be represented by

$$P_{out}(s) = \text{Prob}\{\upsilon(QF) \geq 1\} \tag{C.1}$$

where Prob$\{A\}$ denotes the probability of event A, $\upsilon(.)$ denotes the maximum eigenvalue, Q is a K-dimensional diagonal matrix with the i^{th} non-zero element as $\frac{\gamma_i R_i}{W}$, $i = 1, .., K$, and F is a K by K matrix in which the element at the i^{th} row and the j^{th} column can be expressed as $F_{ij} = \frac{\phi_{ij}^2}{\phi_{ii}^2}$ for $i \neq j$, and $F_{ij} = 0$ for $i = j$.

Owing to the properties of nonnegative matrices, the eigenvalue $\upsilon(QF)$ can be estimated by [76],

$$\begin{aligned}
\upsilon(QF) &= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{K}(QF)_{i,j} \\
&= \frac{1}{\sum_{j=1}^{J} n_s^j}\sum_{j=1}^{J}\left[\frac{R_j\gamma_j}{W}\sum_{m=t_j+1}^{t_j+n_s^j}\sum_{i=1}^{K}F_{mi}\right]
\end{aligned}$$

where $(QF)_{i,j}$ denotes the entry of matrix QF at the i^{th} row and the j^{th} column, and

$$t_j = \begin{cases} \sum_{l=1}^{j-1} n_s^l & \text{if} \quad j > 1 \\ 0 & \text{if } j = 1. \end{cases} \tag{C.2}$$

As shown in [75], by using the central limit theorem, random variable $\upsilon(QF)$ has an approximately Gaussian distribution. Therefore, the outage probability in (C.1) becomes

$$\begin{aligned} P_{out}(s) &= \text{Prob}\{\upsilon(QF) \geq 1\} \\ &= Q\left[\frac{1-E[\upsilon]}{\sqrt{Var(\upsilon)}}\right] \end{aligned} \tag{C.3}$$

where $Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-\frac{x^2}{2}} dx$, $E[\upsilon]$ and $Var(\upsilon)$ denote the expectation and variance of random variable $\upsilon(QF)$, which can be obtained as follows:

$$\begin{aligned} E[\upsilon] &= \sum_{j=1}^{J} \frac{1}{W} \gamma_j R_j n_s^j E[F] \\ Var[\upsilon] &= \frac{1}{K} \sum_{j=1}^{J} n_s^j [\frac{1}{W} \gamma_j R_j]^2 Var[F] \end{aligned} \tag{C.4}$$

where $E[F]$ and $Var[F]$ denote the expectation and variance of F_{ij}, which can be evaluated numerically. Table 5.4 presents these numerical values for a uniform circular array, which are derived in [75].

The accuracy of the above outage probability approximation has been shown in [75].