DESIGN AND ANALYSIS OF OPTIMAL RESOURCE ALLOCATION POLICIES IN WIRELESS NETWORKS

WANG BANG

NATIONAL UNIVERSITY OF SINGAPORE

2004

DESIGN AND ANALYSIS OF OPTIMAL RESOURCE ALLOCATION POLICIES IN WIRELESS NETWORKS

BY

WANG BANG

(*M.Eng., B.Eng., HUST, PRC*)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHYLOSOPHY
DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2004

Dedication

To my Mama and Papa and to my wife Minghua

Acknowledgements

I would like to take this opportunity to express my deepest thanks to many who have contributed to the production of this thesis. Without their support, this thesis could not have been written. My thesis advisor, Associate Professor Chua Kee Chaing, has my sincerest gratitude. Both this thesis and my personal development have been benefited greatly from his guidance, advices, encouragements, rigorous research style. I feel fortunate to have been his student.

I would like to thank the Department of Electrical and Computer Engineering and the National University of Singapore for the kind offer of a research scholarship. Also, I thank Siemens ICM for providing a chance to have worked on an industrial project in Munich Germany. I meet many wonderful colleagues, among whom I specially thank Dr. Robert Kutka and Dr. Hans-Peter Schwefel for their kind help when working in Munich.

I must also thank my parents, my wife and my parents in law for their constant caring and support. My sincere thanks to my wife, Xu Minghua, whose endless and selfless love is always an important part of my life. My deepest thanks go to my parents in China for their prayerful supports in my decision to go on to graduate study in Singapore.

Finally, I would like to express my gratitude to my colleagues and friends in Open Source Software laboratory for providing hearty help and happy hours.

Contents

List of Figures

List of Tables

List of Abbreviations

3G	The Third Generation Mobile Systems
ACOE	Average Cost Optimal Equation
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BS	Base Station
CAC	Connection Admission Control
CDMA	Code Division Multiple Access
CSR	Channel State Report
ETSI	European Telecommunications Standards Institute
FDMA	Frequency Division Multiple Access
FSP	Frame Success Probability
GPS	Generalized Processor Sharing
GSM	Global System for Mobile communication
ITU	International Telecommunication Union
MDP	Markov Decision Processes
MS	Mobile Station
QoS	Quality of Services
RLC	Radio Link Control
SIR	Signal-to-Interference Ratio
SSP	Stochastic Shortest Path
TDMA	Time Division Multiple Access
UTRA	UMTS Terrestrial Radio Access
UMTS	Universal Mobile Telecommunication System
WCDMA	Wideband CDMA

List of Symbols

$*$	Superscript as optimal indication, for example, π^*, optimal policy
a	An action
A	The largest action, e.g., the largest number of packets can be sent in a frame
$\mathcal{A}_s$	Set of available actions in state s
$\mathcal{A}$	System action space, $\mathcal{A} = \bigcup_{s \in \mathcal{S}} \mathcal{A}_s$
B	Buffer limit. $B < \infty$, a finite buffer; $B = \infty$, an infinite buffer
$C(s, a)$	Cost structure when the state is s and the action a is selected.
d_t	Decision rule at epoch t
$\mathbb{E}_s^\pi$	Expected value with respect to policy π conditional on starting state s
f_s	Average frame success probability, either a function or a scalar
$\mathbb{N}$	Set of positive integers
$\mathbb{N}^+$	Set of non-negative integers
$q(i)$	Probability of i packets arriving in a frame
Q	Geometric distribution parameter
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^+$	Set of non-negative real numbers
$\mathcal{S}$	System state space
s	A scalar system state
$\mathbf{s}$	A system state when consisting of more than one component
s_t	State of the system at decision epoch t
$\mathcal{T}$	Set of decision epochs
t	A decision epoch, often used as a subscript

$\mathrm{Tr}(s'\|s,a)$	Transition probability that the system occupies state s' at the next decision epoch if the current state is s and action a is chosen
$V^{\pi}(s)$	Value of policy π with the expected total cost optimal criterion and starting state s
$V_{\rho}^{\pi}(s)$	Value of policy π with the expected total discount cost optimal criterion and starting state s
$\overline{V}^{\pi}(s)$	Value of policy π with the expected total average cost optimal criterion and starting state s
π	Policy $(d_0, d_1, \cdots, d_{T-1})$, $T \leq \infty$
π^*	Optimal policy
Π	Set of all (allowable) policies
μ	Stationary policy $(d, d, \cdots)$ where $d_t = d$ for all t
ρ	Discount factor, $0 \leq \rho < 1$
λ	Average number of arrivals in a frame, $\lambda = \sum_{i=0} iq(i)$

Abstract

Wireless communications have been progressing steadily in recent years. It is expected that data traffic generated by services such as web surfing, file transfer, emails and multimedia message services will be dominant in next generation mobile networks. Radio resource management is very important in that it improves the resource utilization efficiency while meeting Quality of Service (QoS) requirements. This thesis studies the design of optimal resource allocation policies for data services in wireless networks. In particular, this thesis investigates the following resource management issues: power allocation, transmission control and rate allocation. We first study these issues separately from a single user point of view and then jointly from a system viewpoint.

A set of problems is modelled from the stochastic decision theoretic framework and solved by using the Markov decision processes (MDP) mathematical tool. We first consider a power allocation problem for transfer of a file by a single sender in a Rayleigh fading channel. The objective is to minimize the energy required for transfer of the file while meeting a delay constraint. We show how to convert such a constrained stochastic optimization problem with an average delay constraint to a standard Markov decision problem via a Lagrangian approach. It is observed from our numerical results that transmission power can be substantially reduced with optimal policies which exploit knowledge of the channel variations to meet the delay constraint.

We next consider a transmission control problem over a time-varying channel and with general arrival statistics. We show the existence of average cost optimal policies and explore the properties of the optimal policies. The resulting optimal policies are proved to have a structural property: when the buffer occupancy is low, the sender can suspend transmission in some bad channel states to save transmission power; however, when the buffer occupancy exceeds some thresholds, the sender has to transmit even in some bad channel states to avoid increasing the delay cost. We evaluate how the channel characteristic affects the resulting optimal policies via extensive simulations.

We also consider a rate allocation problem. We prove that the resulting optimal policies have a monotone property, i.e., the optimal action is nondecreasing with the system state. We analyze two extreme policies which provide the upper and lower delay bounds based on the stochastic process comparison technique. A class of one-threshold based simple policies are proposed to approximate the optimal policy and a tight delay bound is proved. We also extend the rate control problem against the existence of competitions across users. We then identify the characteristic of the value functions and the property of optimal policies for such an extended problem

When allocating resource among multiple users, fairness among users is also important in addition to system utilization efficiency. We propose a new fairness model, the *fair-effort resource sharing* model, and a simple credit based algorithm to implement the proposed fairness model. According to our fairness model, the resource share (quota) allocated to a user is proportional to the user's effort which is considered as time dependent rather than as fixed. We then present an integrated packet level resource allocation scheme which consists of optimal power allocation, exhaustive instantaneous data rate allocation and fair-effort resource sharing. Numerical results show that fair-effort based fairness is guaranteed with our proposed scheme and that system efficiency is improved compared to a scheme based on the generalized processor sharing fairness model.

Chapter 1

Introduction

Cellular mobile communications have been progressing steadily in recent years, from the first and second generation systems to the third generation systems (3G). The services that can be supported have also evolved from pure voice service to multimedia services, including voice, video and data. It is expected that data traffic generated by services such as web surfing, file transfers, emails and multimedia message service will be dominant in next generation mobile networks. As different services have different Quality of Service (QoS) requirements, e.g., delay, error rate, etc., compared to pure voice services, more flexibility in allocating the radio resources to meet these diverse requirements is needed. However, radio resources are scarce due to the limited radio spectrum. Therefore, how to efficiently utilize/allocate radio resources and simultaneously to provide the required QoS guarantees is an important topic of research to enable mobile networks to support heterogeneous services.

1.1 Cellular Mobile Communications

In cellular mobile communications, the geographical area covered by the whole system is divided into several contiguous small areas (*cells*) in which multiple mobile stations (MS) communicate with a central base station (BS) [65], as shown in Fig. 1.1. When several mobile stations (mobile users) wish to communicate with the base station through a common channel, multiple access techniques are used to coordinate the communications between the mobile stations and the base station. Common multiple

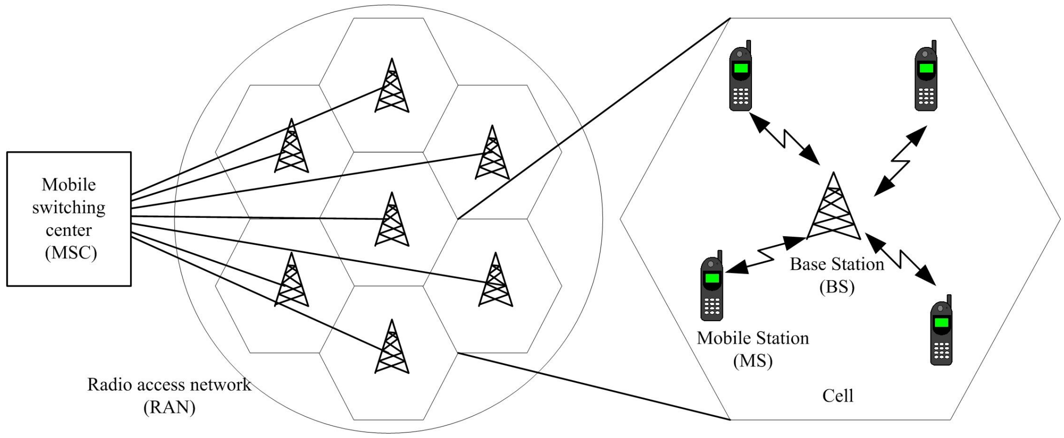

Figure 1.1: System model of cellular mobile communications

access techniques are:

- Frequency Division Multiple Access (FDMA)

  The radio spectrum is divided into separate frequency bands (channels). Each mobile station is assigned a unique frequency channel upon successful request, which is not used by others during the whole course of its communication (connection holding time). Multiple users can communicate with the base station simultaneously by using different frequency bands.

- Time Division Multiple Access (TDMA)

  The time axis is divided into several contiguous timeslots. In each timeslot, only one user can transmit. However, a user can transmit in several consecutive timeslots (in the same frame) to obtain a high transmission rate via slots aggregation. Thus multiple users communicate with the base station through a common frequency channel but in a time-slotted manner.

- Code Division Multiple Access (CDMA)

  Each user is assigned a unique code which the base station uses to separate different users. The codes are used for either modulating the radio waves or changing the carrier frequency, i.e., *spreading* the information radio waves. Multiple users share the same bandwidth and thus can transmit simultaneously.

The same frequencies and timeslots can be reused in different cells by using FDMA and TDMA if the distance between the base stations are large enough and interference of the same frequency bands is negligible. Hence more users can be supported and radio utilization efficiency can be improved in a mobile system. By using CDMA, the information bearing signal is spread over a bandwidth larger than the signal itself. Although it is not spectrally efficient for a single user, a CDMA system becomes bandwidth efficient in the multiple user case since it is possible for multiple users to share the same spreading bandwidth at the same time. Usually, FDMA is used together with TDMA or CDMA to separate the spectrum into smaller bands which are then divided in a time or code division manner. The above fundamental techniques can be used together to form various hybrid schemes.

Some second generation digital cellular systems, such as the Global System of Mobile Telecommunications (GSM), employ a simple form of TDMA scheme that assigns fixed timeslots to mobile users to support digital voice services. Timeslots aggregation can be used to support multi-rate services in second generation systems. Most third generation mobile networks will be based on the Wide-band CDMA (W-CDMA) technique. However the TDMA component has also been incorporated in 3G standards.

1.1.1 3G and UMTS

The third generation mobile communication system (3G) is standardized and defined by International Telecommunication Union (ITU) as IMT-2000 (International Mobile Telecommunication). It comprises a set of standards and recommendations. In Europe, the 3G system is called Universal Mobile Telecommunication System (UMTS) [40], which has been specified by the European Telecommunication Standards Institute (ETSI). The UMTS Terrestrial Radio Access (UTRA) consists of two operational modes, a frequency division duplex (FDD) mode, and a time division duplex (TDD) mode [13]. Wideband Code Division Multiple Access (WCDMA) is used for UTRA FDD and Time Division-Code Division Multiple Access (TD-CDMA) is used for UTRA TDD. UTRA FDD uses different frequency bands for uplink and downlink, separated by the duplex distance, while UTRA TDD utilizes the same frequency for both uplink and downlink. UTRA FDD and TDD are harmonized with respect to the basic system parameters such as carrier spacing, chip rate and frame length and hence FDD/TDD dual mode operation can be facilitated. UMTS is a hybrid system which enables the use of FDMA, TDMA and CDMA and their combinations. For more specifications of UMTS, refer to the ETSI standard series and the text edited by Holma [40].

The most important feature of the UMTS is its high data rate capability, which is usually summarized as 144 kbps for vehicular speeds, 384 for pedestrian speeds and 2Mbps for indoor environments. Other main features include global roaming, diverse services, Internet connection, easy and flexible service bearer configuration, etc. Finally, we note that both circuit switching and packet switching are allowable in UMTS. Hence advanced and flexible QoS can be supported.

1.2 Resource Allocation in Wireless Networks: Challenges and Issues

As mentioned earlier, next generation mobile networks need to support heterogeneous services with different QoS requirements. For example, voice services have strict delay requirements while data services may tolerate some delays. Although this feature lends the 3G networks to efficiently utilize resources, it also complicates the design of resource management policies. On the other hand, due to the hostile transmission medium in wireless communications, the resource allocation policy should also be fine tuned to balance between the transmission quality, e.g., meeting the minimum error rate requirement, and the cost to achieve the quality requirement, e.g., using the least transmission power. In this section, we briefly overview the wireless QoS issue in the context of UMTS, summarize the characteristic of the radio channel and introduce some resource management modules.

1.2.1 Wireless Services and QoS Issues in UMTS

UMTS defines *bearer service* as the abstraction of the capability for information transfer between access points [20]. The information transfer capabilities and transfer qualities are the two main requirements for bearer services. The characterization of a bearer service is made by using a set of characteristics, which include traffic type (realtime/non-realtime), traffic characteristics (uni-/bi-directional, broadcast, multicast), information quality (delay, delay jitter, error rate, data rate) and so on. UMTS allows a user (or application) to negotiate bearer characteristics that are most appropriate for its information transfer. It is also possible to change bearer characteristics via a bearer re-negotiation procedure during an ongoing connection. UMTS uses a layered structure to map an end-to-end network service into several bearer services [21]. The end-to-end QoS is thus split into several parts and each part should be supported by one bearer service. The lowest bearer service that covers all aspects of the radio interface transport is the *radio bearer service*, which uses the UTRA FDD/TDD services.

UMTS defines four kinds of QoS classes (traffic classes) [21]. They are: *conversational*, *streaming*, *interactive* and *background* class. The main distinguishing factor between these QoS classes is how delay sensitive the traffic is. Conversational class is meant for traffic which is very delay sensitive while background class is the most delay insensitive traffic class. The first two classes are those real-time traffic which needs to preserve time relation (variation) between information entities of the stream. The last two classes are those best-effort traffic which needs to preserve payload content. A summary of the major groups of example applications in terms of QoS requirements is shown in Fig. 1.2, in which the delay values represent the one-way delay [20]. Applications may be applicable to one or more groups.

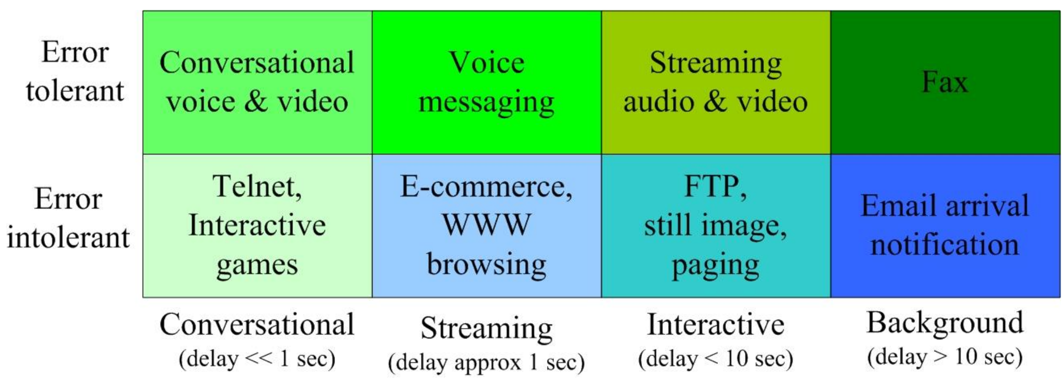

Figure 1.2: UMTS QoS classes and example allocations [22].

In UMTS, the QoS *attributes* define some typical parameters (e.g., delay/loss ratio) for each QoS class. The QoS attributes are used to compose a QoS profile for negotiation of the bearer service between the end user and the network. The specification of UMTS QoS attributes is still ongoing and especially, the bit rate attributes are under discussion. Different classes have different ranges of the value of some QoS attributes. Table 1.1 summarizes the typical values for some main QoS attributes. The *delivery order* indicates whether the service data unit (SDU) can be delivered in-sequence or not. The *residual bit error ratio* indicates the undetected bit error ratio in the delivered SDUs. The *transfer delay* is the maximum delay for the 95th percentile of the distribution of delay for all delivered SDUs during the life time of a bearer service. It is worth noting that the guaranteed bit rate and the transfer delay are not specified for the interactive and background classes according to UMTS specifications.

Table 1.1: Value ranges of UMTS radio bearer QoS attributes (adapted from [21])

QoS attributes	Conversational	Streaming	Interactive	Background
Maximum bit rate	< 2048 kbps	< 2048 kbps	< 2048 kbps	< 2048 kbps
Delivery order	yes/no	yes/no	yes/no	yes/no
Residual BER	$5 \times 10^{-2}, 10^{-2}$, $5 \times 10^{-3}, 10^{-3}$, $10^{-4}, 10^{-6}$	$5 \times 10^{-2}, 10^{-2}$, $5 \times 10^{-3}, 10^{-3}$, $10^{-4}, 10^{-5}, 10^{-6}$	4×10^{-4}, 10^{-5}, 6×10^{-9}	4×10^{-4}, 10^{-5}, 6×10^{-9}
SDU error ratio	$10^{-2}, 7 \times 10^{-3}$, $10^{-3}, 10^{-4}, 10^{-6}$	$10^{-1}, 10^{-2}, 10^{-4}$, $7 \times 10^{-3}, 10^{-5}$	$10^{-3}, 10^{-4}$, 10^{-6}	$10^{-3}, 10^{-4}$, 10^{-6}
Transfer delay	80ms-maximum	250ms-maximum		

The separation of the bearer service and QoS profile enables the flexible allocation and utilization of UMTS network resources. For example, a user (or an application) can request to use a lower data rate to save transmission cost during its connection holding time via the radio bearer negotiation procedure in UMTS. On the other hand, some distinguishing characteristics of wireless communications also complicate the resource management. The following subsection briefly overviews a main distinct feature particular to wireless communications.

1.2.2 Hostile Radio Channel

In a mobile radio environment, radio wave propagation suffers from attenuation between the mobile station and its serving base station. In general, the received signal strength is affected by antenna heights, local reflectors and obstacles. Furthermore, the user mobility pattern, i.e., the speed and the direction, also greatly impacts the received signal strength. In practice, the path loss cannot be assumed to be computed based on a simple free-space and line-of-sight model. However, some engineering models can be used. These engineering models are based on several wave propagation phenomena such as *reflection*, *diffraction* and *scattering* [65]. Reflection from an object typically occurs when the wavelength of an impinging wave is much smaller than the object itself, resulting in the multi-path components. Diffraction causes the wave to bend around obstacles and can be explained by Huygen's principle [65]. When a wave travels in a medium with a large number of elements having smaller dimensions compared to its wavelength, the energy is scattered. Although accurate prediction of ratio propagation is rather difficult, several engineering radio fading models are widely used in cellular mobile communications. The signal fading in a wireless environment is normally considered to contain three components with different time scales of variations. These are the *large-scale path loss*, *medium-scale slow fading* and *small-scale fast fading* [65]. Decreased received power with distance, reflection and diffraction constitute the path loss. These are denoted large-scale since changes appear when moving over hundreds of meters. A mobile station can be shadowed by, e.g., trees and buildings. The local mean received power changes when a user moves just a few tens of meters, i.e., on a medium-scale. Small-scale fast fading or multi-path fading characterizes the effect of multi-path reflections by local scatterers and changes by the order of wavelengths. For example, in the absence of a strong non-fading line-of-sight component, the Rayleigh fading model is often used, in which the envelope S of the received signal follows a Rayleigh distribution: $f_S(s) = \frac{s}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$, $s \geq 0$. Note that the received signal power S^2 follows the exponential distribution in this case.

The transmission quality of a connection (or an application) is closely related to the underlying channel conditions which determine the probability of successful receptions and hence determine the QoS of the connection. Many methods can be used to alleviate the harsh channel conditions, such as power control, error correction coding, interleaving and so on. However, there are always some costs associated with the method for alleviating channel conditions. We consider the following example. In wireless communications, the received signal to noise ratio (SNR, often in the context of TDMA) or signal to interference plus noise ratio (SINR or SIR, often in the context of CDMA) has a one-to-one mapping to the bit error rate (BER) given a fixed transmission scheme, i.e., fixed coding and modulation scheme, etc. Let γ denote the received SNR which can be simply computed as the ratio of the received signal power to the channel noise, S^2/σ^2. Then the famous Shannon capacity of an additive white Gaussian noise (AWGN) channel can be expressed as [61]:

$$C = \log_2(1 + \gamma) \ \text{bit/sec/Hz} \tag{1.1}$$

This can be interpreted by an increase of 3dB in SNR required for each extra bit per second per Hertz. Note that (1.1) can also be interpreted as how to maintain the received SNR for a fixed transmission rate requirement, i.e., adjusting transmission power according to the time-varying channel path gain. At first glance, increasing transmission power can improve the received SNR and hence improve the effective transmission rate of a connection. However, when we consider the transmission power as the cost to achieve the QoS requirements, it is of course better to use the least cost to achieve the same QoS requirements. Hence a better (or an optimal) resource management policy should also address the tradeoff between the QoS requirements and the costs to achieve the QoS. In the next subsection, we briefly introduce some resource management tools considered in this thesis.

1.2.3 Some Management Modules

Radio resource management [40, 96], which has always been an important research area in wireless communications, provides the mechanisms for efficient utilization of the limited and scarce radio resources while guaranteeing the diverse QoS requirements of different services. However, the design of a comprehensive resource management scheme is rather difficult and, sometimes, almost impossible. Nonetheless, we can identify different QoS levels each with appropriate QoS metrics, and further identify some management tools for each level accordingly. In general, we can classify the QoS requirements at three different levels: *class level*, *call level* and *packet level*. This example classification enables us to work at different levels of the QoS hierarchy independent of each other and facilitates us to identify the required management tools for each level. For example, at the call level, the channel allocation scheme and the handoff scheme are important management modules as they determine the call blocking and handoff dropping probabilities, the main call level QoS metrics. In this thesis, we focus on packet level resource management and in particular, we focus on the optimal policy design problem for data services.

At the packet level, we are mainly concerned with the following problems: when to transmit a packet (or when to transmit which packet), how much transmission power should be used and how many information bits (or data packets) should be transmitted in a transmission. Indeed, these problems represent three important modules of resource management at the packet level, i.e., *transmission scheduling*, *power control* and *rate allocation*. These problems can be solved either separately or jointly. Furthermore, these problems can also be solved either from a single user point of view or from the system point of view. For example, a centralized system operator decides which user should transmit next among multiple backlogged users. We next briefly review the main functionality of each module.

Transmission scheduling From a single user's point of view, transmission scheduling determines the times for transmitting the head of line packet in the (sorted) buffer. Transmission scheduling can be used to exploit the variations of a wireless channel in that it can avoid transmitting in poor channel conditions. This may lead to energy savings but increases delay. However, a realtime packet should be transmitted before its deadline. From a system operator's point of view, transmission scheduling is used to decide which user (flow) should transmit next. Hence, transmission scheduling may (partly) determine the quota of the system resources allocated to each user and fairness (e.g., the max-min fairness [8]) among users is a basic objective in this case.

Power control Transmission power determines the probability of a successful reception of a packet. From a single user's point of view, power control is mainly for combatting the hostile radio channel. It, together with transmission scheduling, can achieve energy efficient transmissions. Power control is of particular importance in CDMA network in that it controls the total interference over the air and hence determines the achievable total system throughput.

Rate allocation As mentioned in the previous section, it is possible to change the transmission rate for a connection during its holding time. From a single user's point of view, it can choose to transmit with a high or low rate based on its demand, e.g., its buffer occupancy. From the system's point of view, rate allocation also determines how the system resources will be shared among different users.

In this thesis, we first consider the three management modules separately for a single user and put the three problems in the decision theoretic framework. We then study the three problems jointly and from a system operator's point of view. We review some related works for the two sets of problems in the next section.

1.3 Related Works

1.3.1 Optimal Policy Design

We focus on data services instead of realtime services throughout this thesis. In general, data services generate *elastic traffic* which are more delay tolerant than realtime traffic, cf. Section 1.2.1. At the packet level, delay tolerance often means that there is no strict deadline for a data packet to be transmitted. Hence there is more flexibility in allocating resources to data services. On the other hand, we may also exploit the channel variations for delay tolerant data services in that we may transmit data packets in an *opportunistic* way, e.g., not transmitting in bad channel conditions but waiting for better channel conditions to transmit later. However, it is also not appropriate that we totally neglect any delay requirement for data services. Instead, we can take the delay into consideration via some cost functions and provide statistical delay guarantees. As there may be many solutions to these problems, we need to find an optimal one and design the resource allocation policy accordingly.

A resource allocation policy prescribes the procedure of how to choose different actions, e.g., different transmission powers, according to the observed state, e.g., the channel conditions. Obviously, the design of a policy is determined by the design objective. It is desirable but almost impractical that a policy can perform best in all aspects. It is not uncommon that we have to face tradeoffs between different design objectives, e.g., reducing energy consumption vs. decreasing packet delay. To compare different policies, it is useful to assign some (real) value to each policy. Hence an optimal policy can be defined as the one that has the minimum (or maximum) policy value among all (allowable) policies. When the dynamics of the radio channel and/or the dynamics of the data sources are considered, a policy needs to consider not only the current outcome of the action but also the future action options. In the context of stochastic optimization, a Markov decision process (MDP) [7, 62] is such a useful mathematic tool that can be used for our resource allocation problems in that it not only considers stochastic dynamics but also assigns policy values. We defer the introduction of the Markov decision theory to the next chapter. Note that there may be other methods to compare policies, such as the commonly used linear programming and nonlinear programming methods. For example, A. Sampath et al., in their widely refereed paper [68], have applied the nonlinear programming modelling technique for power control and resource management in a CDMA network and recently, M. Soleimanipour et al. have applied a mixed integer nonlinear programming technique in the design of optimal resource management [74].

In this thesis, we apply MDP theory in policy design for the three allocation problems. Before going into our approaches, we mention some recent related works applying MDP theory in wireless resource allocation policy design at the packet level. In particular, researchers have applied MDP theory in the design of wireless transmission schemes each with a particular context and problem formulation [12, 37, 38, 39, 92, 93, 97, 98, 63, 64, 6, 32]. In [12], a user controls its target SIR for its head of line packet based on the estimated interference over the air in order to maximize a reward function each time it transmits a deadline-constrained packet. The resulting policy provides network layer QoS guarantees while increasing the system achievable total throughput in a saturated CDMA network. In [37, 38, 39], T. Holliday et al. apply the MDP theory to design optimal link adaptation policies for voice traffic in the context of both

TDMA and CDMA networks. The resulting optimal transmission polices prescribe optimal actions in terms of the choice of the modulation scheme, source coding scheme, and the transmission power level for a voice packet before its deadline. In [92, 93], H. Wang and N. Mandayam consider an opportunistic file transfer over a Rayleigh fading channel. The resulting optimal binary power control scheme, i.e., either transmit with fixed power level or not transmit at all, takes care of both the energy constraint and the different delay constraints for a fixed size file transfer. In [97, 98], D. Zhang and K. Wasserman study the energy efficient power control problem for an always backlogged user over a time-varying channel, in which the channel conditions are assumed only partially observable. They prove that under a mild assumption, the resulting optimal policy for such a partially observable MDP problem has a certain structural property. In [63, 64], D. Rajan et al. explore transmission schemes for bursty sources over Gaussian channels. In their work, a packet is considered lost when the buffer overflows, when it is dropped or when it is received in error. They derive optimal transmission schemes to minimize packet loss with constraints on both the average delay and transmit power. In [6], R. Berry and R. Gallager consider the tradeoff between power consumption and packet delay for one way communication (where erroneous packets are lost and not retransmitted) over a fading channel. They show that the optimal power and delay curve is convex and quantify the behavior of the power delay tradeoff in the regime of asymptotically large delay. Finally, in [32], M. Goyal et al. extend the work in [6] to provide upper and lower bounds for a simplified rate allocation policy.

1.3.2 Fair Resource Allocation

In this thesis, we also present an integrated resource allocation policy covering the three management modules from a system operator's point of view. When facing multiple users, another important resource allocation criterion prevails, i.e., *fairness* among the users.

Fairness has always been an important issue in communications, especially in computer networks. In wired networks, packet scheduling, i.e., which packet should be sent next, takes care of the fairness issue. The most often used fairness criterion is *max-min fairness* and the Generalized Processor Sharing (GPS) model [60] is used as the ideal reference model by most known algorithms, e.g., Weighted Fair Queueing (WFQ) [10] and Worst-case Fair weighted Fair Queueing(WF^{2}Q) [5]. Recently, some wireless fair scheduling schemes have been proposed such as Channel-condition Independent packet Fair Queueing (CIF-Q) [56] and Idealized Wireless Fair-Queueing (IWFQ) [49], in which the GPS model has also been used as the fairness reference. Compared to these previous works, we propose a new fairness model that may be more appropriate to wireless communications, especially to soft capacity limited CDMA networks. The proposed fairness model is just a slight modification of the GPS model and incorporates the time varying channel conditions as a factor impacting on the quota of resources allocated to a user.

Based on our proposed fairness model, we then present a detailed packet level resource allocation policy that consists of a series of actions: transmission scheduling, power allocation, and rate allocation in each frame. Recently, many compound resource allocation schemes have been proposed but each with a particular objective and focus, e.g., [2, 4, 34, 35, 57, 58, 59, 67]. M. Arad et al. [2, 4] and Ö. Gürbüz et al. [34, 35] propose detailed packet level resource allocation policies including transmission scheduling and power allocation for multi-service CDMA networks. In their works, data users are allocated the same instantaneous data rate and the simple first-in-first-out (FIFO) transmission scheduling is used. In [57, 58, 59], S. Oh and K. Wasserman propose several resource allocation schemes all based on the maximization of the system total throughput, i.e., the total instantaneous data rate over all data users, in a multi-cell CDMA system. The total throughput is maximized when the allocated instantaneous data rate is inversely proportional to a user's path gain. However, their proposed scheme does not consider fairness among the data users, and hence a backlogged flow with a low path gain may be starved for a long time. In [67], O. Sallent et al. propose a detailed packet level resource allocation scheme for data users. In this work, different instantaneous data rates for data users are allowed but no explicit fairness guarantee is provided. Compared to these works, we allow users to be allocated different instantaneous data rates and provide explicit fairness guarantees among the users. However, these are based on our proposed fairness model.

1.4 Contributions of This Thesis

We apply MDP theory to solve the optimal policy design problems from a single user's point of view for the three resource management modules, viz., power control, transmission scheduling and rate allocation. Though the problems share a common mathematic structure, their contexts are different. We also present a detailed packet level resource allocation policy from a system operator's point of view based on a proposed new fairness model. This section reviews the main work in this thesis. Our contributions are also briefly outlined and compared to the related works.

1.4.1 Optimal Power Allocation Policies

Intuitively, only transmitting in the best channel state and using the least transmission power lead to the most energy efficient transmissions. However, the resulting cost is increased delay. We consider an energy efficient file transfer problem, in which a user needs to decide when to transmit and how much transmission power should be used in each transmission in order to consume the least power while meeting the delay constraints for finishing the file transfer. We model such a file transfer problem as a constrained stochastic optimization problem. We note that our problem can be considered as a dual problem of the one investigated by H. Wang and N. Mandayam [92, 93], which studies how to maximize the probability of a successful file transfer over a Rayleigh fading channel via a binary power control scheme under total energy and transfer delay constraints. Similar to [92, 93], we consider two delay constraints: the average delay constraint and the strict delay constraint. However, we also consider multiple transmission power levels. Furthermore, our objective is to achieve energy efficient file transfer assuming an infinite power budget. We first show how to convert the average delay constrained stochastic optimization problem to a standard Markov decision problem via the Lagrange approach. The resulting optimal policy under the average delay constraint is a stationary one while the resulting optimal policy under the strict delay constraint is time dependent. We present numerical examples to show the resulting optimal policies and to compare the performance of the optimal policies to that of a fixed power persistent transmission policy. The simulation results indicate that the transmission power can be substantially reduced while the delay constraint is still satisfied with the computed optimal policies which exploit the channel variations.

This work is also summarized in our paper [87].

1.4.2 Optimal Transmission Control Policies

We consider a simple transmission control problem, in which the arrival process is included but the action is simplified as either to transmit or not to transmit. The objective is to find the policy that optimally balances different costs such as the delay and transmission power. We prove the existence of stationary average optimal policies for such a Markov decision problem and explore the properties of the optimal policies. In [97, 98], Zhang and Wasserman have explored the structure of the optimal policies for an always backlogged user, i.e., when the channel estimation is in some bad states, the sender suspends transmission and waits for the channel to transit to some good states. Compared to their work, we show that with the arrival dynamics included, the sender has to transmit in some bad channel states when the buffer exceeds some thresholds to avoid increasing the delay cost. Furthermore, we propose an improved policy iteration algorithm to efficiently compute optimal policies, which is based on the property of the optimal policies. We present numerical examples to illustrate how the different cost functions affect the resulting optimal policy and its performance. We compare the performance of the optimal policy with that of a persistent transmission policy. We also provide extensive simulation results that investigate the effect of channel memory on the performance of the optimal policies. These results indicate that increasing the channel memory increases the value of the optimal policy but decreases the system throughput.

This work is also summarized in our papers [91, 90].

1.4.3 Optimal Rate Allocation Policies

Besides choosing the transmission times and adapting transmission powers, a data connection may also adapt its transmission rate during its holding time to achieve cost efficiency while meeting QoS requirements. We investigate the rate allocation problem, in which the arrival process is included but the channel is simplified as time invariant. Some recent works have analyzed the problem of designing a power efficient transmission scheme over a fading channel [6, 32, 63, 64]. Compared to these works, our work simplifies the channel to be time invariant but we consider retransmissions. We show that the optimal policy is monotone under a mild assumption, i.e., a larger transmission rate should be chosen when the buffer occupancy increases. We analyze two extreme policies which provide the upper and lower delay bounds based on the stochastic process comparison technique. A case study with numerical examples is also presented. We propose a class of one-threshold based simple policies and provide a tight upper delay bound for such simple policies. We also propose and apply a modelling technique in the case when a single user has to consider its self-optimization in the presence of other users (interference). The characteristic and the property of the optimal policies for the extended problem are also presented.

This work is also summarized in our papers [89, 85].

1.4.4 Fair-effort Based Resource Allocation

We study the three resource management modules, i.e., transmission scheduling, power control and rate allocation, from the viewpoint of an operator who allocates the system resources among multiple users. We focus on two policy design objectives: fairness among users and system utilization efficiency. Unlike the GPS fairness model, we propose a new fairness model. The nominal weight of a flow is considered time-dependent in our fairness model while it is fixed in the GPS model. By such a simple modification, we can incorporate the (possible) interaction between users and the resource allocation process. We then present a simple credit based algorithm to approximate the proposed fairness model. Based on our fairness model, we present a detailed packet level resource allocation scheme for a CDMA-based wireless network. The scheme consists of resource shares assignment, transmission scheduling, rate and power allocation. We evaluate our proposal via simulations. The simulation results show the advantages of using our fairness model in terms of the system utilization efficiency.

This work is also summarized in our papers [84, 88, 86].

1.5 Thesis Organization

In this chapter, we have presented a brief introduction to cellular mobile communications, some challenges and some resource management modules for the radio resource management problem in next generation mobile systems. The research topics of interest are identified and some related works have been reviewed.

The rest of this thesis is organized as follows. Chapter 2 summarizes the common features of the system models and some Markov decision theory. Chapter 3 studies the optimal power allocation policies for an energy efficient file transfer problem. Chapter 4 considers the transmission scheduling problem in which the arrival process is included but the action is simplified. Chapter 5 investigates the rate allocation problem in which the arrival process is included but the channel is simplified as time invariant. A case study and extensions are also presented in Chapter 5. Chapter 6 deals with the resource allocation problem from a system operator's point of view. Finally, concluding remarks and some future research work are given in Chapter 7.

Chapter 2

System Models and Some Markov Decision Theory

In this chapter, we first describe the common features of the models used in this thesis and then summarize some Markov decision theory used as the theoretical framework for our Markov decision problems.

2.1 Basic System Models

The simplified system architecture of cellular mobile communications illustrated in Fig. 1.1 comprises several cells in the system. In this thesis, we focus on the resource allocation issues in a single cell only, in which multiple mobile stations communicate with the same base station located in the center of the cell. Mobile stations can communicate with the base station simultaneously via the use of CDMA or exclusively via the use of TDMA. Both are considered in this thesis, however, only a particular frequency band is considered and hence FDMA is assumed throughout this thesis. Though we study resource allocation problems each with a different objective from the decision theoretic points of view, the problems share some common features and we summarize them as follows.

2.1.1 Discrete System

In this thesis, we consider a discrete system in which transmissions, transmission decisions, transmission powers and transmission rates are all assumed to be discrete.

We consider a discrete time system in which the time axis is divided into contiguous *frames*[1] of equal duration. Transmissions are synchronized at the frame level, i.e., a transmission, if scheduled, should start at the beginning of a frame. Though a transmission needs not span a whole frame, we sometimes assume so. Transmission decisions, e.g., whether to transmit or not in a frame, are also made at the beginning of a frame and just before the start of the transmission.

When we need to allocate different transmission powers or transmission rates, we assume that the available powers and rates are discrete and finite. This assumption simplifies the problem formulation which will be clear in the next section. However, our problems can be extended to the continuous domain without too many modifications.

2.1.2 Transmission Model

We consider a transmission system that allows the use of different levels of transmission power and transmission rate in different frames as well as retransmissions. The use of different levels of transmission power helps to combat the harsh transmission conditions of the radio channel and interference on the one hand, and allows energy efficiency on the other hand. The ability to transmit with different rates is an important characteristic of next generation mobile systems. For example, *radio bearers* with different transmission rates can be easily set up via the configuration and (re)allocation procedure specified in UMTS [19], while different transmission rates can be achieved by using variable spreading factor and/or multi-code operations. Retransmissions are often used in real systems to improve the transmission efficiency, e.g., the RLC retransmission mode in UMTS [18], especially for data services with some delay tolerance.

In our transmission model, we assume that all errors in a frame can be detected and if an erroneous frame cannot be corrected, the data packet(s) in that frame should

[1]The term frame used in this thesis needs not be the physical radio transmission frame but just a notational classification.

be retransmitted. We then assume that each frame should be either positively or negatively acknowledged, i.e., ACK/NACK should be sent by the receiver via some feedback channels. In UMTS, either a dedicated or common control channel can be used to send the acknowledgements, e.g., the dedicated physical control channel (DPCCH) and the primary/secondary common control channel (CCPCH) defined in [15, 16]. For simplicity, instantaneous and perfect reception of the acknowledgements is assumed and a simple *stop-and-wait* retransmission scheme is employed in our transmission model. Finally, we assume that the receiver has the ability to measure the transmission channels and send perfect channel state reports (CSR) to the sender, although some delay in sending CSR is allowed. The measurement of channel conditions can be achieved using some pilot/training bits in each frame, e.g., the training bits in a GSM frame [65]. In UMTS, a more comprehensive and complicated procedure for physical layer measurements has been defined in [17, 23].

Transmissions over wireless channels are not reliable and hence a frame will be successfully received only with some probability. We let f_s , $0 \leq f_s \leq 1$, denote the average frame success probability (FSP) in this thesis. Note that f_s can be either a function or as simple as a scalar based on the context. The detailed form of f_s depends on the choice of the modulation and channel coding schemes, the interleaving depth, and some other system parameters. The value of f_s can also be obtained via Monte Carlo simulations.

We use the following figures to illustrate our transmission model as an example.

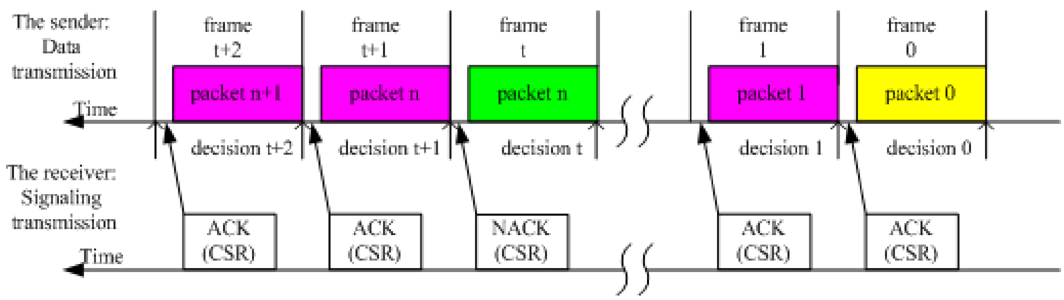

Figure 2.1: Transmission model example 1 – A single user transmits with different transmission powers, represented by different colors in frames.

Fig. 2.1 provides an example of a single user transmitting with different transmission

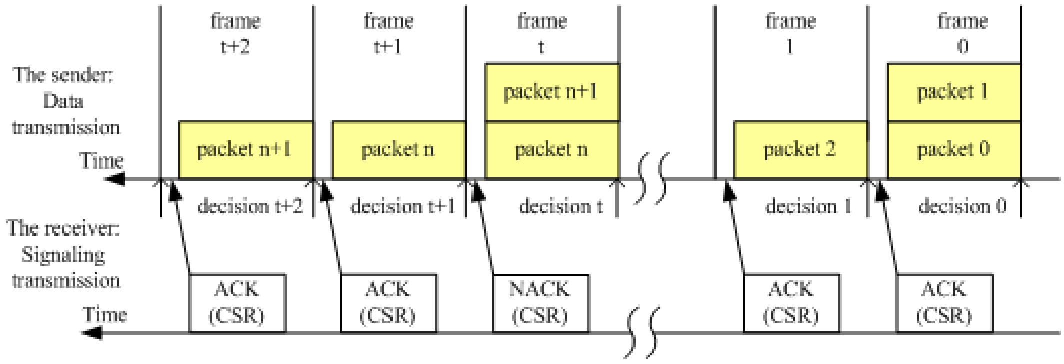

Figure 2.2: Transmission model example 2 – A single user transmits with different transmission rates, represented by different number of packets in frames

powers. At the beginning of a frame, the sender may decide whether or not to transmit in a frame, and if it decides to transmit, which level of transmission power should it use. Note that the transmission of a data packet needs not span the whole frame and so instantaneous acknowledgements can be obtained before the next frame. Fig. 2.2 presents an example that a single user transmits with different transmission rates. In this thesis, we assume that if a frame is negatively acknowledged, then all the data packets in that frame need to be retransmitted. Although we use dedicated control channels to transmit control information in Fig. 2.1 and Fig. 2.2, we note that other methods such as piggybacking are also allowable. Finally, we note that the sender needs to make decisions at the beginning of each frame. In this thesis, we will consider two kinds of decision and optimization problems. One is based on the Markov decision theory focusing on a single user optimization problem. The other is to allocate resources across users while meeting some optimization constraints. We introduce a more general Markov decision model and related theory in the next section and defer the introduction of the second optimization problem to Chapter 6.

2.2 Some Markov Decision Theory

In this thesis, we solve some of the optimal policy design problems based on decision theory. Thus in this section, we provide a brief introduction to Markov decision processes and define the notations that will be used throughout this thesis.

2.2.1 Markov Decision Processes

A Markov decision process (MDP) provides the theoretic foundation and framework for modelling sequential decision making under uncertainty [7, 62]. MDP has been widely adopted as a powerful tool in many fields such as applied mathematics, operations research, economics, management science, stochastic control, and communications engineering. In queueing systems and communication networks, MDP has been applied for the analysis of traffic admission control, flow and congestion control, service rate control and routing (see [1, 76, 77] for comprehensive surveys and references therein).

An MDP model consists of five elements: *decision epochs, states, actions, transition probabilities* and *costs* (or *rewards*). In an MDP, a decision maker needs to take an action at each decision epoch based on the observation of the current state (or the history) of the system. The action chosen in the current decision epoch causes an immediate one-stage cost (or generates a reward) and determines the state at the next decision epoch through a transition probability function. At different decision epochs, the available actions may be different since the system may be in different states. When choosing an action at a decision epoch, the decision maker needs to take into account not only the outcome of the current action but also future decision making opportunities. An MDP is thus a stochastic model for a controlled stochastic process and is often referred to as *stochastic dynamic programming*. If decision epochs are finite (infinite), an MDP is said to be a *finite (infinite) horizon* process. The set of decision epochs can be either a discrete or continuous set, and in the latter case an MDP is termed a *semi-Markov decision process* (SMDP) or *continuous-time Markov decision process* (CTMDP). For analysis, a CTMDP can be converted to an SMDP or discrete time MDP through a standard *uniformization* technique. We will focus on infinite horizon discrete time Markov decision problems in this thesis.

An MDP together with an *optimality criterion* define a Markov decision problem. We introduce several optimality criteria in the next section. A *policy* which consists of a sequence of *decision rules* provides a solution to such a Markov decision problem. A decision rule prescribes a procedure for action selection at a specified decision epoch and hence it is a mapping from the state space to the action space. A decision rule can be *deterministic* or *random* according to how it chooses an action based on certainty or a probability distribution. It can also be *Markovian* or *history dependent* based on whether the action is chosen based on only the current state or the history of the system. A policy is called *stationary* if the decision rules are the same for all decision epochs. In this thesis, we mainly focus on Markovian deterministic stationary policies, which are easy to compute and implement from the engineering points of view. Before going into the next section, we summarize some notations that are used throughout this thesis.

We use $\mathbb{R}$ and $\mathbb{R}^+$ to denote the set of real numbers and the set of non-negative real numbers, respectively. We use $\mathbb{N}$ and $\mathbb{N}^+$ to denote the set of integers and the set of non-negative integers, respectively. As introduced in the previous section, we consider discrete time systems and hence we only consider discrete time MDP models. The decision epochs correspond to the beginning of each frame. The set of decision epochs is denoted as $\mathcal{T}$. We use t to denote a frame and use subscript to denote a decision epoch t, $t = 0, 1, \cdots, T-1$ and $T \leq \infty$. The system state is denoted as $\mathcal{S}$ and an individual state as s or $\mathbf{s}$, where $\mathbf{s}$ is used when a system state consists of more than one component. The state of the system at decision epoch t is then denoted as s_t. We use $\mathcal{A}_s$ to denote the set of available actions in state s and $\mathcal{A}$, $\mathcal{A} = \bigcup_{s\in\mathcal{S}} \mathcal{A}_s$, to denote the system action space. An action in the action space is denoted as a. We use $\mathrm{Tr}(s'|s,a)$ to denote the transition probability that the system occupies state s' at the next decision epoch if the current state is s and action a is chosen. We use $C(s,a)$ to denote the immediate one-stage cost when the system state is s and action a is selected, which is a mapping from the product of state space and action space to real values, i.e., $C(s,a) : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$. A decision rule at decision epoch t is denoted as $d_t(\cdot) : \mathcal{S} \mapsto \mathcal{A}$. A policy is denoted as $\pi = (d_0, d_1, \cdots, d_{T-1})$, $T \leq \infty$ and we use Π to denote the set of all (allowable) policies. For simplicity, we use μ to denote stationary policies.

2.2.2 Optimality Criteria

An optimality criterion determines how to compute the value of a policy and how to determine an optimal policy. Commonly used optimality criteria include *the expected total cost optimal criterion*, *the expected total discounted cost optimal criterion* and *the expected average cost optimal criterion*. Using the expected total discount cost optimal criterion is analytically easier than using the other two criteria. However, our policy design problems are better understood and explained under the expected total cost and the average cost optimal criteria. Since the three criteria are closely related under some conditions, we briefly present their definitions here.

The value of a policy π with the starting state s, $s \in \mathcal{S}$, under the expected total cost optimal criterion is defined as

$$V^\pi(s) = \lim_{T\to\infty} \mathbb{E}_s^\pi \left\{ \sum_{t=0}^{T-1} C(s_t, a_t) \right\} \tag{2.1}$$

$\mathbb{E}_s^\pi$ is the expectation of policy π conditioning on the starting state s. If the limit exists and the interchanging of the limit and expectation is valid, the value of the policy π can be written as

$$V^\pi(s) = \mathbb{E}_s^\pi \left\{ \sum_{t=0}^{\infty} C(s_t, a_t) \right\} \tag{2.2}$$

If $\mathbb{E}_s^\pi \left\{ \sum_{t=0}^{\infty} |C(s_t, a_t)| \right\} < \infty$, it means that reaching some cost-free absorbing state(s) is inevitable (with probability 1) and in this case, the Markov decision problem with the expected total cost optimal criterion is also a stochastic shortest path problem [7].

The value of a policy π with the starting state s under the expected total discount optimal criterion (*discount optimal* for short) is defined as

$$V_\rho^\pi(s) = \lim_{T\to\infty} \mathbb{E}_s^\pi \left\{ \sum_{t=0}^{T-1} \rho^t C(s_t, a_t) \right\} \tag{2.3}$$

ρ is the discount factor and $0 \le \rho < 1$. The limit in (2.3) exists when

$$\sup_{s\in\mathcal{S}} \sup_{a\in\mathcal{A}_s} |C(s,a)| = \text{CONST} < \infty \tag{2.4}$$

When the state space and the action space are finite, we call a cost structure (uniformly) *bounded* if the cost structure satisfies (2.4). When the limit exists and interchanging the limit and expectation is valid, for example when (2.4) holds, we write

$$V_\rho^\pi(s) = \mathbb{E}_s^\pi\left\{\sum_{t=0}^{\infty}\rho^t C(s_t, a_t)\right\} \tag{2.5}$$

Note that $V^\pi(s) = \lim_{\rho\uparrow 1} V_\rho^\pi(s)$ whenever (2.2) holds.

The value of a policy π with the starting state s under the expected average cost optimal criterion (*average optimal* for short) is defined as

$$\overline{V}^\pi(s) = \limsup_{T\to\infty}\frac{1}{T}\mathbb{E}_s^\pi\left\{\sum_{t=0}^{T-1} C(s_t, a_t)\right\} \tag{2.6}$$

The limit supremum is used in (2.6) since the limit sometimes fails to exist. Note that the limit infimum can also be used in (2.6), however, the use of the limit supremum may represent the worst case analysis.

Given an optimality criterion and an initial state, a policy is said to be *optimal* if it has the smallest policy value among all (allowable) policies. We use $\pi^* = (d_0^*, d_1^*, \cdots)$ to denote an optimal policy, where d^* are optimal decision rules. For the total cost optimal criterion, it is defined as

$$V^{\pi^*}(s) = V^*(s) \equiv \inf_\pi V^\pi(s) \tag{2.7}$$

For the discount optimal criterion, it is define as

$$V_\rho^{\pi^*}(s) = V_\rho^*(s) \equiv \inf_\pi V_\rho^\pi(s) \tag{2.8}$$

For the average optimal criterion, it is define as

$$\overline{V}^{\pi^*}(s) = \overline{V}^*(s) \equiv \inf_\pi \overline{V}^\pi(s) \tag{2.9}$$

Note that we may have more than one optimal policy.

Among optimal policies, we are interested in stationary optimal policies (if they exist) which are easy to implement. The next section introduces the conditions for the existence of a stationary optimal policy.

2.2.3 Stationary Optimal Polices

In this thesis, we focus on decision problems in which the state space and the action space are finite, although some analyses later assume an infinite state space for theoretical completeness. Furthermore, the cost structures of our decision problems are bounded, provided that states and actions are finite, and hence (2.4) holds. Therefore, it is easy to verify that stationary discount optimal policies exist for our Markov decision problems. But we need more conditions for the stochastic shortest path and the average optimal problems. We summarize some related MDP theories as follows, and we refer the reader to [7, 62] for more complete discussions.

In a stochastic shortest path problem, there exists (at least) a terminal state that is cost free and absorbing, i.e., whenever the system enters the terminal state, it will stay there forever. We will use the following definition and theorem for such decision problems.

Definition 2.1 (Proper policy, Definition 1.1 in [7]) *A stationary policy μ is said to be* ***proper*** *if, when using this policy, the terminal state will be reached with probability one, regardless of the starting state. A stationary policy that is not proper is said to be* ***improper****.*

If a stationary policy is proper, its policy value is finite provided that the state space and action space are finite and a bounded cost structure is used. Otherwise, there exists at least one state such that the value of any improper policy is infinite starting from such a state.

Theorem 2.1 (Proposition 1.2 in [7]) *If there exists at least one proper policy with finite policy value, then there exists a stationary optimal policy. Furthermore, there exists a bounded function u on $\mathcal{S}$ satisfying*

$$u(s) = \min_{a \in \mathcal{A}_s} \left\{ C(s,a) + \sum_{s \in \mathcal{S}} \mathrm{Tr}(s'|s,a)u(s') \right\} \tag{2.10}$$

A stationary policy realizing the minimum part in (2.10) for all $s \in \mathcal{S}$ is an optimal policy.

The function u is called the *value function* and (2.10) is also known as *Bellman's equation*. In the stochastic shortest path decision problem, $u(s)$ is the optimal policy value, i.e., the minimum expected total cost, if the system follows the optimal policy starting from the state s .

In a discount optimal decision problem, the conditions for the existence of stationary optimal policies are very mild. For the interests of our decision problems, we summarize the related MDP theory[2] in the following theorem.

Theorem 2.2 *Assume that $\mathcal{S}$ is discrete and $\mathcal{A}_s$ is finite for all $s \in \mathcal{S}$, then there exists a stationary discount optimal policy for all $0 \le \rho < 1$. Furthermore, there exists a bounded function u_ρ on $\mathcal{S}$ satisfying*

$$u_\rho(s) = \min_{a \in \mathcal{A}_s} \left\{ C(s,a) + \rho \sum_{s' \in \mathcal{S}} \mathrm{Tr}(s'|s,a) u_\rho(s') \right\} \tag{2.11}$$

A stationary policy realizing the minimum part in (2.11) for all $s \in \mathcal{S}$ is a discount optimal policy.

The function u_ρ is also called the *value function*. In the discount optimal decision problem, $u_\rho(s)$ is the optimal policy value, i.e., the minimum expected total discount cost, if the system follows the optimal policy starting from the state s.

In an average optimal decision problem, however, more conditions are needed for the existence of a stationary optimal policy. Note that in a Markov decision problem, any stationary policy induces a Markov chain on the state space. We will use the following definition.

Definition 2.2 *A Markov chain is called* ***unichain*** *if the Markov chain consists of a single positive recurrent class and a (possibly empty) set of transient states. A Markov decision process is called* ***unichain*** *if the Markov chain induced by every (allowable) stationary policy is unichain.*

The assumption that a Markov decision process is unichain is important to an average optimal decision problem, although the verification may not be as straightforward. If a Markov decision process is unichain, the average cost of a stationary policy can be achieved by the limit in (2.6), and $\overline{V}^*(s)$ is a constant independent of the starting state. The following theorem summarizes the related MDP theory.

[2] See [62] and [7] for complete discussions, for example, see Theorem 6.2.10 in [62] and Proposition 2.2 of Section 1.2 in [7] for reference.

Theorem 2.3 (Theorem 8.4.3 in [62]) *Assume that $\mathcal{S}$ and $\mathcal{A}$ are finite and $C(s,a)$ is uniformly bounded for all s and a. Furthermore, assume that the Markov decision process is unichain. Then there exists a stationary average optimal policy. Moreover, there exists a finite constant J and a bounded function $\overline{u}$ on $\mathcal{S}$ satisfying*

$$J + \overline{u}(s) = \min_{a \in \mathcal{A}_s} \left\{ C(s,a) + \sum_{s' \in \mathcal{S}} \mathrm{Tr}(s'|s,a)\overline{u}(s') \right\} \tag{2.12}$$

A stationary policy realizing the minimum part in (2.12) for all $s \in \mathcal{S}$ is an average optimal policy.

Equation (2.12) is also called the *average cost optimality equation* (ACOE). The constant J is the optimal policy value, i.e., the minimum expected average cost, if the system follows the optimal policy from any starting state. The function $\overline{u}$ is called the *relative value function*. For simplicity, we also call it the value function in this thesis. In an average optimal decision problem, $\overline{u}(s)$ is interpreted as the minimum of the difference between the total expected cost to reach a distinct recurrent state from the state s for the first time and the cost that would be incurred if the cost per frame were the optimal average cost J, when the system follows the optimal policy.

The optimal equations (2.10), (2.11) and (2.12) provide the method to compute optimal policies. When deciding optimal policies, we always break ties by choosing the smallest action for our Markov decision problems. A general value iteration algorithm for computing optimal policies is described in the next section.

2.2.4 Computation of Optimal Policies

Some fairly standard techniques can be used to compute a stationary optimal policy, e.g., *value iteration*, *policy iteration* and *linear programming* (see [62] for details). We briefly introduce the value iteration algorithm as it also provides the method to investigate the characteristic of value functions. The value iteration algorithm of the stochastic shortest path decision problem is structurally similar to that of the discount optimal decision problem and hence we only introduce the latter. The value iteration algorithm is as follows.

Value Iteration Algorithm

1. Set $k = 0$; for all $s \in \mathcal{S}$, set
$$u_\rho^0(s) = 0 \tag{2.13}$$

2. Set $k = k + 1$ and for all $s \in \mathcal{S}$, compute
$$u_\rho^k(s) = \min_{a \in \mathcal{A}_s} \left\{ C(s, a) + \rho \sum_{s' \in \mathcal{S}} \mathrm{Tr}(s'|s, a) u_\rho^{k-1}(s') \right\} \tag{2.14}$$

3. For all $s \in \mathcal{S}$, if
$$\|u_\rho^k(s) - u_\rho^{k-1}(s)\| < \epsilon \tag{2.15}$$
goto (4). Otherwise, goto (2)

4. For each $s \in \mathcal{S}$, choose
$$\mu(s) = \arg\min_{a \in \mathcal{A}_s} \left\{ C(s, a) + \rho \sum_{s' \in \mathcal{S}} \mathrm{Tr}(s'|s, a) u_\rho^k(s') \right\} \tag{2.16}$$
and stop.

In step (3), the function $\|\cdot\|$ is a norm function and $\epsilon > 0$ is a constant specified beforehand. Hence the computed optimal policy is also called ϵ-optimal policy. According to MDP theory [7, 62], it can be shown that the function u_ρ^k converges to the value function u_ρ as $k \to \infty$ for all s, i.e.,

$$\lim_{k \to \infty} u_\rho^k(s) = u_\rho(s), \quad \text{for all} \quad s \in \mathcal{S} \tag{2.17}$$

To compute an average optimal policy, the relative value iteration algorithm can be used to avoid divergence of the value iteration algorithm that may occur in the computation. It is a simple modification of the above algorithm. To apply the relative value iteration algorithm, select a distinct recurrent state $\tilde{s} \in \mathcal{S}$. Note that under the unichain assumption, all recurrent states communicate with the state $\tilde{s}$. As in the value iteration algorithm above, we set $\overline{u}^0(s) = 0$ for all $s \in \mathcal{S}$. Step (2) of the value iteration algorithm is modified to compute

$$w^{k-1}(\tilde{s}) = \min_{a \in \mathcal{A}_{\tilde{s}}} \left\{ C(\tilde{s}, a) + \sum_{s' \in \mathcal{S}} \mathrm{Tr}(s'|\tilde{s}, a) \overline{u}^{k-1}(s') \right\} \tag{2.18}$$

and

$$\overline{u}^k(s) = -w^{k-1}(\tilde{s}) + \min_{a \in \mathcal{A}_s} \left\{ C(s, a) + \sum_{s' \in \mathcal{S}} \mathrm{Tr}(s'|s, a) \overline{u}^{k-1}(s') \right\} \tag{2.19}$$

instead. According to MDP theory [7, 62], it can be shown that

$$\lim_{k\to\infty} w^{k-1}(\tilde{s}) = J \tag{2.20}$$

and for all recurrent states $s \in \mathcal{S}$,

$$\lim_{k\to\infty} \overline{u}^k(s) = \overline{u}(s) \tag{2.21}$$

Finally, we note that if all states communicate, i.e., the Markov chain is ergodic, then the discount optimal problem is related to the average optimal problem by

$$J = \lim_{\rho\to 1}(1-\rho)u_\rho(s) \text{ for all } s \in \mathcal{S} \tag{2.22}$$

Thus we can first investigate the property of a discount optimal decision problem. Then all the results will also apply to the average cost optimality criterion once we identify the unchain property for the average optimal decision problem.

2.3 Summary

In this chapter, we have introduced the basic system model and some Markov decision theory. Some notations are also summarized in this chapter. We will consider a Markov decision modelled file transfer problem in the next chapter.

Chapter 3

Optimal Power Allocation Policies

In this chapter, we consider the power allocation problem in the transfer of a file by a single sender in the presence of a Rayleigh fading channel. The fading channel is modelled by a finite state Markov process. The sender can choose to use different power levels in different channel states. As will be seen, we consider two kinds of delay constraints.

3.1 Channel Model

Many researchers have proposed to use a finite-state Hidden Markov Model (HMM) to model a wireless channel [94, 101]. Through the construction of a finite-state Markov process, the variations of a time-varying channel can be represented via the stationary transition probabilities. In some wireless communication situations, changes of the received signal-to-noise ratio (SNR), i.e., the path gain, occur on a very slow time scale (slow fading) compared with the transmission rate. Thus, it is reasonable to assume that the transmitted symbols in one frame experience the same channel fading. In this chapter, we consider a slow Rayleigh fading channel and use the method introduced in [94] to construct a finite-state Markov channel.

Consider a slow fading channel. We assume that the received SNR remains at a constant level during a frame. The channel can be modelled as an additive white

Gaussian noise (AWGN) channel as follows:

$$y = \sqrt{h}x + n \tag{3.1}$$

where x and y are input and output signals, respectively, n is the AWGN noise and h is the channel fading. For a Rayleigh fading channel, h is exponentially distributed with the probability density function

$$f(h) = \frac{1}{\bar{h}} \exp(-\frac{h}{\bar{h}}), \quad h \geq 0 \tag{3.2}$$

where $\bar{h}$ is the average SNR of the channel. When the background noise is normalized to 1, h may characterize the received SNR. A finite-state Markov channel model for such a Rayleigh fading channel can be constructed as follows. Select a sequence of received SNR thresholds: $\gamma_0 < \gamma_1 < \cdots < \gamma_M$ by which the range of the received SNR is partitioned into a finite number of SNR intervals (M intervals for example). Let $\mathcal{H} = \{h_1, h_2, \cdots, h_M\}$ denote such SNR intervals, in which h_i represents the channel state during a frame, and $0 < h_1 < h_2 < \cdots h_M$ where the greater the index, the better the channel quality. Then the channel is said to be in state h_m if $h \in [\gamma_{m-1}, \gamma_m), m = 1, 2, \cdots, M$. Let $\hat{H}_m$ denote the steady state probability that the channel stays in the state m, i.e.,

$$\hat{H}_m = \int_{\gamma_{m-1}}^{\gamma_m} f(h)dh, \quad m = 1, 2, \cdots, M \tag{3.3}$$

and $\sum_{m=1}^{M} \hat{H}_m = 1$. We assume that the channel state H_t during frame t remains unchanged and that state transitions occur at the boundary of a frame. The channel state transition probabilities are denoted as $h_{ij} \equiv \Pr(H_{t+1} = h_j | H_t = h_i)$. According to the model proposed in [94], a channel state transits only to its neighboring states or stay in the same state. Furthermore, according to [94], the transition probabilities are approximated by the ratio of the expected number of level crossings of the state SNR boundary to the average number of symbols per second in that state, and they are given by

$$h_{i,i+1} = \frac{\Psi(\gamma_{i+1})}{\hat{H}_m R_s}, \quad i = 1, \cdots, M-1 \tag{3.4}$$

$$h_{i,i-1} = \frac{\Psi(\gamma_i)}{\hat{H}_m R_s}, \quad i = 2, \cdots, M, \tag{3.5}$$

where R_s is the symbol rate and Ψ is the expected number of level crossings given by

$$\Psi(z) = \sqrt{\frac{2\pi z}{\bar{h}}} f_d \exp(-\frac{z}{\bar{h}}) \tag{3.6}$$

and f_d is the maximum Doppler shift. This fading channel model has been verified to be precise when the fading process is slow enough (see [94] for more discussions). We let $\mathbf{H} = [h_{ij}]_{M\times M}$ denote the channel state transition matrix and $\mathbf{H}^t = \prod_{i=1}^{t} \mathbf{H}$. For simplicity, the channel states are classified with equiprobability, i.e., all channel states have the same steady state probability, though other classifications are also allowable (see [99] for example).

We assume that the channel coding and the modulation schemes are fixed in all frames. However, the transmission power can be changed in different frames to combat the harsh channel conditions while allowing for energy savings. Hence the probability of successful reception of a frame is dependent on both the transmission power and the channel state, and is denoted as $f_s(a, h_i)$ where a is the transmission power used in a frame and h_i is the current (or estimated) channel state in the same frame. As we consider a discrete system (cf. Section 2.1.1), the transmission powers are chosen from a finite set $\mathcal{A} = \{a_1, a_2, \cdots, a_N\}$, with a higher index denoting a higher power level. In general, we have the following relationship between frame success probabilities.

$$0 \le f_s(a_i, h_1) \le f_s(a_i, h_2) \le \cdots \le f_s(a_i, h_M) \le 1, \quad \text{for all} \ \ i \tag{3.7}$$

and

$$0 \le f_s(a_1, h_i) \le f_s(a_2, h_i) \le \cdots \le f_s(a_N, h_i) \le 1, \quad \text{for all} \ \ i \tag{3.8}$$

(3.7) implies that the probability of a successful transmission is nondecreasing in the channel states given a fixed transmission power; (3.8) implies that the probability of a successful transmission is also nondecreasing in the transmission power given a fixed channel state. However, it is not so easy to determine a detailed form of the function f_s. It is also not easy to compare f_s given any two arbitrary (a, h) pairs. Thus we often use (3.7) and (3.8) for some qualitative analysis and resort to Monte Carlo simulations to find the different f_s values.

3.2 Problem Formulation

We first provide the system model and then formulate the problem of an energy efficient file transfer over a fading channel as a constrained optimization problem in which two delay constraints are considered.

3.2.1 System Model

The system model is shown in Fig. 3.1. We consider a discrete time system (an example

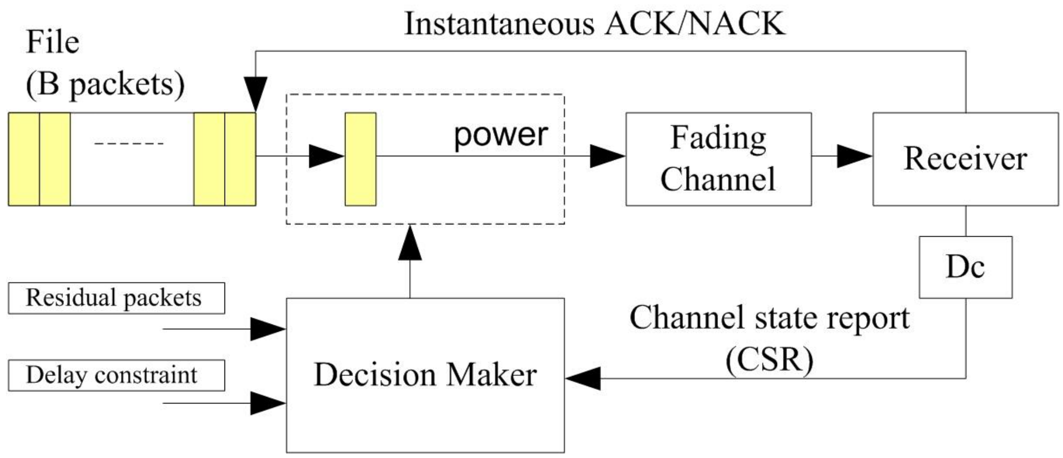

Figure 3.1: System model

of a transmission model has been shown in Fig. 2.1). Note that each frame can transmit exact one (coded) file packet. At the beginning of a frame, the sender first decides to transmit or not based on the observation of the number of residual file packets in the buffer and also the delay constraint. Furthermore, if a transmission is determined then the sender decides the transmission power level needed for transmitting the packet.

The receiver measures the channel conditions via some pilot/training bits in each frame and sends a channel state report (CSR) to the sender over the feedback channel. Perfect reception of a CSR is assumed. However, there may be some delay in sending a CSR. Let D_c denote the delay of the channel state report. $D_c = 0$ denotes that the sender knows the channel state of the current frame before transmission and $D_c = t$ implies that the sender knows the channel state from t frames before and only has an estimate of the current channel state based on the channel transition matrix. We focus on a finite CSR delay.

3.2.2 Energy Efficient File Transfer with Delay Constraints

During the file transfer, the sender has the choice to use different transmission power levels in different frames. This opens up the chance for energy savings. However, using different power levels may also result in different frame success probabilities (even given the same channel state) and hence affects the (average) total transmission time for the file transfer. Thus the tradeoff between power consumption and transfer delay exists, and we explore such a tradeoff. We first model the problem in a direct way, i.e., as a constrained optimization problem, and provide solutions in later sections.

We summarize some notations as follows. The size of the file is B packets. We use the subscript t to denote a frame, and without loss of generality, we assume that the file transfer starts at time $t = 0$ ($t = 0, 1, \cdots$). Let $\mathcal{S}$ denote the system state space and $\mathbf{s}_t = (b_t, h_t)$, $\mathbf{s} \in \mathcal{S}$, denote the system state at the beginning of the frame t, where b_t is the residual file packets in the buffer and h_t is the channel state or the estimate of the channel state. As stated earlier, we consider discrete transmission power levels only. Let $\mathcal{A} = \{0, a_1, a_2, \cdots, a_N\}$ denote the action space where 0 stands for no transmission and a_i, $i = 1, \cdots, N$ represent N possible transmission power levels. Without loss of generality, we assume $0 < a_1 < a_2 < \cdots < a_N$. Let $\pi = (d_0, d_1, \cdots, d_t, \cdots)$ denote the transmission policy where d_t denotes the transmission scheme used in frame t. Note that $d_t(\mathbf{s}_t)$ is a function mapping from the state space to the action space, i.e., $d_t : \mathcal{S} \mapsto \mathcal{A}$. If d_t is the same for all frames, the corresponding policy is said to be *stationary*. For notational simplicity, let $d_t = 0$ denote no transmission is scheduled in a frame and $d_t > 0$ denote a transmission is scheduled with some positive transmission power level. Given an action a in a frame, the transition probabilities are given as

$$\mathrm{Tr}((b', h_j)|(b, h_i), a) = \begin{cases} h_{ij}, & a = 0, b' = b \\ \tilde{f}_s h_{ij}, & a > 0, b > 0, b' = b - 1 \\ (1 - \tilde{f}_s) h_{ij}, & a > 0, b > 0, b' = b \\ 0 & otherwise \end{cases} \tag{3.9}$$

where $\tilde{f}_s$ is the probability of a correct reception of a frame when the CSR is h_i. Since the CSR depends on D_c, $\tilde{f}_s$ is given as

$$\tilde{f}_s = \begin{cases} f_s(a, h_i), & D_c = 0 \\ \sum_{j=1}^{M} h_{ij}^t f_s(a, h_j), & D_c = t > 0 \end{cases} \tag{3.10}$$

where h_{ij}^t is the t-step channel state transition probability, i.e., h_{ij}^t is the element of $\mathbf{H}^t$. Let $T^\pi(\mathbf{s}_0)$ and $P^\pi(\mathbf{s}_0) = \sum_{t=0}^{T^\pi(\mathbf{s}_0)} d_t(\mathbf{s}_t)$ denote the total transmission time (in frames) needed and the total transmission power used for finishing the file transfer in a single realization of the policy π given the starting state $\mathbf{s}_0$, respectively. Note that $T^\pi(\mathbf{s}_0)$, and hence $P^\pi(\mathbf{s}_0)$, is a random variable dependent on both the policy and the initial state. Let $T_{avg}^\pi(\mathbf{s}_0) = \mathbb{E}_{\mathbf{s}_0}^\pi[T^\pi(\mathbf{s}_0)]$ and $P_{avg}^\pi(\mathbf{s}_0) = \mathbb{E}_{\mathbf{s}_0}^\pi[P^\pi(\mathbf{s}_0)]$ denote the average total transmission time and the average total transmission power of policy π, where $\mathbb{E}_{\mathbf{s}_0}^\pi$ is the expectation over policy π conditioning on the initial state $\mathbf{s}_0$. Finally, let $T_{avg}^{min} = min_\pi \mathbb{E}_{\mathbf{s}_0}^\pi[T_{avg}^\pi(\mathbf{s}_0)]$ denote the minimum average time for finishing the file transfer over all policies. Fig. 3.2 illustrates two example realizations (sample paths)

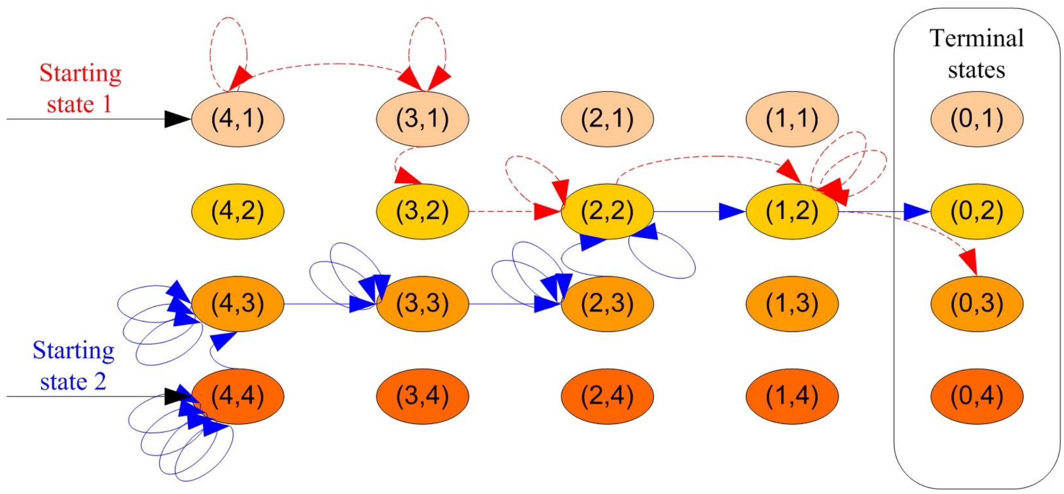

Figure 3.2: Example realizations of file transfer over a Markovian fading channel

of two policies, viz., the dashed and the solid arrow lines. Note that the transition probabilities in each state are dependent on both the selected action and the channel transitions. There are some costs associated with the paths, viz., the time (roughly represented with the number of the arrow lines in Fig. 3.2) and the power used to reach the terminal states. Thus different sample paths have different costs, and a policy determines the possible (cost of a) sample path through which the transmission will pass. Furthermore, the expected total cost of a (proper) stationary policy may be determined by using some suitable averaging functions.

Our objective is to find a policy that minimizes the average total transmission power for finishing the file transfer while satisfying some delay constraints. Let T_D be the delay constraint and assume that $T_D \geq T_{avg}^{min}$ to ensure the existence of a solution. We mainly consider the following two delay constraints. (1) average delay constraint and (2) strict delay constraint.

For the average delay constraint, $T_{avg}^{\pi}(\mathbf{s}_0) \leq T_D$. The mathematical formulation of the average delay constrained problem is given as :

Problem A: Given the initial state $\mathbf{s}_0$, find a policy π^* such that

$$\begin{aligned} P_{avg}^{\pi^*}(\mathbf{s}_0) = \min_{\pi} \left\{ P_{avg}^{\pi}(\mathbf{s}_0) \right\} \\ \text{subject to: } T_{avg}^{\pi^*}(\mathbf{s}_0) \leq T_D \end{aligned} \tag{3.11}$$

Note that the above definition is a typical constrained optimization problem. A direct solution is not easy to find. However, Problem A can be reformulated as a stochastic shortest path problem [7] and solved accordingly. Furthermore, the resulting optimal policy is a stationary policy. We discuss this in the next section. Note that the case of no delay constraint can be considered as a special case of the average delay constraint, which may also provide the lower bound of $P_{avg}(\mathbf{s}_0)$. We also discuss this in the next section.

For the strict delay constraint, $T^{\pi}(\mathbf{s}_0) \leq T_D$. The mathematical formulation of the strict delay constrained problem is given as:

Problem B: Given the initial state $\mathbf{s}_0$, find a policy π^* such that

$$\begin{aligned} P_{avg}^{\pi^*}(\mathbf{s}_0) = \min_{\pi} \left\{ P_{avg}^{\pi}(\mathbf{s}_0) \right\} \\ \text{subject to: } T^{\pi^*}(\mathbf{s}_0) \leq T_D \end{aligned} \tag{3.12}$$

Problem B can be reformulated as a finite horizon Markov decision process problem [62] and solved accordingly. But the resulting optimal policy is not a stationary policy and is dependent on the frame index.

3.3 Optimal Policy with Average Delay Constraint

3.3.1 The Stochastic Shortest Path Problem

We have briefly introduced the Stochastic Shortest Path (SSP) problem in Section 2.2.1 and we refer the reader to Bertsekas [7] for more discussions. In an SSP problem, the *termination state* should be specified. Furthermore, it should be assumed that the value of a *proper policy* is finite and the value of an *improper policy* is infinite. For our problem, the termination states can be defined to be when there is no residual file packet in the buffer, cf. Fig. 3.2. The proper policy can be defined as the policy that only contains proper decision rules which require the sender to transmit at least on some channel states (e.g., the best channel state). They are defined as follows.

Definition 3.1 (Terminal states) *The terminal states are defined as the states with no residual file packets in the buffer. Accordingly, let* $\mathcal{S}_F = \{(b, h_i)|b = 0\}$ *denote the set of terminal states.*

Definition 3.2 (Proper decision rules) *When the system state is* $\mathbf{s} = (b, h)$ *such that the buffer is not empty, i.e.,* $b > 0$ *and the current estimate of the channel state is in the best state* h_M*, a proper decision rule requires that a transmission should be scheduled, i.e.,* $d(b, h_M) > 0$ *if* $b > 0$*.*

The definition of proper decision rules can be relaxed by specifying more channel states in which the sender has to transmit when the buffer is not empty. Note that the definition does not exclude the situation in which a transmission is scheduled when the estimated channel is not in the best state. The proper policy thus only consists of the proper decision rules and we denote the set of proper policies as Π^p. As the channel transition matrix $\mathbf{H}$ is ergodic and irreducible, there is a positive probability to transit from a state (b, h_M) to a state $(b-1, h_M)$, (it may transit like $(b, h_M) \rightarrow (b-1, \cdot) \rightarrow (b-1, h_M)$) and hence the terminal states are reachable with probability 1 for all proper policies. To compute the average total transmission time, we define one frame delay cost as:

$$C_d(\mathbf{s}_t) = \begin{cases} 1, & \mathbf{s}_t \notin \mathcal{S}_F \\ 0, & \mathbf{s}_t \in \mathcal{S}_F \end{cases} \tag{3.13}$$

Note that (3.13) implies that the delay cost is not dependent on the particular number of residual file packets but is for the whole file. Given a policy π and an initial state $\mathbf{s}_0$, the average total transmission time can be computed as:

$$T_{avg}^{\pi}(\mathbf{s}_0) = \mathbb{E}_{\mathbf{s}_0}^{\pi}\left[\sum_{t=0}^{\infty} C_d(\mathbf{s}_t)\right] \tag{3.14}$$

Note that for proper policies, the terminal states are reachable with probability 1 and hence $T_{avg}^{\pi}(\mathbf{s}_0) < \infty$ for all $\pi \in \Pi^p$. An example of an improper policy can be that the sender does not transmit in all states. It is easy to see that for such an improper policy, the terminal states are not reachable and the average total transmission time of improper policies is infinite. Obviously, the improper policies are not of our interests. Hence we can restrict the search for optimal policies for Problem A only on the set of proper policies. To compute the average total transmission power, we define the power usage cost function in one frame as:

$$C_p(d_t(\mathbf{s_t})) = \begin{cases} d_t(\mathbf{s}_t), & \mathbf{s} \notin \mathcal{S}_F \\ 0, & \mathbf{s} \in \mathcal{S}_F \end{cases} \tag{3.15}$$

Given a policy π and an initial state $\mathbf{s}_0$, the average total transmission power can be computed as:

$$P_{avg}^{\pi}(\mathbf{s}_0) = \mathbb{E}_{\mathbf{s}_0}^{\pi}\left[\sum_{t=0}^{\infty} C_p(d_t(\mathbf{s}_t))\right] \tag{3.16}$$

Problem A can be converted into a family of unconstrained optimization problems through a Lagrangian approach [50]. For every $\beta > 0$, the Lagrange multiplier, define a mapping $C(\beta, \mathbf{s}, d(\mathbf{s})) : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}^+$ as:

$$C(\beta, \mathbf{s}_t, d_t(\mathbf{s}_t)) = \beta C_d(\mathbf{s}_t) + C_p(d_t(\mathbf{s}_t)) \tag{3.17}$$

Further, given the initial state $\mathbf{s}_0$, define a corresponding Lagrangian functional for a policy $\pi \in \Pi^p$ as:

$$V^{\pi}(\mathbf{s}_0, \beta) = \mathbb{E}_{\mathbf{s}_0}^{\pi}\left[\sum_{t=0}^{\infty} C(\beta, \mathbf{s}_t, d_t(\mathbf{s}_t))\right] \tag{3.18}$$

Clearly, (3.18) has a similar mathematical definition as that of a Markov decision problem with an expected total optimal criterion, cf. Section 2.2.2 and (2.2). Note that as $T_{avg}^{\pi}(\mathbf{s}_0)$ and $P_{avg}^{\pi}(\mathbf{s}_0)$ are finite for any proper policy $\pi \in \Pi^p$ and infinite for any improper policy, so is $V^{\pi}(\mathbf{s}_0, \beta)$. Now we define a stochastic shortest problem as:

Problem A': Given the initial state $\mathbf{s}_0$, find a policy $\pi^* \in \Pi^p$ such that

$$V^{\pi^*}(\mathbf{s}_0, \beta) = \min_{\pi \in \Pi^p} \{V^{\pi}(\mathbf{s}_0, \beta)\} \tag{3.19}$$

The solution to Problem A is closely related to that of Problem A'. The following proposition provides sufficient conditions under which an optimal policy for Problem A' is also optimal for Problem A.

Proposition 3.1 *Given the initial state $\mathbf{s}_0$ and for some $\beta > 0$, let $\pi^* \in \Pi^p$ be the optimal policy solving Problem A'. Further, if π^* meets the minimum average delay constraint, i.e., $T_{avg}^{\pi^*}(\mathbf{s}_0) = T_D$, then the policy π^* is also optimal for Problem A.*

Proof: For briefness, we omit the dependence on the initial state. If the optimal policy $\pi^* \in \Pi^p$ solves Problem A', then for any policy $\pi \in \Pi^p$, we have

$$V^{\pi^*}(\beta) = \beta T_{avg}^{\pi^*} + W_{avg}^{\pi^*} \leq \beta T_{avg}^{\pi} + W_{avg}^{\pi} = V^{\pi}(\beta) \tag{3.20}$$

Furthermore, as the solution to Problem A can only be from the set of proper policies (otherwise the average transmission time is infinite), we have $T_{avg}^{\pi} \leq T_D$. As π^* also meets the constraint, i.e., $T_{avg}^{\pi^*} = T_D$, the inequality (3.20) and the fact $\beta > 0$ readily imply that:

$$W_{avg}^{\pi^*} \leq W_{avg}^{\pi} + \beta(T_{avg}^{\pi} - T_{avg}^{\pi^*}) \Rightarrow W_{avg}^{\pi^*} \leq W_{avg}^{\pi} \tag{3.21}$$

for all policy $\pi \in \Pi^p$. Hence the policy π^* is also optimal for Problem A. ■

Indeed, β can be seen as the weight between the delay cost and the power consumption cost. The use of a smaller β indicates that the transmission power is emphasized more than the delay consideration, and a larger β puts more emphasis on the transmission delay aspect. Consider a special case in which $\beta = 0$ for Problem A'. This can correspond to the case of no delay constraint and it may be shown that the resulting optimal policy consumes the minimum power. Without any delay constraint, the sender can transmit by using the minimum power only on the best channel state. Although such a transmission policy minimizes the total power, it may incur a very large file transfer delay which is obviously undesirable.

Problem A' can be solved by using some standard techniques, such as value iteration, cf. Section 2.2.4, or policy iteration [7], and the optimal policy can be found by iteratively adjusting the value of β to meet the average delay constraint.

3.3.2 Numerical Examples

We consider a slow Rayleigh fading channel. We set the channel average SNR at 10dB when the normalized noise is equal to one. We classify the channel states by partitioning the channel into SNR intervals each with the same steady state probability. The channel transition matrix is constructed according to the method introduced in Section 3.1. The maximum Doppler shift is set as 50 Hz. An example of the channel transition matrix is shown in Table 3.1. The forward error correction (FEC) code

Table 3.1: Channel transition matrix ($h_{ij} = 0$ for all $|i-j| > 1$, $f_d = 50$Hz, $R_s = 62000$ symbols/second,$M = 8$, $\bar{h} = 10$dB)

state i	1	2	3	4	5	6	7	8
$h_{i,i-1}$	-	0.0051	0.0069	0.0068	0.0064	0.0060	0.0055	0.0050
$h_{i,i}$	0.9949	0.9880	0.9863	0.9868	0.9876	0.9885	0.9895	0.9950
$h_{i,i+1}$	0.0051	0.0069	0.0068	0.0064	0.0060	0.0055	0.0050	-

BCH[31,21, 2] is used and the Quadrature Phase Shift Keying (QPSK) modulation scheme is employed. The duration of a frame is set at 5ms. The length of the file is set at 5040 bits. Each frame transmits a 620-bit packet (including the error correction bits) and hence there are altogether 12 coded file packets to transmit. As a numerical illustration, we first consider 5 transmission powers and the action space is set as $\mathcal{A} = \{0, 6, 8, 10, 12, 14\}$. The average frame success probabilities $f_s(a, h_i)$ are obtained from Monte Carlo simulations. Table 3.2 shows some values of the frame success probabilities from the Monte Carlo simulations.

We first present two example optimal policies computed via the value iteration algorithm, which are shown in Tables 3.3 and 3.4, respectively. From the numerical examples, we have the following observations. First, the sender uses lower power levels when the channel state is good. Obviously, this is based on the fact that the sender

Table 3.2: Channel states and average frame success probabilities (frame length = 5 ms, $R_s = 62000$ symbols per second, BCH[31,21,2] and QPSK)

State	SNR (dB)	$f_s(6,\cdot)$	$f_s(8,\cdot)$	$f_s(10,\cdot)$	$f_s(12,\cdot)$	$f_s(14,\cdot)$
1	$(-\infty, 1.26)$	0.0	0.0	0.0	0.0	0.0
2	$[1.26, 4.59)$	0.0	0.0	0.0	0.0136	0.1902
3	$[4.59, 6.72)$	0.0	0.0	0.0907	0.5398	0.8831
4	$[6.72, 8.41)$	0.0	0.0545	0.6493	0.9550	0.9976
5	$[8.41, 9.92)$	0.0053	0.5357	0.9612	1.0	1.0
6	$[9.92, 11.42)$	0.1541	0.9326	0.9992	1.0	1.0
7	$[11.42, 13.18)$	0.7304	0.9976	1.0	1.0	1.0
8	$[13.18, \infty)$	0.9800	1.0	1.0	1.0	1.0

Table 3.3: Optimal actions ($\beta = 1.0$, $D_c = 0$)

	buffer											
channel	12	11	10	9	8	7	6	5	4	3	2	1
8	6	6	6	6	6	6	6	6	6	6	6	6
7	8	8	8	8	8	8	8	8	8	8	8	8
6	8	8	8	8	8	8	8	8	8	8	8	8
5	10	10	10	10	10	10	10	10	10	10	10	10
4	12	12	12	12	12	12	12	12	12	12	12	12
3	14	14	14	14	14	14	14	14	14	14	14	14
2	0	0	0	0	0	0	0	0	14	14	14	14
1	0	0	0	0	0	0	0	0	0	0	0	0

can benefit (energy-saving) from knowing (or estimating) the channel. We then observe that the optimal actions are almost not sensitive to the number of residual file packets, e.g., when the channel is in the best state, the optimal action is the same, i.e., using the least transmission power, regardless of the number of the residual packets. This is because the delay cost is not based on the number of packets in the buffer but for the whole file instead. However, we also observe that the sender has to transmit even when the channel is in bad states when the number of buffered packets is small, e.g, $d(4,2) = 14, d(3,2) = 14, \cdots$, in Table 3.3. This is reasoned as follows. In general, the fewer the buffered file packets, the less average transmission time is needed to

Table 3.4: Optimal actions ($\beta = 0.01$, $D_c = 0$)

	buffer											
channel	12	11	10	9	8	7	6	5	4	3	2	1
8	6	6	6	6	6	6	6	6	6	6	6	6
7	8	8	8	8	8	8	8	8	8	8	8	8
6	8	8	8	8	8	8	8	8	8	8	8	8
5	0	0	0	0	0	0	0	0	10	10	10	10
4	0	0	0	0	0	0	0	0	0	12	12	12
3	0	0	0	0	0	0	0	0	0	0	0	14
2	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0

finish the file transfer. Also energy saving may not be significant when the number of file packets is small. Hence trading off between the (possible) energy-savings and the average finishing times, it may be better to use high transmission power levels during bad channel states to avoid incurring excessive delays in completing the file transfer. Finally, when the delay cost is small compared with the power consumption, i.e., when a small β is used, the sender can choose not to transmit in the worst channel states. This is illustrated by the larger number of non-transmissions, i.e., $d(b, h) = 0$, in Table 3.4 compared with Table 3.3. However, the average time for finishing the file transfer also increases. This is clearly shown in the following two tables.

Table 3.5 and Table 3.6 present the Monte Carlo simulation results for the average total transmission power and the average total transmission delay (frames) of the two optimal policies. The right most columns of the two tables are the averages over all initial channel states. Note that the average total costs can be derived as the average total power plus β times the average total delay. It is clearly seen from the two tables that the average total transmission powers and delays are heavily dependent on the initial channel state. This is because when starting from a bad channel state, the sender has to wait for the channel to transit to some better states in order to reduce transmission power and hence, save energy. However, the sender still has to transmit in some bad channel states to meet the delay constraint. We also note that the average power consumption (averaging over all initial channel states) can be reduced greatly

Table 3.5: Average total power consumption and average total transmission delay (frames) under different initial channel states ($\beta = 1.0$, $D_c = 0$)

	initial channel state								
	1	2	3	4	5	6	7	8	average
Costs	704.69	508.96	223.18	158.53	137.32	116.40	107.82	86.46	255.42
Powers	192.33	192.22	192.06	145.98	124.84	103.57	95.76	74.22	140.13
Delays	512.36	316.75	31.12	12.55	12.49	12.28	12.17	12.03	115.22

Table 3.6: Average total power consumption and average total transmission delay (frames) under different initial channel states ($\beta = 0.01$, $D_c = 0$)

	initial channel state								
	1	2	3	4	5	6	7	8	average
Costs	127.14	125.16	122.34	117.92	111.67	103.3	95.87	74.32	109.72
Powers	102.84	102.79	102.62	102.34	102.31	102.1	95.74	74.20	98.12
Delays	2430	2237	1952	1508	882.3	52.3	12.12	12.02	1135

by relaxing the delay constraint, which can be seen from the right most columns of the two tables. An extreme case could be that the energy saving is maximized by transmitting with the least power and in the best channel state only[1]. Finally, this example also clearly illustrates why it is better to consider file transfer as *background* traffic, as suggested in UMTS [21].

Fig. 3.3 compares the average total cost, the average total power and the average transmission delay (all are averaged over all initial states) of the optimal policy with those of a persistent transmission policy, i.e., the sender knows nothing about the channel and it transmits persistently in all frames with the same power. The labels on the x-axis are the transmission powers used in all frames of the corresponding persistent transmission policy. Obviously, using variable transmission power levels based on the

[1]However, this may impact on the end-to-end QoS when TCP is employed, since the trigger of TCP retransmission may also result in undesirable waste of the whole communication network resources, e.g., the bandwidth consumption due to undesirable TCP retransmission between the mobile station and a remote file server.

channel state has lower total cost than using a fixed transmission power without any information about the channel state. We note that the average total transmission

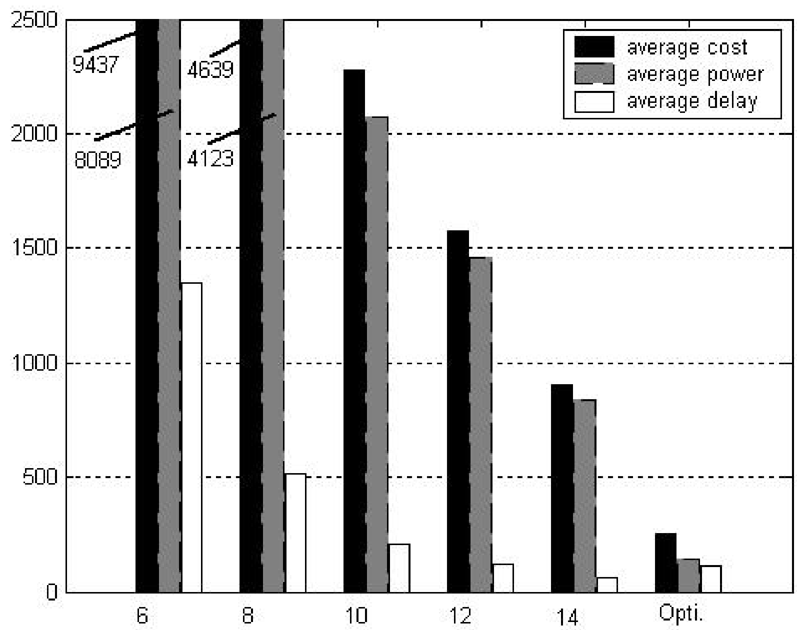

Figure 3.3: Performance comparison of different persistent policies with the optimal policy (channel states = 8, available actions $\{0, 6, 8, 10, 12, 14\}$, $\beta = 1.0, D_c = 0$)

delay in the persistent policy with a power level of 14 is less than that of the optimal policy. However, this is achieved by using a much higher transmission power level and hence the total cost is still much higher than that of the optimal policy. Indeed, when the sender knows the channel fading level exactly, it can use a power level that is the (multiple) inverse of the channel fading [2]. This is a form of *water-filling* power allocation policy to approximate the achievable channel capacity (the Shannon capacity, cf. (1.1)). Note that the solutions to the Problem A' (cf. Table 3.5 and Table 3.6) are structurally similar to water-filling power allocation. Goldsmith has investigated the achievable capacity of a fading channel via dynamic water-filling power (and rate) allocation [29, 30]. Her approach is from the information theoretic point of view and is based on the asymptotic analysis. Hence her approach does not consider the delay

[2]This helps to maintain a constant average frame success probability, however, it needs the assumption of continuous transmission powers.

cost, and also the tradeoff between the delay cost and power consumption. However, our approach is from a decision theoretic point of view and takes the delay cost into consideration as well.

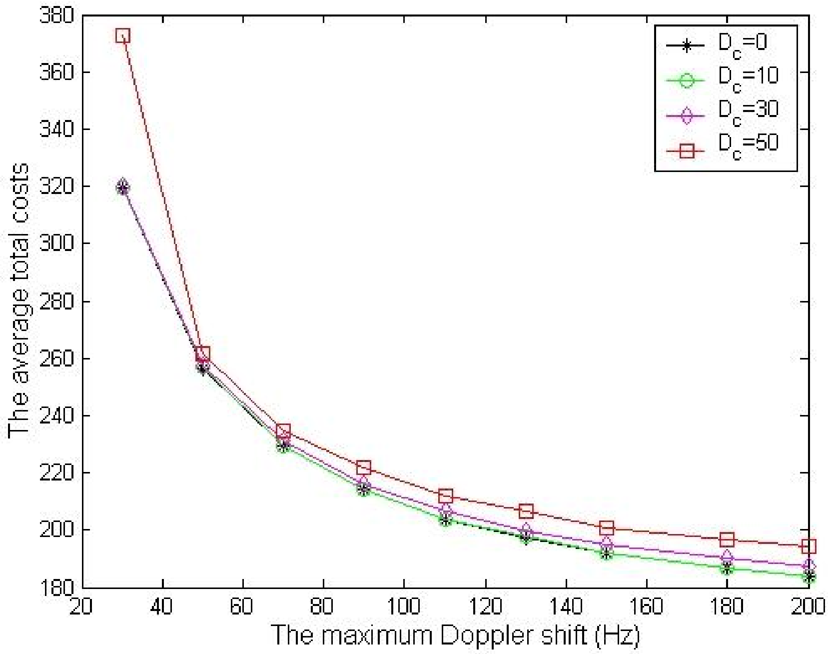

Figure 3.4: The average total costs of different optimal policies (channel states = 8, $\mathcal{A}$=$\{0, 6, 8, 10, 12, 14\}$ and $\beta = 1.0$).

We next investigate the effect of the variation of the channel transition probabilities and the delay incurred by channel state reports on the performance of the optimal policies. The maximum Doppler shift is used to characterize the variation of a slow Rayleigh fading channel. Figs. 3.4, 3.5 and 3.6 plot the average total costs, the average total power and the average total delay, respectively, of different optimal policies that are computed from the different values of the maximum Doppler shift and the delay incurred by channel state reports, D_c. All values in the figures are averaged over all starting states for each optimal policy and are from Monte Carlo simulations. From Fig. 3.4, we observe that the average total costs (i.e., the optimal policy values) decease monotonically with the maximum Doppler shift. This is because a higher maximum Doppler shift increases the transition probabilities between channel states, cf. (3.4), (3.5) and (3.6). Consequently, the sender incurs less delay costs in waiting

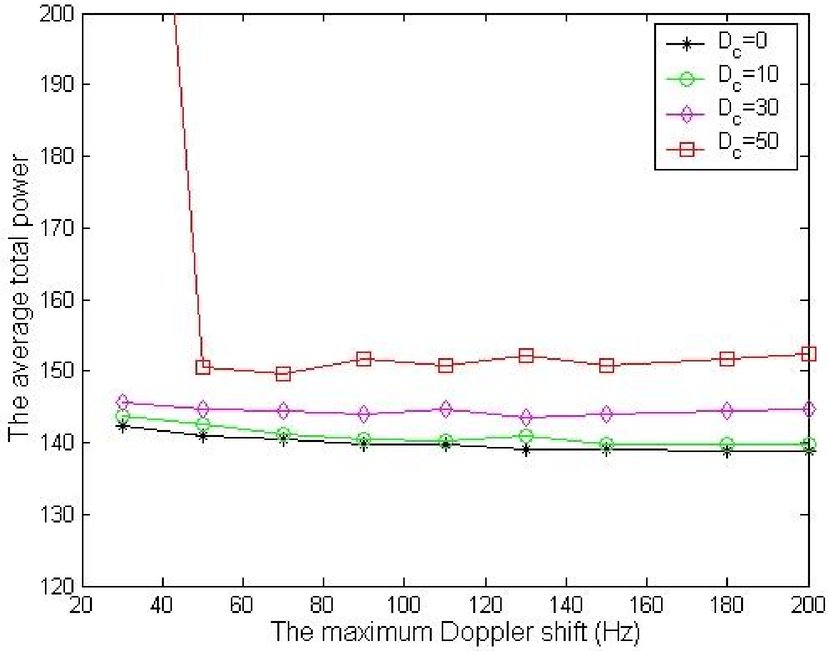

Figure 3.5: The average total powers of different optimal policies (channel states = 8, $\mathcal{A}$=$\{0, 6, 8, 10, 12, 14\}$ and $\beta = 1$).

for good channel states (with high SNR intervals) to transmit in, which also requires lower transmission levels. Clearly, this shows the benefit of using a channel dependent transmission policy. From Fig. 3.4, we also observe that the larger D_c, the higher the optimal policy values given the same maximum Doppler shift. When $D_c > 0$, the optimal policies are computed based on the inaccurate (outdated) system states information which causes misinterpretation of the channel states. In this case, the resulting optimal policies require the sender to transmit in more bad channel states when D_c increases. Hence the average total power increases greatly with the larger delay of the channel state report, as shown in Fig. 3.5. On the other hand, since the sender transmits in more bad channel states, the average total time for finishing the file transfer is reduced, as shown in Fig. 3.6. However, the total costs (i.e., power plus delay) still increase with the delay of the channel state reports, as shown in Fig. 3.4. This also suggests that the outdated information results in inaccurate (optimal) policies and may degrade the performance of the policies in our case.

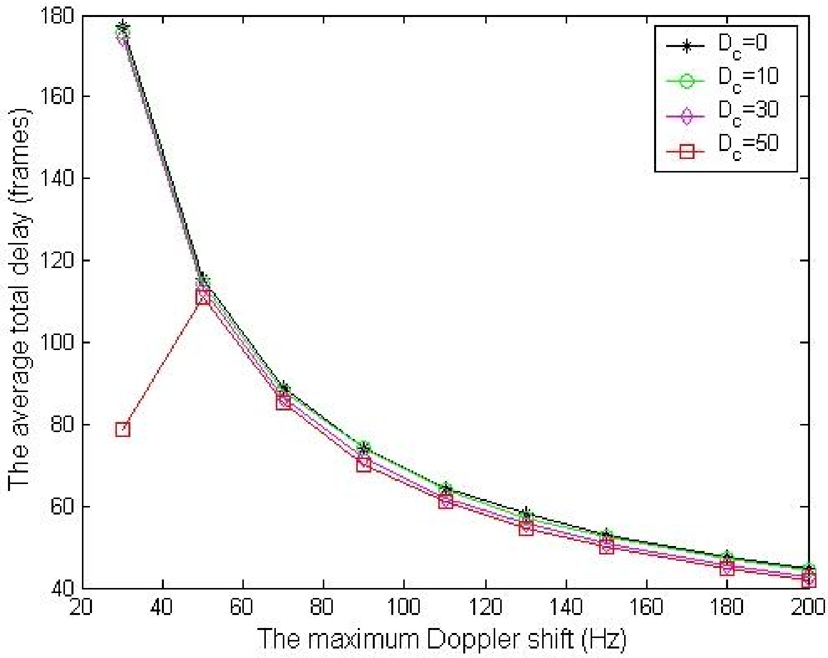

Figure 3.6: The average total delay of different optimal policies (channel states = 8, $\mathcal{A}$=$\{0, 6, 8, 10, 12, 14\}$ and $\beta = 1$).

3.4 Optimal Policy with Strict Delay Constraint

We now consider the problem with a strict delay constraint, i.e., Problem B (3.12). The goal is to find the optimal power allocation policy that minimizes the power consumption used to finish a file transfer under a strict delay constraint.

3.4.1 The Finite Horizon Dynamic Programming Problem

The period of the file transfer is required to be less than T_D frames. We thus consider an optimal policy design problem only spanning over T_D frames and reformulate Problem B as a finite horizon dynamic programming problem. Note that we do not need to assume the proper decisions as there always exists a deterministic Markovian optimal policy for a finite horizon dynamic programming problem with a finite state space and a finite set of actions (see Theorem 4.2.2. and Proposition 4.4.3 of Puterman [62]). Similar to the solution for Problem A', to compute the average total transmission power, we define a one frame power usage cost function as:

$$C_p(d_t(\mathbf{s_t})) = \begin{cases} d_t(\mathbf{s}_t), & \mathbf{s} \notin \mathcal{S}_F \ \text{ and } \ t < T_D \\ 0, & \textit{otherwise} \end{cases} \tag{3.22}$$

where $\mathcal{S}_F$ is the set of terminal states defined as before (see Definition 3.1). At the end of a transmission period, it is possible that some file packet(s) may not be transmitted. Hence, to compute optimal policies, we need to define a terminal cost function to allocate a cost (i.e., negative reward) when there are still file packet(s) left in the buffer after T_D frames. We consider the following simple terminal cost function.

$$C_f(b) = c_0 b^{c_1}, \quad c_0 > 0, c_1 \geq 1 \tag{3.23}$$

Let $V^\pi(\mathbf{s}_0, T_D)$ denote the expected total cost over a T_D-frame decision making horizon if policy π is used and the starting state is $\mathbf{s}_0$. This is defined by

$$V^\pi(\mathbf{s}_0, T_D) = \mathbb{E}^\pi_{\mathbf{s}_0}\left[\sum_{t=0}^{T_D-1} C_p(d_t(\mathbf{s_t})) + C_f(\mathbf{s}_{T_D})\right] \tag{3.24}$$

Now Problem B can be considered as the solution to the following standard finite horizon dynamic programming problem which is to find a policy π^* which has the minimum policy value:

$$V^{\pi^*}(\mathbf{s}_0, T_D) = V^*(\mathbf{s}_0, T_D) \equiv \min_\pi V^\pi(\mathbf{s}_0, T_D) \tag{3.25}$$

Again, some fairly standard techniques, e.g., *the backward induction algorithm* (see Puterman [62] page 92), can be used to solve such a problem and to provide the optimal policy. Note however that the computed optimal policy is not a stationary policy but is dependent on the frame index. We provide some numerical examples in the next subsection.

3.4.2 Numerical Examples

As an illustration, we consider the same channel model used in Section 3.3.2. The maximum Doppler shift is set at 50Hz and other parameters are set as before. For $D_c = 0$ and $T_D = 30$, Figs. 3.7 and 3.8 plot the optimal actions when there are 5 packets and 10 packets, respectively, left in the buffer at all decision epochs. Note that

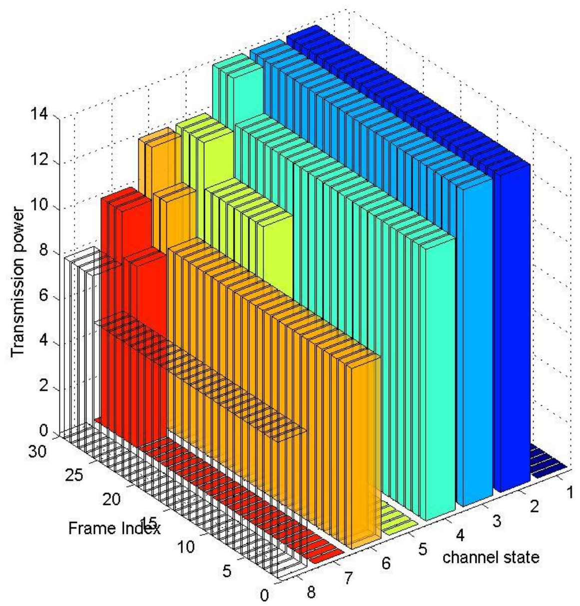

Figure 3.7: Optimal actions when there are 5 packets left in the buffer for all decision epochs. ($c_0 = 500$ and $c_1 = 2$)

the optimal actions now are dependent on the frame index. Compared to the average delay constraint, cf. Table 3.3, the sender has to transmit in some bad channel states, e.g., channel states 2 and 3, with higher transmission power levels during the whole decision period. Under the strict delay constraint, the sender knows that transmitting even with a very lower frame success probability, e.g., $f_s(14, 2) = 0.19$, is better than not transmitting to avoid the much higher terminal cost at the end of the transmission period. On the other hand, the optimal policy still exploits the channel variation to some extent in order to save energy. For example, from Fig. 3.7, we see that the sender does not transmit when the channel is in state 7 and there are 5 packets in the buffer during the first 25 frames. In this case, the sender can benefit from waiting. If the channel transits to a better state, i.e., state 8 in this case, the sender can save transmission power. If the channel transits to a worse state, i.e., state 6 in this case, the sender only needs to use the same transmission power level (8 in this case) in the worse state as in the current state since the frame success probabilities are very close in the two channel states. This is seen from Table 3.2 that $f_s(6, 7) = 0.73$, $f_s(6, 8) = 0.98$, and

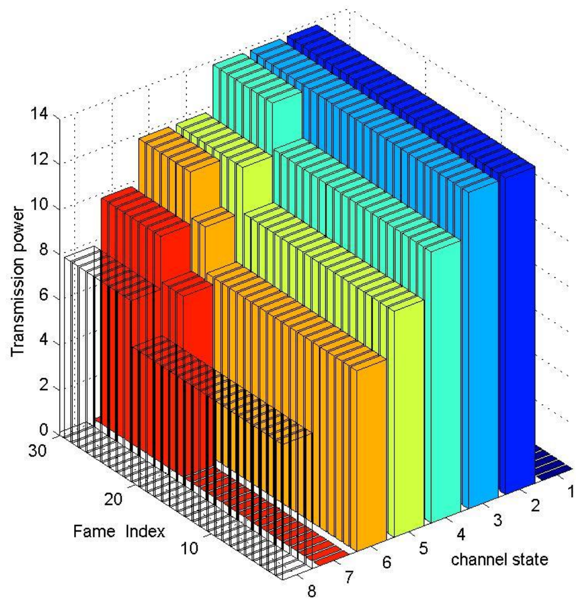

Figure 3.8: Optimal actions when there are 10 packets left in the buffer for all decision epochs. ($c_0 = 500$ and $c_1 = 2$)

$f_s(8, 6) = 0.9326$ and $f_s(8, 7) = 0.9976$. However, we note that this assumes that the sender has full knowledge of the channel variations when computing optimal policies. Comparing Figs. 3.7 and 3.8, we see that the sender is more selective in transmitting when there are fewer residual packets, i.e., there are more states that the sender does not transmit in Fig. 3.7. This is because the sender has more time left to finish the file transfer.

Fig. 3.9 compares the performance of the computed optimal policy (via the backward induction algorithm) with some fixed power persistent transmission policy. Again, the labels on the x-axis are the transmission power levels used. The success probability is the probability that the file transfer can be finished within the decision period, e.g., 30 frames in this case, averaged over all initial channel states. The normalized power consumption is the average transmission power per frame used during the file transfer for each policy, normalized with the maximum available transmission power. Obviously, increasing the transmission power level helps to mitigate the effect of the fading channel and hence increases the success probability. However, from Fig. 3.9, we

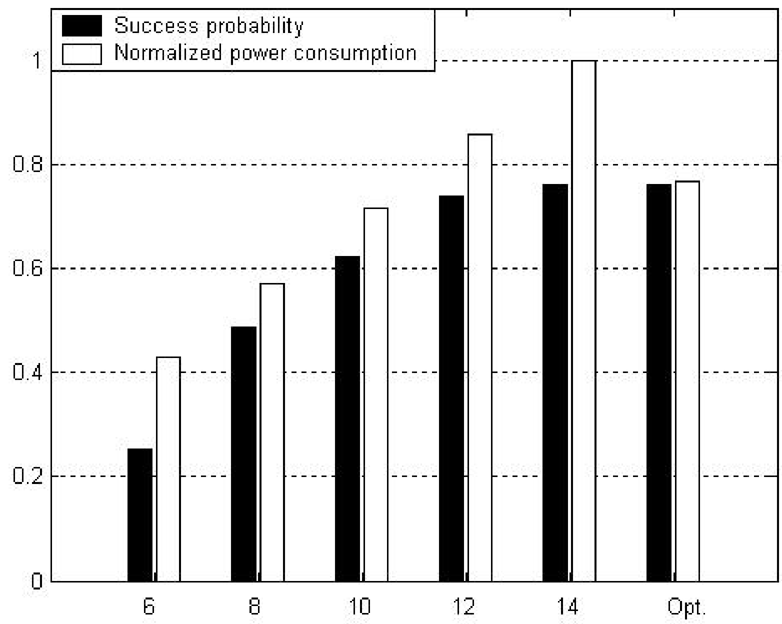

Figure 3.9: Comparisons between different policies within the decision period (T_D = 30).

clearly see that the success probability of the optimal policy is similar to the persistent policy with a power level 14 but the power consumption is significant less. This again suggests that the optimal policy can exploit and benefit from knowing the channel state variations.

3.5 Summary

In this chapter, we have studied the problem of transferring a finite size file over a slow Rayleigh fading channel with two types of delay constraint. The sender decides, at the beginning of each frame, whether or not to transmit a file packet. If it transmits, the transmission power level to use is based on the system (channel and buffer) state. The goal is to minimize the power consumption while meeting the delay constraint. We have shown how to convert such a constrained optimization problem to a standard Markov decision problem. We have presented simulation results that indicate that transmission power can be substantially reduced with optimal policies which exploit knowledge of the channel variations to meet the delay constraints.

We have not considered source dynamics in this chapter. In the next chapter, we will consider a simple transmission control problem in which both source dynamics and channel dynamics are considered.

Chapter 4

Optimal Transmission Control Policies

In this chapter, we study a transmission control problem for a single data user over a time-varying channel. The arrival dynamics and the channel dynamics are considered, however, the action is simplified as a binary action, i.e., either to transmit or not to transmit. The objective is to find a policy that optimally balances different costs. We investigate the characteristic and structure of optimal policies. Numerical examples and some discussions are provided.

4.1 Problem Formulation

The system model is shown in Fig. 4.1. We consider a discrete time system. At the beginning of each frame, the sender needs to decide whether or not to transmit a packet in the frame based on the observation of the buffer occupancy and the channel state information. In each frame, there is a batch of data packets of the same size arriving at the sender's buffer. Arriving packets are queued in a first-in-first-out buffer that can hold at most B packets. If the buffer is full, arriving packets are discarded, i.e., the buffer overflows. We assume that batch arrivals in different frames are independent and cannot be transmitted in the same frame. Let $q(i), i = 0, 1, \cdots$, denote the probability of i packets arriving in a frame and $\sum_i q(i) = 1$. Let λ denote the average number of arrivals in a frame; $\lambda = \sum iq(i)$. To guarantee a stable queueing system, we require

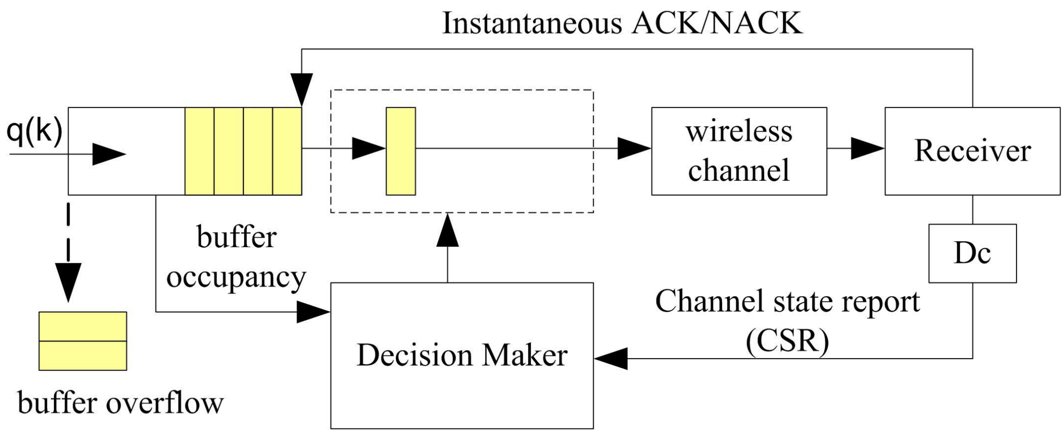

Figure 4.1: System model

$\lambda < 1$ which implies $q(0) > 0$. The wireless channel is modelled again as a finite-state Markov channel with M states and as in the previous chapter, we use $\mathbf{H} = [h_{ij}]_{M\times M}$ to denote the channel state transition matrix. We further assume that the Markov chain over the channel states is ergodic and irreducible. As fixed transmission power is used, the frame success probability is a function of the channel state, h_i, only and is denoted as $f_s(h_i)$. The system state is denoted again as $\mathbf{s} = (b, h_i)$, where $0 \le b \le B$ and $1 \le i \le M$. The action space now contains only two actions $\mathcal{A} = \{0, 1\}$, where $a = 0$ means no transmission and $a = 1$ means to transmit a packet with a fixed power level. The transition probability is then given as:

$$\mathrm{Tr}((b+z-1, h_j)|(b, h_i), a) = \begin{cases} q(z-1)h_{ij}, & b \ge 0, a = 0 \\ q(z-1)h_{ij}(1-f_s(h_i)) + q(z)h_{ij}f_s(h_i) & b > 0, a = 1 \\ 0 & otherwise \end{cases} \tag{4.1}$$

where $z \ge 0$ and $q(z < 0) = 0$. Note that we assume $b + z \le B$. If $b + z > B$, we can redistribute the *excess probability* to state $(B, \cdot)$ by using the *augmentation procedure* introduced in [73].

Compared to the problem in the previous chapter, here, we consider the arrival dynamics but simplify the set of actions. The simplification leads to some rigorous qualitative analysis which will be presented later. Since two stochastic processes, i.e., the buffer evolution and the channel process, have been included in this model, the analysis of multi-level or continuous transmission powers becomes much more complicated. However, extension to multi-level or continuous transmission powers can be obtained via asymptotic analysis against the information theoretic framework of A. Goldsmith [29, 30], for example, where for simplicity, no delay is considered.

The transmission control problem under consideration here can be formulated for different purposes. For example, the objective can be to minimize the transmission power, or to minimize the buffer overflow probability or to maximize the transmission efficiency while meeting an average delay requirement. In the previous chapter, we have introduced how to convert a constrained optimization problem into a standard MDP problem (cf. Section 3.2 and 3.3.1) and how to solve them via the Lagrangian approach. Indeed, different objectives can be represented via different cost functions and a particular objective can be emphasized by adjusting its cost function. Here we use the following cost structure:

$$C((b,h),a) = \begin{cases} c_0 b, & a = 0 \\ c_0 b + c_1 - c_2 f_s(h), & a = 1 \end{cases} \tag{4.2}$$

where $c_0, c_1, c_2 > 0$ are constants. The above cost structure indicates that a user pays some usage costs, i.e., power, for the transmission and some holding costs, i.e., delay, for queueing packets in the buffer. However, it benefits from a successful transmission with some probability, i.e., $-c_2 f_s(h)$. Note that the cost structure is uniformly bounded. By adjusting c_0, c_1, c_2, we can adjust the balance between transmission throughput, energy efficiency and delay cost. In general, the larger $c_0/c_1/c_2$ is, the more emphasis is placed on the delay/energy/throughput efficiency. However, in order to encourage transmission, a packet should be transmitted when the channel is in the best state [1]. Accordingly, we require $c_2 f_s(h_M) > c_1$ (at least). We will use the average optimality criterion to form a MDP modelled problem in the next section.

4.2 Average Cost Optimal Policy

We use the average cost optimality criterion as it implies the cost is not sensitive to when the cost is incurred and the computed policy value is independent of the starting

[1] This is a rather conservative requirement, which can be relaxed if we emphasize more on transmission.

state. To apply the ACOE (2.12) to compute optimal policies, we need to identify the unichain property. We use Definition (3.2) here again and let Π^p denote the set of policies that consist of only proper decision rules. The following proposition shows that the Markov decision process in our problem is unichain.

Proposition 4.1 *The Markov decision process under any transmission policy $\pi \in \Pi^p$ is unichain.*

Proof: As the system state is described by a two-tuple, i.e., $\mathbf{s} = (b, h)$, any stationary policy induces a two-dimensional Markov chain. Under our channel state model, we note that a system state (b, h_i) communicates with all other states $(b, \cdot)$ for all fixed b. Now consider states $(\cdot, h_M)$. Under the requirement of proper decision rules, a packet should be transmitted with a positive power level, and it will be received with a positive probability $f_s(h_M) > 0$. Furthermore, since $q(0) > 0$, it is always possible for the process to transit from state (j, h_M) to state $(j-1, h_M)$. On the other hand, as $q(k) > 0$, $k > 0$, it is also possible for the process to transit from $(j-1, h_M)$ to (j, h_M). As j is chosen arbitrarily, all states $(\cdot, h_M)$ can communicate with a distinguished state $(0, h_M)$. Based on the above arguments, all states communicate with state $(0, h_M)$. Therefore, all states form a single aperiodic positive class containing state $(0, h_M)$ and the proposition follows from the unichain definition. ■

The state and the action are finite and the immediate cost is uniformly bounded. Furthermore, as our transmission control problem is unichain under the proper decision rules definition, all conditions in Theorem 2.3 are satisfied and we can use ACOE (2.12) to compute an optimal policy.

4.3 Property of Optimal Policies

The computation effort for optimal policies increases as the state space increases, i.e., the buffer limit B and/or the number of channel states M increase. Though the available actions are few, the computation effort could be high. For example, when the value iteration algorithm (cf. Section 2.2.4) is used, there are altogether $2 \times B \times M$ equations that need to be solved at each step. Hence in this section we explore the property of optimal policies in order to reduce the computation effort. As the stationary average cost optimal policies exist (cf. Section 4.2), the ACOE (2.12) for our problem is rewritten as follows by using the transition probability (4.1) and the cost structure (4.2).

$$\begin{aligned} \overline{u}(0,h_i) &= -J + \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \overline{u}(z,h_j), & b = 0 \\ \overline{u}(b,h_i) &= -J + \min\left\{X(b), X(b) + \Delta(b,h_i)\right\} & b > 0 \end{aligned} \tag{4.3}$$

where

$$X(b) = c_0 b + \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \overline{u}(b+z,h_j) \tag{4.4}$$

$$\Delta(b,h_i) = c_1 - c_2 f_s(h_i) - f_s(h_i) U(b) \tag{4.5}$$

and

$$U(b) = \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \left[\overline{u}(b+z,h_j) - \overline{u}(b+z-1,h_j)\right] \tag{4.6}$$

We use the superscript k to index the kth step value functions in the value iteration algorithm, e.g., $\overline{u}^k(b,h_i)$, and we define $X^k(b)$, $\Delta^k(b,h_i)$ and $U^k(b)$ accordingly, for example, $X^k(b) = c_0 b + \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \overline{u}^k(b+z,h_j)$. Recall that a function $g(x) : \mathbb{N}^+ \mapsto \mathbb{R}$ is defined to be convex if for all $x = 1, 2, \cdots$,

$$g(x+1) + g(x-1) \geq 2g(x). \tag{4.7}$$

The following lemma states the convexity of $\overline{u}$.

Lemma 4.1 *$\overline{u}(b,h_i)$ is a convex function of b for each fixed h_i, $i = 1, \cdots, M$.*

Proof: The proof is based on the relative value iteration algorithm (see Section 2.2.4) and proceeds by induction. For notational simplicity, we let J^{k-1} denote $w^{k-1}(\tilde{s})$ used in the iteration algorithm, i.e., (2.18) and (2.19). Note that J^{k-1} is a constant in each step to compute $\overline{u}^k$ in (2.19). For $k = 0$, we set $\overline{u}^0(b,h_i) = 0$ for all b and h_i.

For $k = 1$, we have $\overline{u}^1(b,h_i) = \min\{c_0 b, c_0 b + c_1 - c_2 f_s(h_i)\}$ and hence $\overline{u}^1(b,h_i)$ is a convex function in b for each fixed h_i, $i = 1, \cdots, M$.

Now assume that $\overline{u}^k(b,h_i)$ is convex for some $k \geq 1$ and for each fixed h_i, we have

$$\overline{u}^k(b+1,h_i) + \overline{u}^k(b-1,h_i) \geq 2\overline{u}^k(b,h_i) \text{ for all } b > 1. \tag{4.8}$$

We show that it also holds for $k+1$. According to the iteration algorithm, we have

$$\begin{aligned}
\overline{u}^{k+1}(b+1,h_i) &= -J^k + \min\left\{X^k(b+1), \quad X^k(b+1) + \Delta^k(b+1,h_i)\right\},\\
\overline{u}^{k+1}(b,h_i) &= -J^k + \min\left\{X^k(b), \quad X^k(b) + \Delta^k(b,h_i)\right\},\\
\overline{u}^{k+1}(b-1,h_i) &= -J^k + \min\left\{X^k(b-1), \quad X^k(b-1) + \Delta^k(b-1,h_i)\right\}.
\end{aligned}$$

In the state $(\cdot, h_i)$, when $\Delta^k(\cdot, h_i) \geq 0$, the optimal action is not to transmit and when $\Delta^k(\cdot, h_i) < 0$, the optimal action is to transmit. We consider the following cases.

Case (1): The optimal action in state $(b+1, h_i)$ is not to transmit. This implies

$$\Delta^k(b+1,h_i) = c_1 - c_2 f_s(h_i) - f_s(h_i)U^k(b+1) \geq 0$$

From the induction assumption (4.8), we have

$$\begin{aligned}
U^k(b+1) &= \sum_{z=0} \sum_{j=1}^{M} q(z)h_{ij}\left[\overline{u}^k(b+z+1,h_j) - \overline{u}^k(b+z,h_j)\right]\\
&\geq \sum_{z=0} \sum_{j=1}^{M} q(z)h_{ij}\left[\overline{u}^k(b+z,h_j) - \overline{u}^k(b+z-1,h_j)\right]\\
&= U^k(b)\\
&\geq \sum_{z=0} \sum_{j=1}^{M} q(z)h_{ij}\left[\overline{u}^k(b+z-1,h_j) - \overline{u}^k(b+z-2,h_j)\right]\\
&= U^k(b-1)
\end{aligned} \tag{4.9}$$

Hence $\Delta^k(b,h_i) \geq 0$ and $\Delta^k(b-1,h_i) \geq 0$. The optimal actions in states (b,h_i) and $(b-1,h_i)$ are also not to transmit. In this case, we have

$$\begin{aligned}
& \overline{u}^{k+1}(b+1,h_i) + \overline{u}^{k+1}(b-1,h_i) - 2\overline{u}^{k+1}(b,h_i)\\
= \;& X^k(b+1) + X^k(b-1) - 2X^k(b)\\
= \;& \sum_{z=0} \sum_{j=1}^{M} q(z)h_{ij}\left[\overline{u}^k(b+z+1,h_j) - \overline{u}^k(b+z,h_j)\right]\\
& \quad - \sum_{z=0} \sum_{j=1}^{M} q(z)h_{ij}\left[\overline{u}^k(b+z,h_j) - \overline{u}^k(b+z-1,h_j)\right]\\
\geq \;& 0
\end{aligned} \tag{4.10}$$

The inequality is from the induction assumption (4.8) of the convexity of u^k.

Case (2): The optimal action in state $(b+1,h_i)$ is to transmit and the optimal action in state (b,h_i) is not to transmit. Via similar arguments in Case 1, the optimal action in state $(b-1,h_i)$ is also not to transmit. Furthermore, that the optimal action in state (b,h_i) is not to transmit implies

$$\Delta^k(b,h_i) = c_1 - c_2 f_s(h_i) - f_s(h_i)U^k(b) \geq 0. \tag{4.11}$$

In this case, we have

$$\begin{aligned}
& \overline{u}^{k+1}(b+1,h_i)+\overline{u}^{k+1}(b-1,h_i)-2\overline{u}^{k+1}(b,h_i) \\
= \;& X^k(b+1)+\Delta^k(b+1,h_i)+X^k(b-1)-2X^k(b) \\
= \;& c_1-c_2f_s(h_i)+(1-f_s(h_i))\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}^k(b+z+1,h_j)-\overline{u}^k(b+z,h_j)\right] \\
& \quad -\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}^k(b+z,h_j)-\overline{u}^k(b+z-1,h_j)\right] \\
\geq \;& f_s(h_i)\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}(b+z,h_j)-\overline{u}(b+z-1,h_j)\right] \\
& \quad +(1-f_s(h_i))\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}^k(b+z+1,h_j)-\overline{u}^k(b+z,h_j)\right] \\
& \quad -\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}^k(b+z,h_j)-\overline{u}^k(b+z-1,h_j)\right] \\
= \;& (1-f_s(h_i))\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}^k(b+z+1,h_j)-\overline{u}^k(b+z,h_j)\right] \\
& \quad -(1-f_s(h_i))\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}^k(b+z,h_j)-\overline{u}^k(b+z-1,h_j)\right] \\
\geq \;& 0
\end{aligned} \tag{4.12}$$

The first inequality is from (4.11) and the second inequality is from the induction assumption (4.8) of the convexity of $\overline{u}^k$.

Case (3): The optimal actions in states $(b+1,h_i)$ and (b,h_i) are to transmit and the optimal action in state $(b-1,h_i)$ is not to transmit. That the optimal action in state (b,h_i) is to transmit implies

$$\Delta^k(b,h_i)=c_1-c_2f_s(h_i)-f_s(h_i)U^k(b)\leq 0 \tag{4.13}$$

In this case, we have

$$\begin{aligned}
& \overline{u}^{k+1}(b+1,h_i)+\overline{u}^{k+1}(b-1,h_i)-2\overline{u}^{k+1}(b,h_i) \\
= \;& X^k(b+1)+\Delta^k(b+1,h_i)+X^k(b-1)-2\left[X^k(b)+\Delta^k(b,h_i)\right] \\
= \;& (1-f_s(h_i))\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}^k(b+z+1,h_j)-\overline{u}^k(b+z,h_j)\right] \\
& \quad -(1-f_s(h_i))\sum_{z=0}\sum_{j=1}^{M}q(z)h_{ij}\left[\overline{u}^k(b+z,h_j)-\overline{u}^k(b+z-1,h_j)\right]-\Delta^k(b) \\
\geq \;& 0
\end{aligned} \tag{4.14}$$

The inequality is from (4.13) of $-\Delta^k(b)\geq 0$ and the induction assumption (4.8) of the convexity of $\overline{u}^k$.

Case (4): The optimal actions in states $(b+1, h_i)$, (b, h_i) and $(b-1, h_i)$ are to transmit. In this case, we have

$$\begin{aligned}
& \overline{u}^{k+1}(b+1, h_i) + \overline{u}^{k+1}(b-1, h_i) - 2\overline{u}^{k+1}(b, h_i) \\
= \; & X^k(b+1) + \Delta^k(b+1, h_i) + X^k(b-1) + \Delta^k(b-1, h_i) - 2\left[X^k(b) + \Delta^k(b, h_i)\right] \\
= \; & (1 - f_s(h_i)) \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \left[\overline{u}^k(b+z+1, h_j) - \overline{u}^k(b+z, h_j)\right] \\
& - (1 - f_s(h_i)) \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \left[\overline{u}^k(b+z, h_j) - \overline{u}^k(b+z-1, h_j)\right] \\
& + f_s(h_i) \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \left[\overline{u}^k(b+z, h_j) - \overline{u}^k(b+z-1, h_j)\right] \\
& - f_s(h_i) \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \left[\overline{u}^k(b+z-1, h_j) - \overline{u}^k(b+z-2, h_j)\right] \\
\geq \; & 0
\end{aligned} \tag{4.15}$$

Again, the inequality is from the induction assumption (4.8) of the convexity of $\overline{u}^k$.

From (4.10), (4.12), (4.14), (4.15) and from the induction argument, we conclude that $\overline{u}^{k+1}(b, h_i)$ is a convex function of b for all $k \geq 0$, and from the MDP result of value iteration algorithm, $\overline{u}(b, h_i) = \lim_{k\to\infty} \overline{u}(b, h_i)$ is a convex function of b for each fixed h_i, $i = 1, \cdots, M$. This also completes the proof. ■

From Lemma 4.1, we then have the following corollary and proposition.

Corollary 4.1 *$\Delta(b, h_i)$ is a nonincreasing function of b for each fixed h_i, $i = 1, \cdots, M$.*

Proposition 4.2 *If there exists a buffer threshold $b^{(i)}$ such that the optimal action is to transmit in the state $(b^{(i)}, h_i)$ given some channel state h_i, then for all states $(b > b^{(i)}, h_i)$, the optimal action is to transmit too.*

Proof: From the proposition assumption, $\Delta(b^{(i)}, h_i) \leq 0$. The rest of the proof is clear from Corollary 4.1 and (4.3). ■

Proposition 4.2 suggests that an optimal policy has a structural property which reduces the computation effort for optimal policies. There may exist many buffer state thresholds (assuming that the buffer limit is large enough) each corresponding to a particular channel state. In [98], Zhang and Wasserman have proved that an optimal policy has a *back-off* structure for an always backlogged user, i.e., whenever the estimated channel state is in some bad channel states, the optimal policy is not to transmit and wait until the channel transits to some good states. However, with the arrival process included in our model, the sender still has to transmit in some bad states to avoid increasing the holding cost whenever the buffer occupancy exceeds some thresholds. We next show the property of optimal policies related to the channel states.

Lemma 4.2 *$\overline{u}(b, h_i)$ is a nondecreasing function of b for each fixed h_i, $i = 1, \cdots, M$.*

Proof: Again, the proof is based on the relative value iteration algorithm (see Section 2.2.4) and proceeds by induction. For $k = 0$, we set $\overline{u}^0(b, h_i) = 0$ for all b and h_i.

For $k = 1$, we have $\overline{u}^1 = \min\{c_0 b, c_0 b + c_1 - c_2 f_s(h_i)\}$ and hence $\overline{u}^1(b, h_i)$ is nondecreasing in b for each fixed h_i, $i = 1, \cdots, M$.

Now assume that $\overline{u}^k(b, h_i)$ is nondecreasing in b for some $k \geq 1$ and for each fixed h_i, we have

$$\overline{u}^k(b+1, h_i) \geq \overline{u}^k(b, h_i) \quad \text{for all} \quad b \geq 1. \tag{4.16}$$

We show that it also holds for $k + 1$. Again, we consider the following cases.

Case (1): The optimal action in state $(b + 1, h_i)$ is not to transmit. Then from (4.9), the optimal action in state (b, h_i) is also not to transmit. In this case, we have

$$\begin{aligned} & \overline{u}^{k+1}(b+1, h_i) - \overline{u}^{k+1}(b, h_i) \\ = \; & X^k(b+1) - X^k(b) \\ = \; & c_0 + \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \left[\overline{u}^k(b+1, h_i) - \overline{u}^k(b, h_i)\right] \\ \geq \; & 0 \end{aligned} \tag{4.17}$$

The inequality is from the nonnegative c_0 and the induction assumption (4.16)

Case (2): The optimal action in state $(b + 1, h_i)$ is to transmit and the optimal action in state (b, h_i) is not to transmit. That the optimal action in state (b, h_i) is not to transmit implies

$$\begin{aligned} & \Delta^k(b, h_i) = c_1 - c_2 f_s(h_i) - f_s(h_i) U^k(b) \geq 0 \\ \Rightarrow \; & c_1 - c_2 f_s(h_i) \geq f_s(h_i) \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \left[\overline{u}^k(b+z, h_j) - \overline{u}^k(b+z-1, h_j)\right] \\ \Rightarrow \; & c_1 - c_2 f_s(h_i) \geq 0 \end{aligned} \tag{4.18}$$

The last inequality follows from the induction assumption (4.16). In this case, we have

$$\begin{aligned}
& \overline{u}^{k+1}(b+1,h_i) - \overline{u}^{k+1}(b,h_i) \\
= \; & X^k(b+1) + \Delta^k(b+1,h_i) - X^k(b) \\
= \; & c_0 + c_1 - c_2 f_s(h_i) + (1-f_s(h_i)) \sum_{z=0} \sum_{j=1}^{M} q(z) h_{ij} \left[\overline{u}^k(b+z+1,h_j) - \overline{u}^k(b+z,h_j)\right] \\
\geq \; & 0
\end{aligned} \tag{4.19}$$

The inequality follows from the nonnegative c_0, (4.18), $1-f_s(h_i) \geq 0$ and the induction assumption (4.16).

Case (3): The optimal actions in the states $(b+1,h_i)$ and (b,h_i) are to transmit. In this case, we have

$$\begin{aligned}
& \overline{u}^{k+1}(b+1,h_i) - \overline{u}^{k+1}(b,h_i) \\
= \; & X^k(b+1) + \Delta^k(b+1,h_i) - X^k(b) - \Delta^k(b,h_i) \\
= \; & c_0 + (1-f_s(h_i)) \sum_{z=0} \sum_{j=1}^{M} \left[\overline{u}^(b+z+1,h_j) - \overline{u}^(b+z,h_j)\right] \\
& \quad + f_s(h_i) \sum_{z=0} \sum_{j=1}^{M} \left[\overline{u}^(b+z,h_j) - \overline{u}^(b+z-1,h_j)\right] \\
\geq \; & 0
\end{aligned} \tag{4.20}$$

Again, the inequality follows from the nonnegative c_0 and the induction assumption (4.16).

From (4.17), (4.19), (4.20) and from the induction argument, we conclude that $\overline{u}^{k+1}(b,h_i)$ is a nondecreasing function of b for all $k \geq 0$, and from the MDP result of value iteration algorithm, $\overline{u}(b,h_i) = \lim_{k\to\infty} \overline{u}^k(b,h_i)$ is a nondecreasing function of b for each fixed h_i, $i = 1,\cdots,M$. This also completes the proof. ■

Note that $f_s(h_i)$ is considered as a nondecreasing function of the channel state h_i (cf. Section 3.1 (3.7)). From Lemma 4.2 and the nondecreasing $f_s(h_i)$, we then have the following corollary and proposition.

Corollary 4.2 *Suppose that $f_s(h_i)$ is a nondecreasing function of the channel state h_i, then $\Delta(b,h_i)$ is a nonincreasing function of h_i, $i = 1,\cdots,M$ for each fixed $b > 1$.*

Proof: From Lemma 4.2, $\overline{u}(b+z,h_j) - \overline{u}(b+z-1,h_j) \geq 0$. The rest of the proof is clear from the nonnegative increasing function $f_s(h_i)$. ■

Proposition 4.3 *If there exists a channel state threshold h_z such that the optimal action is to transmit in the state (b, h_z) given some buffer occupancy $b > 0$, then for all states $(b, h_i), i \geq z$, the optimal action is to transmit too.*

The proof of Proposition 4.3 is rather straightforward and we omit it here. Proposition 4.3 is intuitively clear and it suggests that if the sender transmits in a channel state, it should also transmit in all channel states better than this one.

Besides the value iteration algorithm, the policy iteration algorithm [7, 62] can also be used to compute an optimal policy. We propose the following modified unichain policy iteration algorithm for our problem which simplifies the computation of optimal policies. Puterman (see [62], page 386) has proposed a general modified policy iteration algorithm. Our algorithm is based on his algorithm, however, the property of optimal policies has been exploited in the policy improvement. Our algorithm, *the modified unichain policy iteration algorithm*, is described as follows.

The modified unichain policy iteration algorithm

1. Set $k = 0$, $\overline{u}^0(b, h_i) = 0$ for all (b, h_i). Specify $\epsilon > 0$ and an integer $N \geq 1$.

2. (Policy Improvement)

   2a. Set $i = 1$, $b = 1$; Goto (2b).

   2b. Compute $\Delta^k(b, h_i)$. If $\Delta^k(b, h_i) \geq 0$, goto (2c); otherwise goto (2d).

   2c. Set $d^k(b, i) = 0$ and $b = b + 1$. If $b \leq B$, goto (2b); otherwise goto (2e).

   2d. Set $d^k(b', i) = 1$ for all $b' \geq b$. Goto (2e)

   2e. Set $b = 1$ and $i = i + 1$. If $i \leq M$ goto (2b), otherwise goto (3).

3. (Partial Policy Evaluation)

   3a. Set $n = 1$ and for all $\mathbf{s}$ compute

$$v^n(\mathbf{s}) = C(\mathbf{s}, d^k(\mathbf{s})) + \sum_{\mathbf{s}'} \text{Tr}(\mathbf{s}'|\mathbf{s}, d^k(\mathbf{s}))\overline{u}^k(\mathbf{s}). \tag{4.21}$$

   3b. If

$$\|\overline{u}^k - v^n\| \leq \epsilon \tag{4.22}$$

goto (4). Otherwise goto (3c).

3c. If $n = N$ goto (3d); otherwise compute v^{n+1} by

$$v^{n+1}(\mathbf{s}) = C(\mathbf{s}, d^k(\mathbf{s})) + \sum_{\mathbf{s}'} \mathrm{Tr}(\mathbf{s}'|\mathbf{s}, d^k(\mathbf{s}))v^n(\mathbf{s}). \tag{4.23}$$

and set $n = n + 1$, goto (3c).

3d. Set $k = k + 1$ and $\overline{u}^k = v^n$ and goto (2).

4. (Policy Identification)

4a. Set $i = 1$, $b = 1$; goto (4b).

4b. Compute $\Delta^k(b, i)$. If $\Delta^k(b, i) \geq 0$, goto (4c); otherwise goto (4d).

4c. Set $d_\epsilon(b, i) = 0$ and $b = b + 1$. If $b \leq B$, goto (4b); otherwise goto(4e);

4d. Set $d_\epsilon(b', i) = 1$ for all $b' \geq b$; goto (4e).

4e. Set $b = 1$ and $i = i + 1$. If $i \leq M$, goto (4b); otherwise goto (5).

5. Print the $\epsilon-$optimal policy and stop.

This algorithm combines features of both policy iteration, e.g., step 2 and value iteration, e.g., step 3. Furthermore, in step 2 and step 4, the property of optimal policies, i.e., Proposition 4.2, has been exploited to reduce the computation of optimal policies. We next provide some numerical examples.

4.4 Numerical Examples

The channel model used in this section is based on the Gudmundson [33] model for a mobile with a constant velocity. We classify the channel states in terms of channel gain, i.e., the received SNR. We assume that fast fading is averaged over a frame via perfect interleaving and only shadowing is considered in the simulations. As suggested by Gudmundson [33], we model the log-normal shadowing as a Gaussian process (in dB units). The channel gain (SNR) is modelled as a log-normal random process with mean 7 dB and autocorrelation function $R(k) = \sigma^2(0.3)^{\alpha|t|}$, where $\sigma^2 = 4.3$ dB and α is proportional to the velocity of a user. We split the range of channel gains into a finite number of intervals (states) and make each state equiprobable. The transition probabilities are determined by suitably integrating over the conditional probability density function. For example, we consider to construct a 4-state Markov channel. The quantization intervals are given as: $(0, 5.6013], (5.6013, 7], (7, 8.3985], (8, 3985, \infty)$. The transition probability is given by

$$\Pr(h(t+1) = h_j | h(t) = h_i) = \frac{\Pr(h(t+1) = h_j, h(t) = h_i)}{\Pr(h(t) = h_i)} \tag{4.24}$$

where the joint cumulative distribution function of $(h(t+1), h(t))$ is a bivariate normal distribution with the correlation matrix $\begin{pmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{pmatrix}$ and $\rho_{12} = \rho_{21} = 0.3^{\alpha}$. We develop a routine to compute the probability of the bivariate normal probabilities, which is based on the method introduced by A. Genz [27]. For example, the channel transition matrix at $\alpha = 0.6$ is

$$\mathbf{H} = \begin{bmatrix} 0.4729 & 0.2784 & 0.1719 & 0.0768 \\ 0.2784 & 0.2930 & 0.2568 & 0.1718 \\ 0.1719 & 0.2568 & 0.2928 & 0.2785 \\ 0.0768 & 0.1718 & 0.2785 & 0.4729 \end{bmatrix} \tag{4.25}$$

For simplicity, we assume an uncoded system using BPSK modulation with 512-bit packets. The frame success probabilities $f_s(h_i)$ are simply taken as the middle values of each SNR interval. For example, the frame success probabilities are $f_s(h_1) = 0.4, f_s(h_2) = 0.88, f_s(h_3) = 0.95, f_s(h_4) = 1.00$ for a 4-state Markov channel. We only consider a simple Bernoulli arrival process. The arrival probability is set as $q(0) = 0.5$ and $q(1) = 0.5$. The buffer limit is set as 80 in all simulations.

We now provide some numerical examples to illustrate the performance of the optimal policies. The purpose of the simulations is two folds. We first evaluate how the different cost weights affect the resulting optimal policies and then compare the performance of an optimal policy with that of a persistent transmission policy. We then evaluate how the channel characteristics affect the resulting optimal policies. For the first purpose, we consider a 4-state Markov channel. We fix $c_1 = 100.0$ and $c_2 = 200$ and use different values of c_0 in the simulations. Recall that the value of c_0 influences the delay costs. Table 4.1 shows the computed optimal policies (i.e., the buffer occupancy that corresponds to the optimal action being transmission) with different values of c_0. In Table 4.1, there are four channel states with 1 representing the worst and 4 the best

Table 4.1: Optimal Transmission Policies ($B = 80$)

	Channel state			
c_0	1	2	3	4
10	[4, 80]	[1, 80]	[1, 80]	[1, 80]
5	[7, 80]	[1, 80]	[1, 80]	[1, 80]
1	[31,80]	[2, 80]	[1, 80]	[1, 80]
0.5	[60,80]	[3, 80]	[1, 80]	[1, 80]
0.1	-	[7,80]	[1, 80]	[1, 80]

channel. From Proposition 4.2, if the optimal action is to transmit in a state (b, h_i), then for all states $(b' > b, h_i)$ the optimal actions are to transmit too. For example, with $c_0 = 5$ and the channel state being 1, the optimal action is to transmit whenever there are 7 or more than 7 packets in the buffer. There may exist multiple thresholds each corresponding to a particular channel state, for example, when $c_0 = 0.5$. We see from Table 4.1 that if the optimal action is to transmit in a state (b, h_i), then for all states $(b, h_j), j \geq i$ the optimal actions are to transmit too (cf. Proposition 4.3). Also as shown in Table 4.1, the optimal policies based on our average cost modelling have a similar structure as the "back-off" optimal policies in [98], i.e., the controller tends to suspend a transmission when the channel state is in the worst state. However as the arrival process is included in our model, in some bad channel states the controller still has to transmit to avoid increasing the holding cost whenever the buffer occupancy exceeds some thresholds.

We next compare an optimal transmission policy with a *persistent transmission* policy. In the persistent transmission policy, the action is to transmit whenever the buffer is not empty. In the simulations, immediate costs are computed and collected in each frame based on the system state at the beginning of the frame and the action prescribed by the optimal policy and the persistent transmission policy. In the simulations, we also count the buffer state and record the delay of a successfully transmitted packet in each frame. The simulation results including average costs, average delay and the probability of buffer occupancy are then computed by averaging over 10 runs of simulations each with 100000 frames. Table 4.2 shows the average costs per unit time (per frame) of the optimal policy and the persistent transmission policy. Note that

Table 4.2: Average Costs Comparison

c_0	10	5	1	0.5	0.1
Optimal (computed)	-32.6504	-39.4137	-45.4993	-46.5073	-47.6982
Optimal (simulated)	-32.7216	-39.3334	-45.5844	-46.5301	-47.7238
Persistent(simulated)	-30.0616	-32.9291	-36.2007	-36.5783	-37.0761

the negative cost can be considered as the positive reward from the transmission. The simulated average cost of the optimal policy (given a fixed c_0) is close to the computed one, and is smaller than that of the persistent transmission policy. When the buffer occupancy is less than some threshold, the sender can choose not to transmit if the current channel state is poor under the optimal policy. In such a situation, suspension is more profitable than transmission and the sender can transmit either when the channel transits to a better state or when the buffer occupancy exceeds a threshold. However, this may be at the expense of increased delay. Table 4.3 compares the goodput, the average buffer occupancy and the average delay of the optimal policy and the persistent transmission policy. The goodput is defined as the ratio of error-free transmissions over all transmissions. Note that the goodput defined here also can be used as a measure of energy (or transmission) efficiency since the transmission power is fixed

Table 4.3: Goodput, Occupancy and Delay Comparison

	Goodput		Occupancy		Delay	
c_0	opti.	pers.	opti.	pers.	opti.	pers.
10	0.9195	0.7977	1.2947	0.8733	3.0894	2.2466
5	0.9372	0.7972	1.4627	0.8739	3.4254	2.2478
1	0.9507	0.7978	1.9332	0.8742	4.3664	2.2484
0.5	0.9554	0.7973	2.3645	0.8740	5.2290	2.2480
0.1	0.9638	0.7978	4.3487	0.8736	9.1974	2.2472

in our problem. The average buffer occupancy is the average number of packets in the buffer. It is computed as $\overline{N} = \sum_{j=0}^{B} jp^{\pi}(j)$, where $p^{\pi}(j)$ is the stationary probability of j packets in the buffer when the policy π is applied. Since stationary optimal policies exist, the stationary probability distribution of buffer occupancy exists. Note that in our simulations, overflow does not happen as the arrival rate ($\lambda = \sum_j jq(j) = 0.5$) is small and the buffer limit is large. Hence by Little's theorem, the average delay of a packet can also be computed as $\overline{T} = \frac{1}{2} + \overline{N}/\lambda$ (unit in frame duration). It is seen from Table 4.3 that the optimal policy has a larger goodput compared with the persistent transmission policy. This is because the optimal policy can exploit the time-varying channel and the delay tolerance of data users. It is also seen from Table 4.3 that the larger the goodput of the optimal policy, the larger the average delay. However different design objectives (e.g., goodput, delay) can be easily met by adjusting the values of the different costs parameters.

We next investigate how the channel characteristic affects the resulting optimal policy and its performance. The channel dynamics can be totally represented via the channel transition matrix $\mathbf{H}$. The frequency at which the channel changes state depends on the values of the diagonal elements of $\mathbf{H}$. As the values of the diagonal elements increase, the less frequent the channel changes state, and the stronger the dependence structure, or the channel memory, becomes. We hence use the following definition for the channel memory which has also been used by D. Zhang and K. Wasserman in [98] for a Markov channel.

Definition 4.1 *The channel memory ξ of $\mathcal{H}$ is defined to be the second dominant eigenvalue of the channel transition probability matrix* $\mathbf{H}$.

For example, the probability transition matrix of the two-state Markov channel (or the Gilbert-Elliott (GE) channel [28]) is

$$\mathbf{H} = \begin{bmatrix} 1-g & g \\ b & 1-b \end{bmatrix} \tag{4.26}$$

and the second eigenvalue of $\mathbf{H}$ is $\xi = 1 - g - b$, which has been used as the channel memory in [53]. The case of $\xi = 0$ corresponds to a memoryless channel, and $\xi = 1$ to a decomposable Markov chain, where the channel remains in its initial state for all time (infinite memory). Furthermore, when $\mathbf{H}$ is *monotone* [2] and irreducible, it has, apart from its simple eigenvalue at 1, a second non-negative eigenvalue at $\xi < 1$ such that all other eigenvalues have modulus not exceeding ξ (see [42] for proof). As the Markov channels used in our problem are monotone, we investigate the resulting optimal policies for different channels in terms of the channel states and the channel memory. We study 2-, 4- and 8-state Markov channels of different channel memories. Furthermore, in all the following numerical examples we fix the arrival probabilities as $q(0) = 0.5$ and $q(1) = 0.5$, and the weights as $c_0 = 1$, $c_1 = 100$ and $c_2 = 200$.

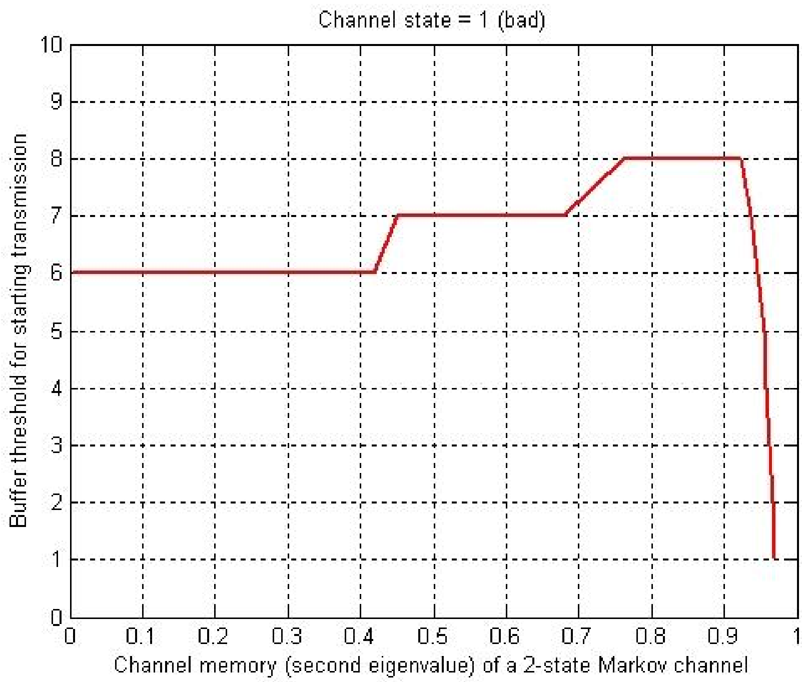

Figure 4.2: Buffer threshold for starting transmission in channel state 1 of a 2-state Markov channel as a function of channel memory.

Figs. 4.2, 4.3 and 4.4 plot the buffer thresholds (optimal policies) for starting transmission in different channel states as a function of channel memory for the 2-, 4- and 8-state Markov channels, respectively. For those channel states that are not plotted in the figures, the buffer thresholds for starting transmission are all from 1, i.e., when the buffer is not empty. Though the buffer thresholds are not monotone increasing or de-

[2] A stochastic matrix is said to be monotone if its probability of row vectors are stochastically nondecreasing. A brief introduction of monotone stochastic matrix is provided in Section 5.5.3.

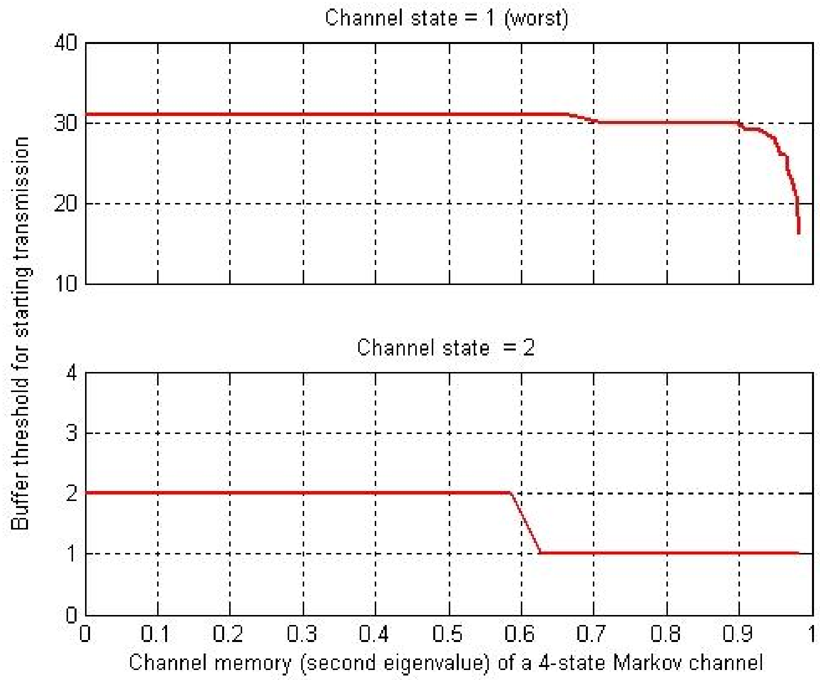

Figure 4.3: Buffer threshold for starting transmission in channel state 1 and 2 of a 4-state Markov channel as a function of channel memory.

creasing with the channel memory, e.g., Fig. 4.2, it is observed that the sender becomes aggressive when the channel memory is very high (e.g., $\xi \geq 0.9$) in that it has to send even when the buffer occupancy is low in the bad state(s). We explain it as follows. Note that an optimal policy is computed given a specific channel. When the channel memory increases, the sojourn time in the current channel state increases. However, the holding cost could increase regardless of the current channel state as new arrivals are independent of the channel and could happen in each frame. Therefore, to reduce the (possibly) increasing holding cost due to new arrivals, the sender cannot wait too long for the channel to transit from a bad state to a good state and has to transmit at the low buffer occupancy. We note that our results are different from the results reported by D. Zhang and K. Wasserman. In [98], D. Zhang and K. Wasserman considered the optimal transmission control problem for an always backlogged user whose buffer occupancy dynamics are not considered (i.e., the holding cost is not based on the buffer occupancy but a constant) and reported that the optimal backoff time increases as the channel memory increases. As the buffer dynamics are not considered and only

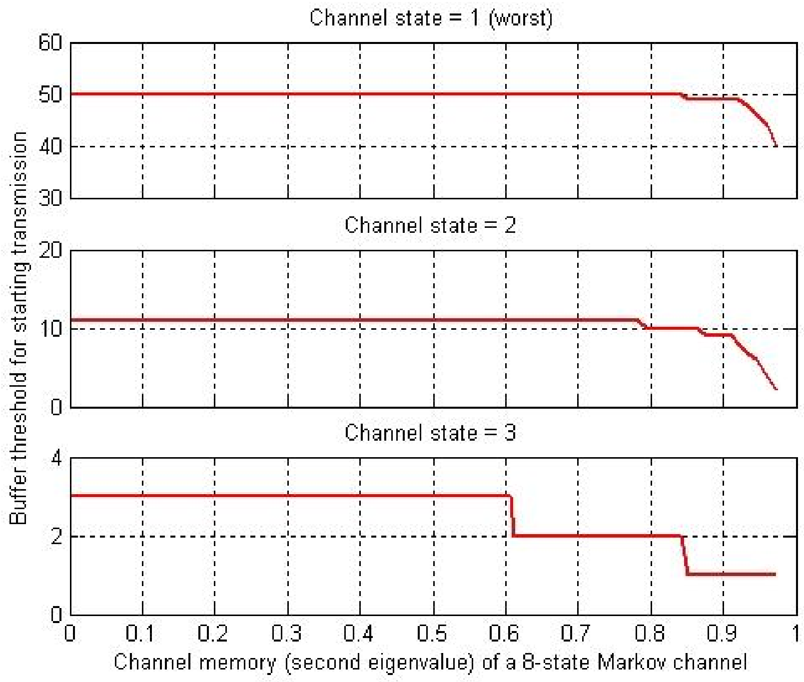

Figure 4.4: Buffer threshold for starting transmission in channel state 1, 2 and 3 of a 8-state Markov channel as a function of channel memory.

fixed penalty is charged for no transmission in [98], the sender needs not to balance the increased holding cost and hence the optimal backoff time (or in other words, the optimal waiting time for channel transiting from a bad state to a good state) increases with the channel memory.

Figs. 4.5, 4.6 and 4.7 plot the average cost as a function of channel memory for the optimal policies and a persistent transmission policy with 2-, 4- and 8-state channels, respectively. We first note that the optimal policies have lower costs than that of the persistent policy. We then note that the cost increases with the channel memory for both the optimal policies and the persistent policy. This is due to the fact that a channel with a larger memory is less frequent to transit to other states. It is less *opportunistic* for a sender to exploit the channel variation with the increase of channel memory. This leads to the increased transmission failures and the increased holding cost, and hence the total cost increases. Finally, we note that the cost with a channel with more states is better than with a channel with fewer states (e.g., the cost of 8-state channel is less than that of a 4-state channel), which is due to a finer quantization of

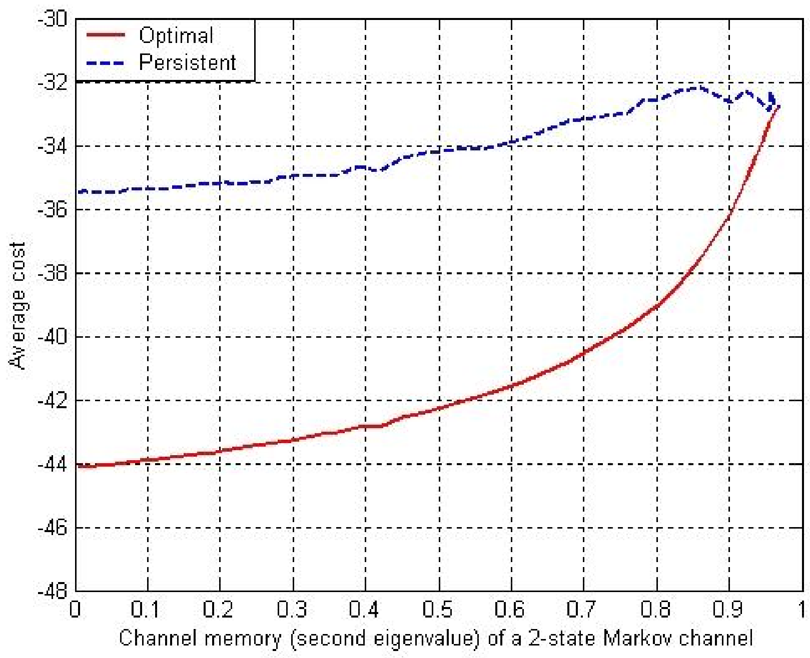

Figure 4.5: Cost for the 2-state Markov channel as a function of channel memory.

the channel.

Figs. 4.8, 4.9, and 4.10 (Figs. 4.11, 4.12 and 4.13) plot the goodput (the average buffer occupancy) as a function of channel memory for the optimal policies and the persistent policy with 2-, 4- and 8-state channels, respectively. The goodput of the optimal policy is higher than that of the persistent policy as the sender suspends transmission in bad states. However, the higher goodput is at the cost of increased buffer occupancy. Again, the fact that the sender stays more time in the current state in channel with a small memory but has to transmit in bad state(s) explains the decrease of the goodput for both the optimal and the persistent policies with the channel memory. The drop in the occupancy in Fig. 4.11 when the memory $\xi > 0.9$ is due to the fact that the corresponding optimal policy has to transmit at very low buffer occupancy, cf. Fig 4.2. The drop in the occupancy in Fig. 4.12 when the memory ξ is above 0.9835 is due to the fact that overflow occurs (0.0014%).

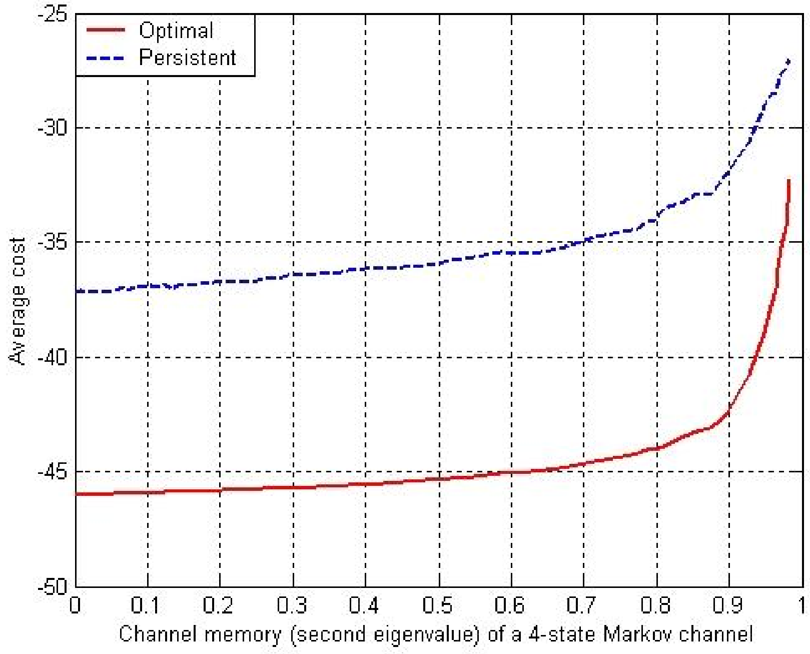

Figure 4.6: Cost for the 4-state Markov channel as a function of channel memory.

4.5 Summary

In this chapter, we have studied a simple transmission control problem for a single user with general arrival statistics. The sender decides, at the beginning of each frame, whether or not to transmit a packet based on the buffer occupancy and the channel state. The goal is to find a policy that optimally balances different costs such as the transmission power and the average delay. We formulate it as a Markov decision problem and show the existence of the stationary average cost optimal policies. We have also shown the properties of the optimal policies, i.e., the threshold-based structure, which helps to reduce computation effort. Numerical examples are then provided to illustrate how to achieve different balance points and to compare the performance of optimal policies with that of the persistent policy.

Besides power and transmission control, a mobile may control its transmission rate during the holding time of a connection. Rate control will be studied in the next chapter.

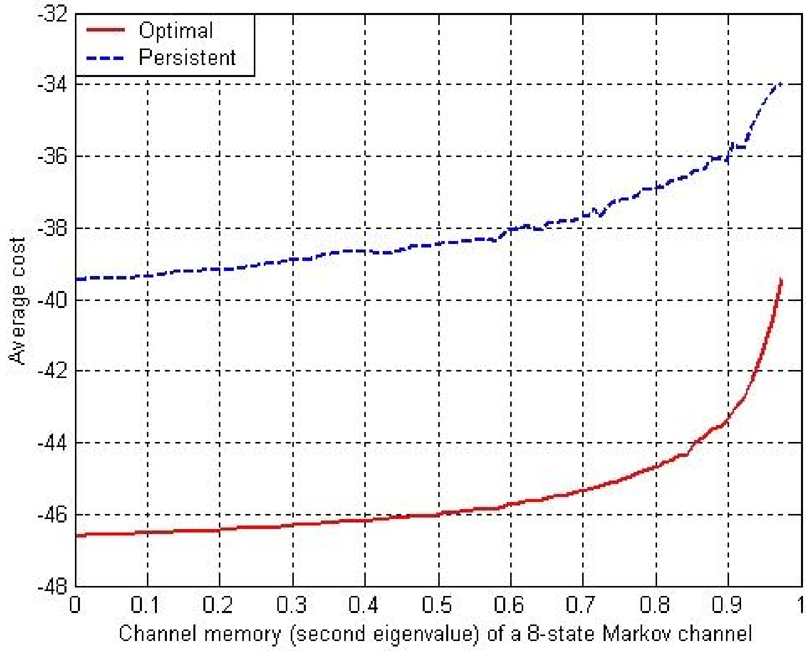

Figure 4.7: Cost for the 8-state Markov channel as a function of channel memory.

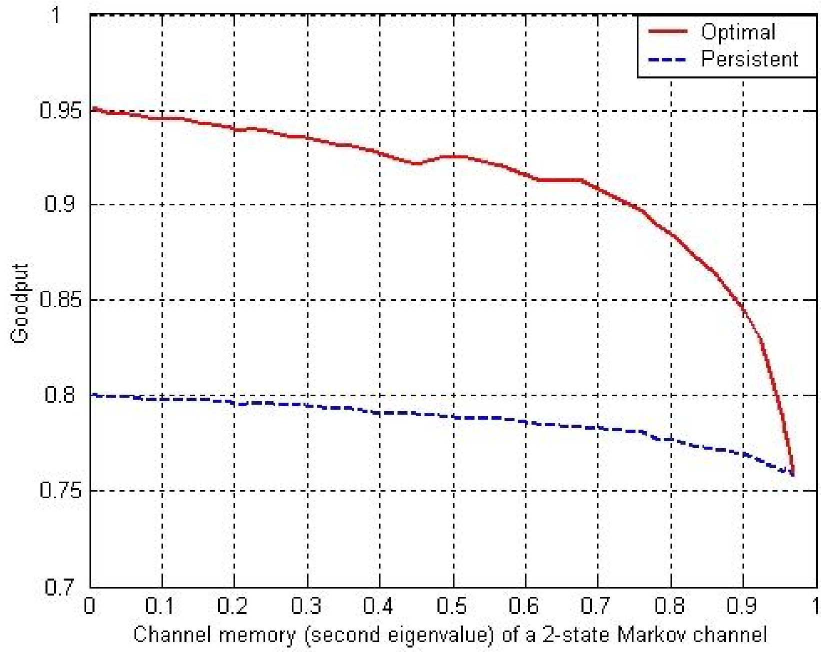

Figure 4.8: Goodput for the 2-state Markov channel as a function of channel memory.

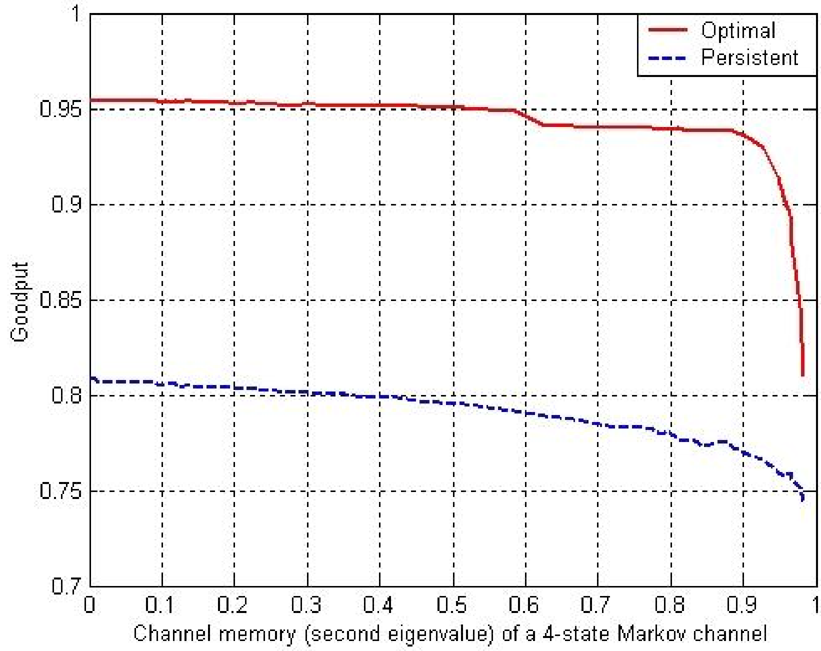

Figure 4.9: Goodput for the 4-state Markov channel as a function of channel memory.

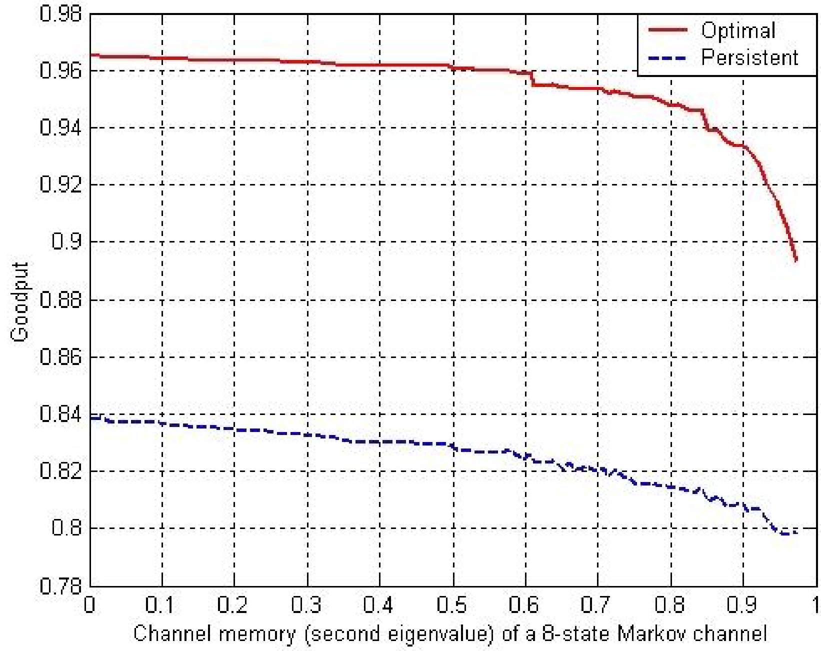

Figure 4.10: Goodput for the 8-state Markov channel as a function of channel memory.

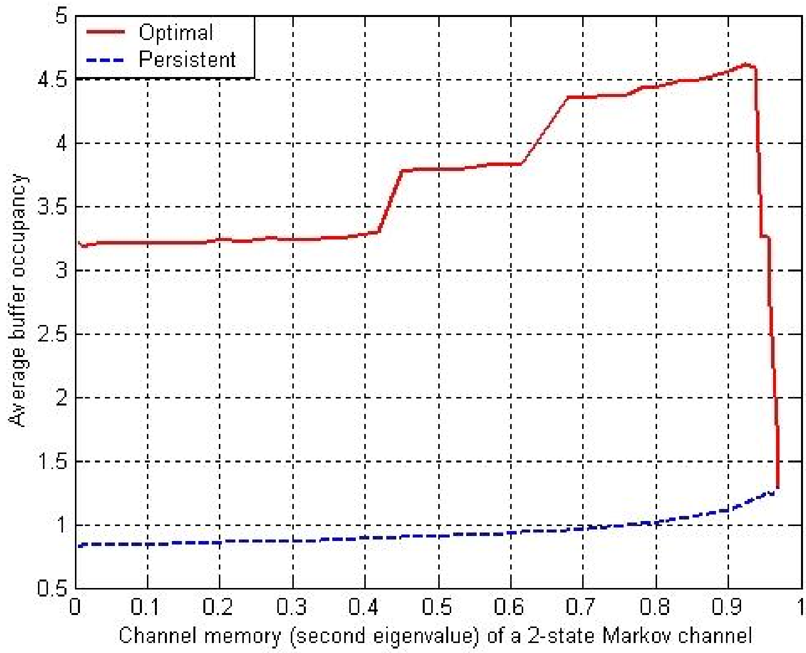

Figure 4.11: Average buffer occupancy for the 2-state Markov channel as a function of channel memory.

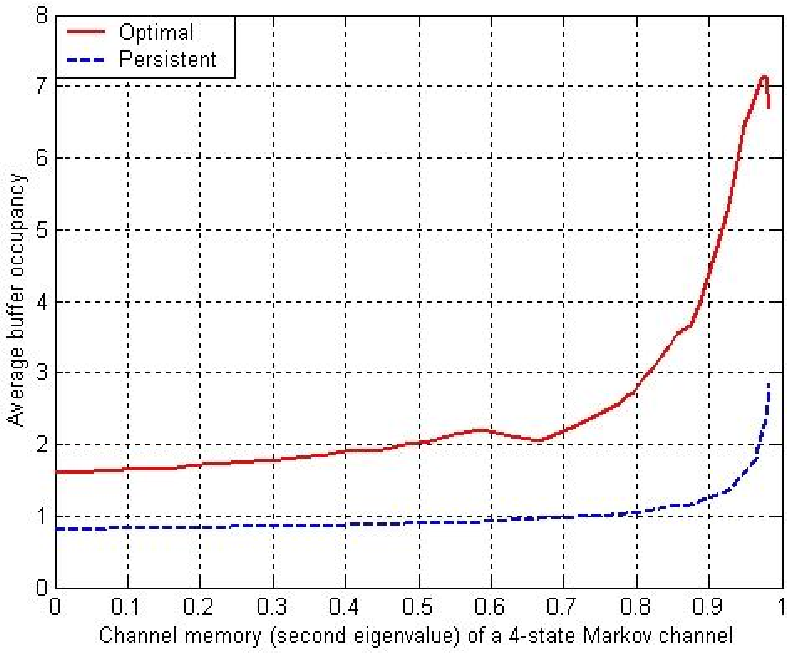

Figure 4.12: Average buffer occupancy for the 2-state Markov channel as a function of channel memory.

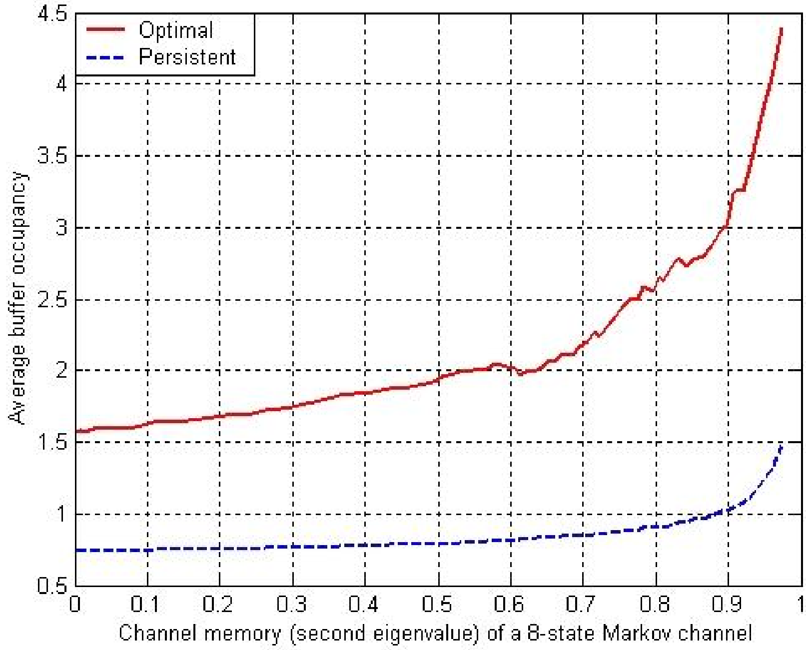

Figure 4.13: Average buffer occupancy for the 2-state Markov channel as a function of channel memory.

Chapter 5

Optimal Rate Allocation Policies

In this chapter, we consider a rate allocation problem for a single data user. A data user needs to pay for the usage of resources, e.g., it pays for its transmission rate. Thus a data user may request different transmission rates during its connection holding time to reduce its resource usage cost but still meeting its QoS requirements. We formulate it as a Markov decision problem. The characteristic and structure of optimal polices are discussed. We show that, based on some mild assumptions, the optimal policies are monotone. Furthermore, we propose a class of simple policies that are easy to implement to approximate the optimal policies. We analyze such simple policies and provide an upper delay bound. Finally, we extend the single user self-optimization problem to consider the situation where multiple users are present. We still use MDP to model and to solve the latter problem but with some proper assumptions. The characteristic of the value function and the property of the optimal policies for the extended problem are discussed.

5.1 Problem Formulation

In this section, we provide the problem formulation of a Markov decision modelled dynamic rate allocation problem for a single data user with general arrival statistics. The system model is shown in Fig. 5.1. We consider a discrete time system and the example of transmission model has been shown in Fig. 2.2. Note that the system model can also be that of a service rate controlled queueing system, i.e., a single server with batch arrivals, adjustable batch service capabilities and a finite buffer. We consider

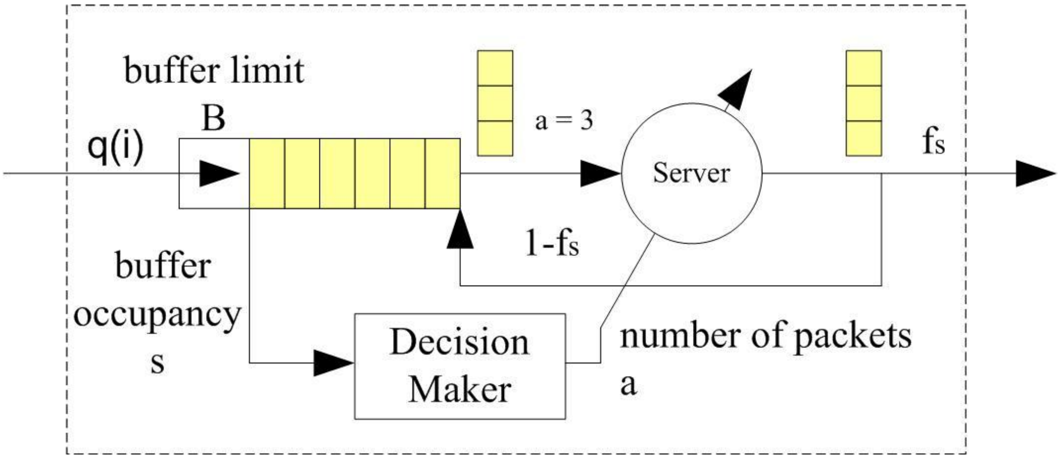

Figure 5.1: System model — A service rate controlled queueing system.

transmitting different number of packets in a frame to approximate the allocation of different rates. Note that variable rates can be implemented using the variable spreading factor operation and/or multi-code operation in a CDMA-based wireless network [1]. In each frame, there is a batch of data packets of the same size arriving at the sender's buffer. Arriving packets are queued in a first-in-first-out buffer that can hold at most B packets. If the buffer is full, arriving packets are discarded, i.e., the buffer overflows. We assume that batch arrivals in different frames are independent and cannot be transmitted in the same frame. Let $q(i)$, $i = 0, 1, \cdots$, denote the probability of i packets arriving in a frame. Let λ denote the average number of arrivals in a frame; $\lambda = \sum_{i=0} iq(i)$. In this chapter, we assume that the underlying power control algorithm is ideal and can achieve the same average frame success probability f_s, $0 < f_s \leq 1$, even if different number of packets are transmitted in a frame. In general, the target SIR determines the frame success probability if perfect power control is assumed, and a larger target SIR is required for achieving the same frame success probability when more packets are sent in a frame. This will be discussed further later. We note that

[1]We note that other methods such as timeslot aggregation in GPRS/EDGE are also applicable for configuring different rates in a TDMA based network. A good overview on rate configuration and adaptation techniques in wireless packet data services can be found in [54].

power control algorithms such as up-down power control used in practical systems[2] have been shown to be able to achieve the SIR target [36, 75]. Thus f_s is considered as a constant in this chapter. At the beginning of each frame, the sender has to decide the number of packets to send in that frame based on the observation of the buffer occupancy. The objective of the sender is to find a stationary discount (average) optimal policy. We now provide a brief Markov decision problem formulation.

The decision epochs correspond to the beginning of each frame. The states are the number of packets queued in the buffer, denoted as s and $0 \leq s \leq B$, and the state space is then $\mathcal{S} = \{0, 1, \cdots, B\}$. The actions are the number of packets to transmit in a frame. In state 0, i.e., $s = 0$, there is no action as there is no packet available to transmit. In state $s \geq 1$, the available actions are from $\mathcal{A}_s = \{1, \cdots, \min\{A, s\}\}$, where A is the largest number of packets that can be transmitted in a frame. From the consideration of a stable queueing system, we require (at least) and hence assume $\lambda < Af_s$. Otherwise, buffer overflow is inevitable. The action space is then $\mathcal{A} = \bigcup_s \mathcal{A}_s$. The above definition of action space requires that at least one packet should be sent in a frame whenever the buffer is not empty. This can correspond to the minimum data rate requirement of a connection. However, this requirement can be relaxed by specifying a probability distribution on the set of actions, i.e., given that the buffer is not empty, there is a probability to choose a number of packets to send. According to our system model, the transition probabilities are then given as:

$$\mathrm{Tr}(s'|s,a) = \begin{cases} q(s') & s = 0 \\ (1-f_s)q(s'-s) + f_s q(s'-s+a) & s \geq 1, \quad a \in \mathcal{A}_s \\ 0 & otherwise \end{cases} \tag{5.1}$$

where $q(i < 0) = 0$. Note that we assume $s' \leq B$. If $s' > B$, we can redistribute the excess probability to state $s = B$ by using the augmentation procedure introduced in [73]. However, as shown later in the next section, selected properties of optimal policies simplify the computation of boundary probabilities and optimal policies. We consider a linear cost structure which consists of two cost functions: the usage cost function C_u and the holding cost function C_h. The usage cost may represent the

[2]For example, in UMTS the fast closed loop power control is designed to operate at a frequency of 1500 Hz [40]

charges that a user pays for its transmission rate and hence it is a function only on the action space. Without loss of generality, we assume that the usage cost function $C_u(a)$ is a nonnegative and nondecreasing function of a. The holding cost may represent the delay of packets and/or the energy that the sender uses to hold the packets and hence it is a function only on the state space. Without loss of generality, we assume that the holding cost function $C_h(s)$ is a nonnegative and nondecreasing function of s. Furthermore, we assume that C_u and C_h are bounded when the state space and the action space are finite. Hence the cost structure in our dynamic rate allocation problem is given as

$$C(s,a) = C_h(s) + C_u(a), \quad s \in \mathcal{S}, \;\; a \in \mathcal{A}_s \tag{5.2}$$

The objective is to find stationary optimal policies. Since the definition of policy values and the definition of optimal policies are the same as those in Section 2.2.2, we omit them here.

The optimality criterion can be either the discount optimal or the average optimal criterion. We prefer to use the average optimal criterion, since it implies that the cost is not sensitive to when the cost is incurred and also its policy value is independent of the starting state. However, as we assume a very general arrival probability distribution, it may not be easy to identify the unichain property for our Markov decision problem. On the other hand, it is easy to verify that all conditions in Theorem 2.2 are satisfied for this dynamic rate allocation problem. Thus a stationary discount optimal policy exists and the computation method introduced in Section 2.2.4 can be used to compute a stationary discount optimal policy for our rate allocation problem. However, based on a mild assumption, we find that optimal policies have a monotone property that can be used to simplify the computations. The next section discusses this monotone property of optimal policies. It is based on the discount optimality criterion, however, it also applies to the average optimality criterion once we identify the unichain property. A case study that deals with average optimal policies will be provided later.

5.2 Monotone Optimal Policies

A policy is *monotone* if its decision rules are monotone (nondecreasing/nonincreasing) in the state space. In our dynamic rate allocation problem, this means that the decision rules require that a larger number of packets are sent in a frame (i.e., using a higher transmission rate) when the number of packets in the buffer increases. If an optimal policy is monotone, we can simplify its computation and implementation. For example, when the largest action has to be used for a state s', then it should be used for all state $s > s'$ assuming that an optimal policy is monotone. An example is of the following *multiple-threshold-based policy*.

$$d(s) = \begin{cases} a_{min} & \text{if } \ s < z_1 \\ a_{avg} & \text{if } \ z_1 \leq s < z_2 \\ a_{max} & \text{if } \ s \geq z_2 \end{cases}$$

where z_1 and z_2, $z_1 < z_2$, are some buffer occupancy thresholds and a_{min}, a_{avg} and a_{max}, $a_{min} < a_{avg} < a_{max}$, are distinct transmission rates. A more simplified threshold-based policy will be analyzed in detail later. An intuitive explanation here is that the sender needs to transmit more packets in a frame (i.e., using a higher transmission rate) to reduce delay when its buffer occupancy increases. On the other hand, it needs to transmit fewer packets (i.e., using a smaller transmission rate) to reduce resource usage costs when its buffer occupancy decreases. If we can establish that optimal policies are monotone, then such a multi-threshold policy is also optimal and the problem reduces to that of determining the proper thresholds. We note that a multi-threshold like radio bearer allocation policy has also been suggested for UMTS in [22].

Many researchers including Puterman [62], Bertsekas [7], Serfozo & Lu [48], and Stidham & Weber [78, 79] have proposed and proved necessary conditions for the existence of a monotone optimal policy for Markov decision problems, though their problem contexts are different. Puterman proposed a set of general conditions for the existence of monotone optimal policies. The main idea of Puterman's monotone conditions (e.g., see Theorem 4.7.4 and Theorem 4.7.5 in [62]) is to establish the *superadditive* (or *subadditive*) property of the value function over the state space and the action space. Both Bertsekas and Serfozo & Lu provide a proof for monotone optimal service rate control policies in a continuous time M/M/1 queueing system. Bertsekas' proof is based on showing the convexity of the value function while Serfozo & Lu show the superadditive property of the value function. Stidham & Weber provide the monotone conditions for general cases in which the induced Markov chain is *left-skip-free* in a Markov decision problem.

We note that the context of our dynamic rate allocation problem defined in the previous section is more general than those above. However, based on an additional but mild assumption, we prove that the optimal policies for our dynamic rate allocation problem also have a monotone property. Our proof follows a similar line with that of Bertsekas, i.e., to prove the convexity of the value function, but our problem takes a more general form. As the state space and the action space are discrete in our problem, we use the following definition for a convex function on the set of nonnegative integers. A function $g(x) : \mathbb{N}^+ \mapsto \mathbb{R}$ is defined to be convex if for all $x = 1, 2, \cdots$,

$$g(x+1) + g(x-1) \geq 2g(x) \tag{5.3}$$

The following lemma provides an alternative characterization of the convexity restricted to nonnegative integers. We denote $\lceil x \rceil$ to be the smallest integer greater than x and $\lfloor x \rfloor$ to be the largest integer smaller than x.

Lemma 5.1 *Let $g : \mathbb{N}^+ \to \mathbb{R}$ be a function defined on $\{0, 1, \cdots, \}$, then the following claims are equivalent:*
(i) *g is convex.*
(ii) $g(x_1) + g(x_2) \geq g(\lceil \frac{x_1+x_2}{2} \rceil) + g(\lfloor \frac{x_1+x_2}{2} \rfloor)$ *for all $x_1, x_2 \in \mathbb{N}^+$.*

Proof: Note that (ii) implying (i) is straightforward. By letting $x_1 = x + 1$ and $x_2 = x - 1$, (ii) implies (5.3) and hence g is convex. We now show that (i) implies (ii). That g is convex also implies $g(x+1) - g(x)$ is non-decreasing in $x \in \mathbb{N}^+$ and hence $g(x+1+k) - g(x+k) \geq g(x) - g(x-1)$ for all $k \geq 0$. This follows by iterating (5.3) with k steps, i.e., iterating $g(x+1) - g(x) \geq g(x) - g(x-1)$ and summing up the k inequalities.

We show that (i) implies (ii) by an induction argument. Without loss of generality, we assume that $x_1 \geq x_2$. First note that (ii) holds with equality for all $x_1, x_2 \in \mathbb{N}^+$

with $x_1 = x_2$ and $x_1 = x_2 + 1$. It is also easy to verify that (i) implies (ii) for all $x_1, x_2 \in \mathbb{N}^+$ with $x_1 = x_2 + 2$. Now assume that (i) implies (ii) for all $x_1, x_2 \in \mathbb{N}^+$ with $x_1 = x_2 + k$ for some $k \geq 3$. We then show by induction that (ii) must be true for any x_1, x_2 with $x_1 = x_2 + k + 1$. As g is convex and $g(x+1) - g(x)$ is nondecreasing in x, we have

$$g(x_1) - g(x_1 - 1) = g(x_2 + k + 1) - g(x_2 + k) \geq g(x_2 + 1) - g(x_2)$$

for all $k \geq 1$. Let $x'_1 = x_1 - 1$ and $x'_2 = x_2 + 1$. Note that $x'_1 = x'_2 + k - 1$. Thus by induction, we have

$$g(x_1 - 1) + g(x_2 + 1) \geq g(\lceil \frac{x_1 + x_2}{2} \rceil) + g(\lfloor \frac{x_1 + x_2}{2} \rfloor)$$

Combining the above two inequalities, we have

$$g(x_1) + g(x_2) \geq g(\lceil \frac{x_1 + x_2}{2} \rceil) + g(\lfloor \frac{x_1 + x_2}{2} \rfloor)$$

We then have that (ii) also holds for all x_1, x_2 with $x_1 = x_2 + k + 1$ and hence the lemma follows from the induction. ■

Our monotone proof is based on showing that the value function is convex, which is accomplished by the following assumption and lemma. We use the discount optimal value function in our proof. However, we note that similar arguments also apply to the average optimal relative value function once we identify the unichain property. Furthermore, we assume that the buffer is infinite to eliminate any buffer boundary effect, which is equivalent to assuming that no buffer overflow occurs.

Assumption 5.1 *The usage cost function $C_u(a)$ is a convex function on the action space and the holding cost function $C_h(s)$ is a convex function on the state space.*

Lemma 5.2 *Assumption 5.1 holds. Assume that the buffer is infinite, i.e., $s \in \{0, 1, \cdots\}$, the discount optimal value function $u_\rho(s)$ is a convex function of s.*

Proof: The proof is based on the value iteration algorithm (see Section 2.2.4) and proceeds by induction arguments. For $k = 0$, we set $u_\rho(s) = 0$ for all $s \in \mathcal{S}$.

For $k = 1$, we have

$$u_\rho^1(s) = \min_{a \in \mathcal{A}_s} \{C_h(s) + C_u(a)\}$$

and hence $u^1_\rho(s)$ is convex as $C_h(s)$ is assumed as a convex function of s.

Now assume that $u^k_\rho(s)$ is convex for some $k \geq 1$. We show that it also holds for $k+1$. By substituting the general transition probabilities representation of $\mathrm{Tr}(s'|s,a)$ in (2.14) with the one defined in (5.1), the $(k+1)$th step of the value function can be written as

$$\begin{aligned} & u^{k+1}_\rho(s) \\ = \ & \min_{a\in\mathcal{A}_s}\left\{C_h(s)+C_u(a)+\rho(1-f_s)\sum_{i=0}q(i)u^k_\rho(s+i)+\rho f_s\sum_{i=0}q(i)u^k_\rho(s+i-a)\right\} \end{aligned} \tag{5.4}$$

For any $s \geq 2$, let a^*_{s+1} and a^*_{s-1} be the optimal actions that realize the minimum part of (5.4) for $u^{k+1}_\rho(s+1)$ and $u^{k+1}_\rho(s-1)$, respectively. Then we have

$$\begin{aligned} & u^{k+1}_\rho(s+1)+u^{k+1}_\rho(s-1) \\ = \ & C_h(s+1)+C_h(s-1)+C_u(a^*_{s+1})+C_u(a^*_{s-1}) \\ & +\rho(1-f_s)\textstyle\sum_i q(i)u^k_\rho(s+i+1)+\rho f_s\sum_i q(i)u^k_\rho(s+i+1-a^*_{s+1}) \\ & +\rho(1-f_s)\textstyle\sum_i q(i)u^k_\rho(s+i-1)+\rho f_s\sum_i q(i)u^k_\rho(s+i-1-a^*_{s-1}) \\ \geq \ & 2C_h(s)+C_u(\lceil\tfrac{a^*_{s+1}+a^*_{s-1}}{2}\rceil)+C_u(\lfloor\tfrac{a^*_{s+1}+a^*_{s-1}}{2}\rfloor)+2\rho(1-f_s)\textstyle\sum_i q(i)u^k_\rho(s+i) \\ & +\rho f_s\textstyle\sum_i q(i)u^k_\rho(s+i-\lceil\tfrac{a^*_{s+1}+a^*_{s-1}}{2}\rceil)+\rho f_s\sum_i q(i)u^k_\rho(s+i-\lfloor\tfrac{a^*_{s+1}+a^*_{s-1}}{2}\rfloor) \\ \geq \ & 2u^{k+1}_\rho(s) \end{aligned}$$

The first inequality follows from the convexity of C_u, C_h and u^k_ρ (from the induction hypothesis) and the application of Lemma 5.1. The second inequality follows from the definition of u^{k+1}_ρ.

Thus we have the convexity of u^{k+1}_ρ for all $k \geq 0$ from the induction argument, and from the MDP result of value iteration algorithm (2.17), $u_\rho(s) = \lim_{k\to\infty} u^k_\rho(s)$ is convex in s. This also completes the proof. ■

Indeed, $u_\rho(s)$ is also a nonnegative and nondecreasing[3] function of s. Surprisingly though, the monotone proof does not need such a condition. The proof above in essence shows that the convexity of the cost structure can be *propagated* to the value function via the optimal equation. The following proposition summarizes our monotonicity result.

[3]Proposition 5.5 provides a proof of nondecreasing value function for an extended problem, where the same procedure is also applicable to prove the nondecreasing $u_\rho(s)$.

Proposition 5.1 *Assume that buffer overflow does not occur. Then the stationary discount optimal policy of our dynamic rate allocation problem is monotonically increasing in the state space. Mathematically, this means that*

$$d^*(s+1) \geq d^*(s), \quad s \geq 1 \tag{5.5}$$

Proof: The proof proceeds by contradiction.

For states s and $s+1$, $s \geq 1$, let a_s^* and a_{s+1}^* be the corresponding optimal actions, respectively. Assuming $a_s^* > a_{s+1}^*$, we show that this assumption contradicts with the convexity of $u_\rho(s)$. As the optimal actions realize the minimum part of the right hand side of optimal equation(2.14), via some simple algebraic calculations and for state s, we have

$$C_u(a_s^*) + \rho f_s \sum_i q(i) u_\rho(s+i-a_s^*) < C_u(a_{s+1}^*) + \rho f_s \sum_i q(i) u_\rho(s+i-a_{s+1}^*)$$

Note that the strict inequality holds since we always break ties by choosing the smallest action when choosing an optimal action. Similarly, for state $s+1$, we have

$$C_u(a_{s+1}^*) + \rho f_s \sum_i q(i) u_\rho(s+i+1-a_{s+1}^*) \leq C_u(a_s^*) + \rho f_s \sum_i q(i) u_\rho(s+i+1-a_s^*)$$

Combining the above two inequalities, we have

$$\begin{aligned} & \textstyle\sum_i q(i) \left[u_\rho(s+i+1-a_{s+1}^*) - u_\rho(s+i-a_{s+1}^*)\right] \\ < \; & \textstyle\sum_i q(i) \left[u_\rho(s+i+1-a_s^*) - u_\rho(s+i-a_s^*)\right] \end{aligned} \tag{5.6}$$

From Lemma 5.2, u_ρ is convex in s and hence $u_\rho(s+1) - u_\rho(s)$ is nondecreasing in s. Thus

$$\sum_i q(i) \left[u(s+i+1-a_s^*) - u(s+i-a_s^*)\right]$$

is also nondecreasing in s. Then from (5.6), it follows that $s+i-a_{s+1}^* \leq s+i-a_s^*$. But this contradicts with the assumption that $a_s^* > a_{s+1}^*$ and hence the assumption is not true. This also completes the proof. ∎

We note that similar techniques to prove Lemma 5.2 and Proposition 5.1 also apply to the average optimal decision problem whenever we identify the unichain property, or whenever we use the average cost optimal equation (2.12) to compute stationary average optimal policies. Indeed, we can prove that the relative value function $\overline{u}$ is convex by using the same argument as that in Lemma 5.2, provided that the ACOE holds. We now summarize the result in the following corollary but omit the proof here.

Corollary 5.1 *Assumption 5.1 holds. Further, assume that there exists stationary average optimal policies satisfying the ACOE (2.12). Then the average optimal policy is monotone in the state space.*

The unichain assumption is a strong assumption to prove the existence of a stationary average optimal policy. We note that weaker but sufficient conditions can be used in the existence proof instead of verifying the unichain property when the state space is finite. This will be discussed in the next section.

5.3 A Case Study

In this section, we provide a case study of the dynamic resource allocation problem. To this end, we assume that the packet arrivals in a frame follow a geometric distribution with the parameter Q, $0 < Q < 1$, i.e.,

$$q(i) = Q(1-Q)^i, \quad i = 0, 1, \cdots \tag{5.7}$$

As an example, we first show that stationary average optimal policies exist in this special case. We then discuss the choice of cost functions and provide some numerical examples for this case study.

5.3.1 Existence of Stationary Average Optimal Policies

In Section 2.2.3, the conditions for the existence of a stationary average optimal policy have been stated in Theorem 2.3. The first three conditions of Theorem 2.3 are easily verified from our problem definition, i.e., the state space and the action space are finite and the cost structure is uniformly bounded. However, it is not an easy task to examine the unichain property (though it seems straightforward) for all possible stationary policies as there may have altogether B^A stationary policies to be examined. Instead, we use the following theorem from Bertsekas (see Proposition 2.6 of Section

4.2 [7]) to verify that a stationary average optimal policy exists and it also satisfies the ACOE in our particular problem.

Theorem 5.1 (Bertsekas [7] page 198) *Assume that the state space $\mathcal{S}$ is finite. Further, for every two states s and s', there exists a stationary policy $\pi = (d, d, \cdots)$ (depending on s and s') such that, for some $t < \infty$, $\text{Tr}(s_t = s'|s_0 = s, \pi) > 0$. Then the optimal average cost is a constant independent of the starting state for all $s \in \mathcal{S}$, i.e.,*

$$\overline{V}^*(s) \equiv J, \quad \textit{for all } s \in \mathcal{S} \tag{5.8}$$

Furthermore, the ACOE given in (2.12) holds and an optimal stationary policy realizes its minimum for all $s \in \mathcal{S}$.

Recall that $\overline{V}^*(s)$, defined by (2.9), is the optimal policy value under the average optimal criterion. Indeed, the condition in Theorem 5.1 is a weaker version of the unichain property, which requires the existence of at least *one* stationary policy under which all states communicate when the state space is finite. The following proposition summarizes the existence of a stationary average optimal policy for our particular problem.

Proposition 5.2 *There exists a stationary average optimal policy satisfying the ACOE for our dynamic rate allocation problem. Further, the average optimal policy value is a constant independent of the starting state.*

Proof: The proof is based on showing that the conditions in Theorem 5.1 are satisfied in our problem.

Consider a stationary policy π' that always transmits one packet whenever the buffer is not empty, i.e., $\pi' = (d', d', \cdots)$ where $d'(s) = 1$, for all $s \geq 1$. We denote $d'(0) = 0$. Recall that in this case study, the batch arrivals follow a geometric distribution and $q(i) > 0$ for all $i \geq 0$. For any state s and state $s-1$, $s \geq 1$, we have $\text{Tr}(s|s-1, d'(s-1)) = q(1) > 0$, $s = 1$ and $\text{Tr}(s|s-1, d'(s-1)) = (1-f_s)q(1) + f_s q(2) > 0$, $s > 1$, and $\text{Tr}(s-1|s, d'(s)) = f_s q(0) > 0$. Thus state s communicates with state $s-1$ under policy π'. As s is arbitrarily chosen, we conclude that all states communicate with each other. Thus all states consist of a closed recurrent class under policy π' and the induced Markov chain is ergodic and irreducible, i.e., the conditions in Theorem 5.1 are satisfied. We then conclude that a stationary average optimal policy satisfying the ACOE exists for our problem. ■

We may have a similar analysis for any other stationary policy that consists of the decision rules such that the sender has to send (at least one) packet(s) whenever the buffer is not empty. And indeed the Markov decision problem is unichain if $d(s) > 0$ for all $s \geq 1$. Furthermore, we may also prove that a stationary optimal policy exists and the ACOE applies if the state space is denumerable infinite, i.e., the buffer is infinite. However this needs other conditions, namely, $\frac{1-Q}{Q} < f_s A$ (for a stable queueing system) and there exists a finite constant N and a non-negative integer n such that $C_h(s) \leq Ns^n$. The existence proof becomes more complicated in the case of a denumerable infinite state space. We do not discuss it further here and the interested reader can refer to Sennott [73].

5.3.2 Choice of Cost Functions

A stationary policy does not depend on the decision epochs and hence we can denote it as a vector $\mu = (\mu(1), \mu(2), \cdots, \mu(B))$ with the action $\mu(s) \in \mathcal{A}_s$ associated with the state s, $1 \leq s \leq B$. We let $\mu(0) = 0$. We have shown in Proposition 5.2 that the induced Markov chain of our problem is ergodic and irreducible. Thus, the steady state probability distribution of the buffer occupancy exists under a stationary policy μ and we denote it as $p^\mu = (p^\mu(0), p^\mu(1), \cdots, p^\mu(B))$ [4]. The buffer overflow probability associated with policy μ is denoted as P_o^μ and given as $P_o^\mu = p^\mu(B)$. Although buffer overflow is admissible, it is undesirable. Given a large enough but finite buffer limit B, we are more interested in a stationary policy resulting in no buffer overflow. Let $\overline{\mu} \equiv \mathbb{E}[\mu]$ denote the average number of packets to send in a frame under policy μ (i.e., the average service rate of policy μ) which can be computed as

$$\overline{\mu} = f_s \sum_{s=1}^{B} \mu(s) p^\mu(s) \tag{5.9}$$

[4] It may be hard to derive a closed form of p^μ for a general stationary policy. However we can get p^μ via simulations.

To avoid buffer overflow, we require that $\lambda < \overline{\mu}$. Thus we define *allowable* stationary polices where the average arrival rate is less than the average service rate. When overflow is taken into consideration, the effective average packet arrival rate (for those packets actually accepted and queued in the buffer) is given as

$$\overline{\lambda}^{\mu} = (1 - P_o^{\mu})(\frac{1}{Q} - 1) \tag{5.10}$$

The average number of packets in the system can be computed as:

$$\overline{N}^{\mu} = \sum_{s=1}^{B} s p^{\mu}(s) \tag{5.11}$$

By Little's Theorem, the average waiting time of a packet is then given as:

$$\overline{W}^{\mu} = \frac{1}{2} + \overline{N}^{\mu}/\overline{\lambda}^{\mu} = \frac{1}{2} + \frac{Q/(1-Q)}{1-P_o^{\mu}} \sum_{s=0}^{B} s p^{\mu}(s) \tag{5.12}$$

The constant $\frac{1}{2}$ in (5.12) accounts for the fact that a batch arrival can occur anywhere within a frame but the packets can only be transmitted after the current frame. Also the unit of $\overline{W}^{\mu}$ is in terms of frame time.

As the Markov chain induced by a stationary policy μ is ergodic and irreducible, we have an alternative method to compute the average cost associated with policy μ, i.e.,

$$\sum_{s=0}^{B} p^{\mu}(s)[C_u(\mu(s)) + C_h(s)] \tag{5.13}$$

Eq. (5.13) is the objective function of the corresponding dual linear programming problem formulation of the Markov decision problem (see Section 4.3.3 of Bertsekas [7] for example). It is clearly seen from (5.13) that we can use a linear holding cost function to (partly) characterize the average delay of a packet. Furthermore, we will use a simple nonlinear resource usage cost function. A discussion for such a usage function will be given in Section 5.5.1. Then, we suggest the following cost structure:

$$C(s,a) = C_u(a) + C_h(s) = c_0 a^{c_1} + s\ , \quad c_0 > 0,\ c_1 \geq 1 \tag{5.14}$$

The parameter c_0 may capture the relationship between the unit usage cost and the unit holding cost. The parameter c_1 is used to capture the effect of the nonlinear increase of the usage cost. Note that the cost functions given in (5.14) satisfy Assumption 5.1.

5.3.3 Average Delay Bounds

In this section, we consider two extreme policies which also provide the average delay bounds among all policies. Define a *least-effort policy* μ_l as that where only one packet is sent whenever the buffer is not empty. μ_l is given as:

$$\mu_l(s) = \begin{cases} 0, & s = 0 \\ 1, & s > 0 \end{cases} \tag{5.15}$$

Define a *most-effort policy* μ_m as that where packets in the buffer or the largest allowable number of packets whenever possible, whichever is smaller, are sent whenever possible. μ_m is given as:

$$\mu_m(s) = \begin{cases} 0, & s = 0 \\ \min\{s, A\}, & s > 0 \end{cases} \tag{5.16}$$

Note that μ_l and μ_m defined above are stationary polices. It is intuitively clear that μ_m minimizes the total number of packets in the buffer and hence has the lowest average waiting time, while μ_l results in the largest average delay. Recall that according to our action space definition, at least one packet should be sent whenever the buffer is not empty. We will show that μ_l and μ_m provide the upper and lower bounds of the average waiting time among all policies (and also the bounds of the buffer overflow probability), respectively, using the comparison method for stochastic processes and related theory (see Stoyan [80] for more references).

We will use the following definition of stochastic orders for random variables and stochastic processes and also a related theorem. First consider two real valued random variables $\hat{X}$ and $\hat{Y}$ defined on a common probability space. Then $\hat{X}$ is said to be *stochastically smaller* than $\hat{Y}$, denoted as $\hat{X} \leq_{st} \hat{Y}$, if $\Pr(\hat{X} > z) \leq \Pr(\hat{Y} > z)$ for all $z \in \mathbb{R}$. Furthermore, $\hat{X}$ is stochastically smaller than $\hat{Y}$ if and only if $\mathbb{E}[g(\hat{X})] \leq \mathbb{E}[g(\hat{Y})]$ for all nondecreasing functions $g : \mathbb{R} \mapsto \mathbb{R}$. In particular, if $g(x) = x$ and $\hat{X}$ is stochastically smaller than $\hat{Y}$, then $\mathbb{E}[\hat{X}] \leq \mathbb{E}[\hat{Y}]$. Now consider two discrete processes: $\mathbf{X} = \{X_t\}_{t=0}^{\infty}$ and $\mathbf{Y} = \{Y_t\}_{t=0}^{\infty}$. Let $\mathcal{R} = \mathbb{R}^{\mathbb{N}^+}$ denote the space of all real valued sequences. We say that the process $\mathbf{X}$ is *stochastically smaller* than the process $\mathbf{Y}$, denoted as $\mathbf{X} \leq_{st} \mathbf{Y}$, if $\Pr(g(X) > z) \leq \Pr(g(Y) > z)$ for every $z \in \mathbb{R}$, where $g : \mathcal{R} \mapsto \mathbb{R}$ is measurable and $g(X) \leq g(Y)$ for every $X, Y \in \mathcal{R}$ such that $X_t \leq Y_t$ for all $t \in \mathbb{N}^+$.

The following theorem from [80] provides alternative characterizations of the stochastic orders between two processes.

Theorem 5.2 (Stoyan 1983 [80]) *Consider two discrete time stochastic processes* $\mathbf{X} = \{X_t\}_{t=0}^{\infty}$ *and* $\mathbf{Y} = \{Y_t\}_{t=0}^{\infty}$. *The following three statements are equivalent.*

(i) $\mathbf{X} \leq_{st} \mathbf{Y}$

(ii) $\Pr(g(X_0, \cdots, X_k) > z) \leq \Pr(g(Y_0, \cdots, Y_k) > z)$ *for all* $z \in \mathbb{R}$, $k \in \mathbb{N}^+$, *and for all* $g : \mathbb{R}^t \mapsto \mathbb{R}$, *measurable and such that* $X_i \leq Y_i, 0 \leq i \leq k$, *implies that* $g(X_0, \cdots, X_k) \leq g(Y_0, \cdots, Y_k)$.

(iii) *There exists two stochastic processes* $\mathbf{X}' = \{X'_t\}_{t=0}^{\infty}$ *and* $\mathbf{Y}' = \{Y'_t\}_{t=0}^{\infty}$ *on a common probability space with the same probability laws as* $\mathbf{X}$ *and* $\mathbf{Y}$, *respectively, such that* $X'_t \leq Y'_t$ *almost surely (a.s.) for every* $t \in \mathbb{N}^+$.

We note that if we can prove that the total number of packets in the buffer under policy μ_m is stochastic smaller than that under some other policy π, then the expected average number of packets in the buffer under μ_m is smaller than that under policy π. The construction method in Theorem 5.2 (iii) is also known as *stochastic coupling*. Another application of Theorem 5.2 is as follows, which needs stronger assumptions. If we can find two random variables $\hat{X}$ and $\hat{Y}$ on a common probability space such that when t goes to infinity, the processes $\mathbf{X}$ and $\mathbf{Y}$ converge in law (in distribution for example) to $\hat{X}$ and $\hat{Y}$, denoted as $\mathbf{X} \rightarrow_t \hat{X}$ and $\mathbf{Y} \rightarrow_t \hat{Y}$, respectively, then $\mathbf{X} \leq_{st} \mathbf{Y}$ implies that $\hat{X} \leq_{st} \hat{Y}$ and $\mathbb{E}[\hat{X}] \leq \mathbb{E}[\hat{Y}]$. Finally, we note that actually, policy μ_m is stochastically smaller than any other policy, not restricted to stationary policies only. The following lemma states that μ_m minimizes in the stochastic ordering sense the total number of packets in the system.

Lemma 5.3 *Let* $\mathbf{S}^{\pi} = \{S_t^{\pi}\}_{t=0}^{\infty}$ *be the buffer process with the starting state* s_0 *when a policy* π *is applied. Let* $\mathbf{S}^{\mu_m} = \{S_t^{\mu_m}\}_{t=0}^{\infty}$ *be the corresponding buffer process when policy* μ_m *is used. Then*

$$\mathbf{S}^{\mu_m} \leq_{st} \mathbf{S}^{\pi} \tag{5.17}$$

Proof: The proof consists of two steps. For any arbitrary policy π, there exists a policy π_0 such that π_0 acts in the same way as policy μ_m at $t = 0$. Further, the corresponding buffer processes satisfy $S_t^{\pi_0} \leq S_t^{\pi}$ almost surely (a.s.) for $t = 0, 1, \cdots$. Then in the second step, we show that such a construction can be repeated and finally leads to our result according to Theorem 5.2.

We construct a policy π_0 and couple the buffer occupancy realizations under π and π_0 as follows. We construct a policy π_0 that acts the same as μ_m at time 0 and acts almost the same as policy π for $t = 1, \cdots$ in the sense that it either sends the same number of packets as π does or all the packets in the buffer, whichever is smaller. Furthermore, without loss of generality, we assume that policy π successfully transmits packet(s) in frame 0. Then policy π_0 is constructed with a successful transmission in frame 0 accordingly. If π_0 and π take the same action at $t = 0$, then the buffer processes are actually identical and hence we have $S_t^{\pi_0} = S_t^{\pi}$. If π sends fewer packets than π_0 at $t = 0$, then we have $S_1^{\pi_0} \leq S_1^{\pi}$ at $t = 1$. We can easily see that if $S_t^{\pi_0} \leq S_t^{\pi}$ holds at t, it also holds at $t+1$ via an induction argument. Hence we can have $S_t^{\pi_0} \leq S_t^{1}$ almost surely (a.s) for all t from the induction argument.

Now we repeat the construction, and we can show there exists a policy π_1 that agrees with π_0 at the first frame, agrees with μ_m at the second frame, and agrees again with π_0 at the third and all subsequent frames. We then have $S_t^{\pi_1} \leq S_t^{\pi_0}$ almost surely (a.s.) for $t = 0, 1, \cdots$. We can repeat the above argument k times and obtain policies π_i, $i = 0, 1, \cdots, k$ such that for the corresponding processes, we have

$$S_t^{\pi_k} \leq S_t^{\pi_{k-1}} \leq \cdots \leq S_t^{\pi_0} \leq S_t^{\pi} \text{ a.s.}, \ t = 0, 1, \cdots \tag{5.18}$$

For frames $0, 1, \cdots, k$ and a function g as in Theorem 5.2 (ii), consider policies π_k and μ_m. By construction, the total number of packets in the system under policy π_k, $S_0^{\pi_k}$, $S_1^{\pi_k}$, $\cdots$, $S_k^{\pi_k}$ have the same joint probability distribution with $S_0^{\mu_m}$, $S_1^{\mu_m}$, $\cdots$, $S_k^{\mu_m}$ under policy μ_m. Hence for all z, we have

$$\Pr(g(S_0^{\pi_k}, S_1^{\pi_k}, \cdots, S_k^{\pi_k}) > z) = \Pr(g(S_0^{\mu_m}, S_1^{\mu_m}, \cdots, S_k^{\mu_m}) > z) \tag{5.19}$$

From (5.18), $S_t^{\pi_k} \leq S_t^{\pi}$ almost surely (a.s.) for all $t = 0, 1, \cdots$. Therefore, we have

$$\Pr(g(S_0^{\pi_k}, S_1^{\pi_k}, \cdots, S_k^{\pi_k}) > z) \leq \Pr(g(S_0^{\pi}, S_1^{\pi}, \cdots, S_k^{\pi}) > z) \tag{5.20}$$

From Theorem 5.2(ii) and (5.19) and (5.20), we conclude $\mathbf{S}^{\mu_m} \leq_{st} \mathbf{S}^{\pi}$. ■

Although policy μ_m minimizes the expected total number of packets in the buffer, it may not necessarily be the average cost optimal one when the resource usage cost is taken into consideration. This is clear with the numerical example in the next section. Via similar arguments in Lemma 5.3, we have the following lemma for the least-effort policy and we omit the proof here.

Lemma 5.4 *Let $\mathbf{S}^\pi = \{S_t^\pi\}_{t=0}^\infty$ be the buffer process with the starting state s_0 when a policy π is applied. Let $\mathbf{S}^{\mu_l} = \{S_t^{\mu_l}\}_{t=0}^\infty$ be the corresponding buffer process when policy μ_l is used. Then*

$$\mathbf{S}^\pi \leq_{st} \mathbf{S}^{\mu_l} \tag{5.21}$$

We note that the queueing system under policy μ_l may not be a stable one if the arrival rate $\lambda > 1$. In such a case, policy μ_l also results in the largest buffer overflow probability according to Lemma 5.4. Indeed, since the buffer is assumed finite and the arrival process is assumed to be identically and independently distributed, it can be shown that under any stationary policy, the system may eventually reach a statistical equilibrium or steady-state regime (cf. Proposition 5.2 that the induced Markov chain is ergodic and irreducible). Thus we may safely assume that for any stationary policy μ, there exists some random variable $\hat{S}^\mu$ such that S_t^μ converges in law to $\hat{S}^\mu$ as t goes to infinity, denoted as $S_t^\mu \to_t \hat{S}^\mu$. Furthermore, if $S_t^\mu \to_t \hat{S}^\mu$ in distribution, then $\mathbb{E}[g(S_t^\mu)] \to \mathbb{E}[g(\hat{S}^\mu)]$ as $t \to \infty$ for all bounded continuous functions g. We then summarize the results of this section in the following proposition.

Proposition 5.3 *Assume that the buffer process S_t^μ under any stationary policy μ converges in law as t goes to infinity. Then the average waiting time of policy μ is upper and lower bounded as $\overline{W}^{\mu_m} \leq \overline{W}^{\mu} \leq \overline{W}^{\mu_l}$. Further, the buffer overflow probability is upper and lower bounded as: $P_o^{\mu_m} \leq P_o^{\mu} \leq P_o^{\mu_l}$.*

Proof: From Lemma 5.3 and 5.4, we have $\hat{S}^{\mu_m} \leq_{st} \hat{S}^{\mu} \leq_{st} \hat{S}^{\mu_l}$. Then by Little's law, we know that the average waiting time of policy μ is upper and lower bounded by that under policy μ_l and μ_m, respectively. The second claim follows from the definition of stochastic orders for random variables. ■

5.3.4 Numerical Examples

We provide some numerical examples in this section. We compute optimal policies under different settings of the usage function (5.14) and compare the optimal policies with other policies. We set $Q = 0.2$ (hence on average 4 packets arrive in a frame) and $f_s = 0.95$ in all simulations. The available actions are to send from 1 to 10 packets, i.e., $\mathcal{A} = \{1, 2, \cdots, 10\}$. The buffer limit is set as $B = 200$ in the computation of optimal policies but we will examine the buffer overflow probabilities by assuming a smaller buffer.

Fig. 5.2 shows the computed optimal policies under a fixed $c_1 = 1.6$ but with different choices of c_0. As expected, the optimal policies are monotonically increasing with the buffer occupancy. Note that c_0 may (partially) represent the ratio of the cost

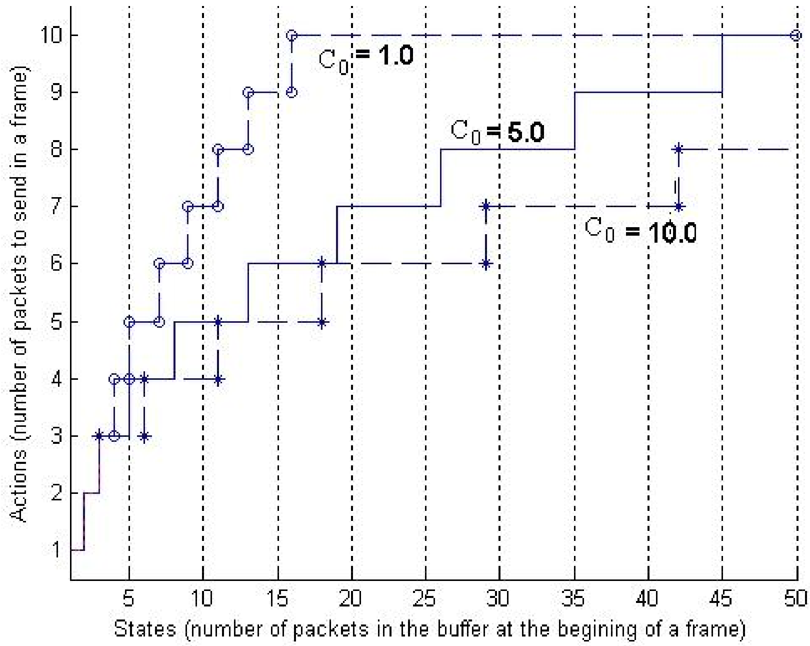

Figure 5.2: Optimal policies with respect to different c_0

of transmitting a packet to the cost of holding a packet. From Fig. 5.2, we see that given a buffer state (e.g., 20 packets in the buffer), fewer packets are transmitted when c_0 is larger. In other words, with increasing unit usage cost (price per unit resource), the user prefers to hold more packets in its buffer and defers the transmission of more packets in order to be total cost optimal.

We next compare the performance of different policies summarized in Table 5.1. Policies μ_1, μ_2 and μ_3 are the average cost optimal policies but with different available

Table 5.1: Different Rate Allocation Policies

μ_1	optimal policy with available actions $\{1, \cdots, 10\}$
μ_2	optimal policy with available actions $\{1, 5, 10\}$
μ_3	optimal policy with available actions $\{1, 10\}$
μ_4	least-effort policy with a single action $\{1\}$
μ_5	most-effort policy with available actions $\{1, \cdots, 10\}$
μ_6	dull policy with a single action $\{5\}$

actions. Policy μ_3 (μ_2) may represent the situation in which only the minimum and maximum (average) transmission rates are available. Policy μ_4 always transmits one packet whenever the buffer is not empty. Note that buffer overflow is inevitable under policy μ_4 when $Q = 0.2$. Policy μ_5 is one that either depletes the buffer when $s < 10$ or transmits with the largest rate when $s \geq 10$. Policy μ_6 is one that always transmits 5 packets when $s \geq 5$ and does not transmit when $s < 5$.

We set $c_0 = 5.0$, $c_1 = 1.6$ and $B = 80$ in the simulations. The buffer size is chosen in order to examine the buffer overflow probability. In the simulations, immediate costs are computed and collected in each frame based on the system state at the beginning of the frame and the action prescribed by the policies. In the simulations, we also count the buffer state and record the delay incurred by successfully transmitted packets in each frame. The simulation results include the average cost, the average delay and the probability of buffer overflow. They are computed by averaging over 10 runs of simulations each with 100000 frames.

Fig. 5.3 shows the average total costs (including the average usage and holding costs), usage costs and holding costs of the different policies. We see that policy μ_1 has the smallest average total cost among all the policies. In particular, the average total cost of μ_1 is smaller than those of μ_2 and μ_3 which are also optimal policies but have fewer available actions. This indicates that more options (more available actions) is better. Note that the queueing system is not stable under policy μ_4 and its buffer occupancy increases rapidly towards the buffer size. Thus policy μ_4 is not an *allowable*

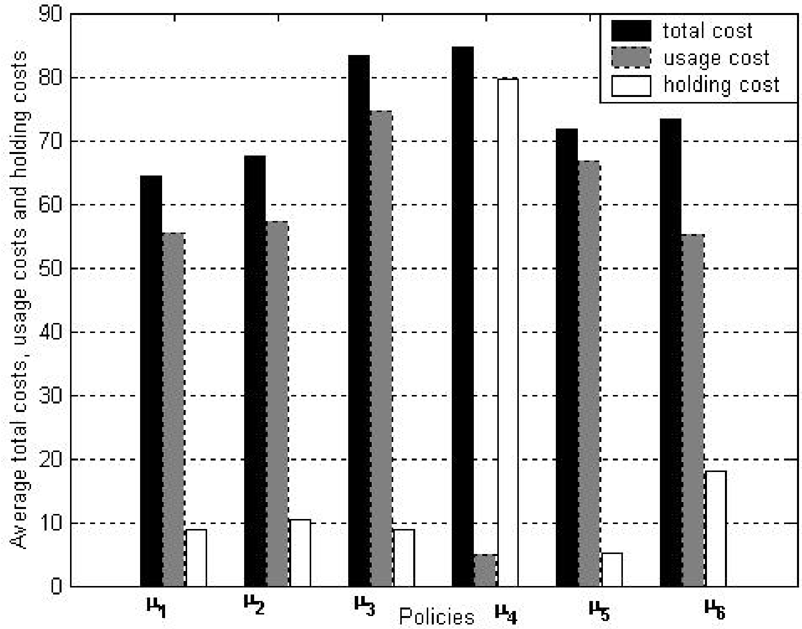

Figure 5.3: Average costs of different policies

policy as it results in a non-stable queueing system. Except policy μ_4 the rest of the policies are allowable policies as the induced queueing system is stable and their average holding costs will not tend to infinity as the buffer size tends to infinity. From Fig. 5.3, we see that although policy μ_5 has the smallest holding cost among all the policies, its average total cost is larger than that of policy μ_1 due to its high usage cost. From Fig. 5.3, we also see that the average total cost of policy μ_2 is close to that of policy μ_1. This indicates that policy μ_1 can be well approximated by a threshold-based policy with fewer available actions to simplify implementation.

Table 5.2: Performance comparison of different policies

	μ_1	μ_2	μ_3	$\mu_4(\mu_l)$	$\mu_5(\mu_m)$	μ_6
$\overline{W}$	2.73	3.07	2.70	84.32	1.76	4.59
$\overline{N}$	8.91	10.27	8.80	79.68	5.03	17.96
$P_o(10^{-4})$	0.0175	0.0325	0.0	7620	0.0	41.2

Table 5.2 compares the average waiting time $\overline{W}$, the average number of packets in the buffer $\overline{N}$ and the average buffer overflow probability P_o for the different policies,

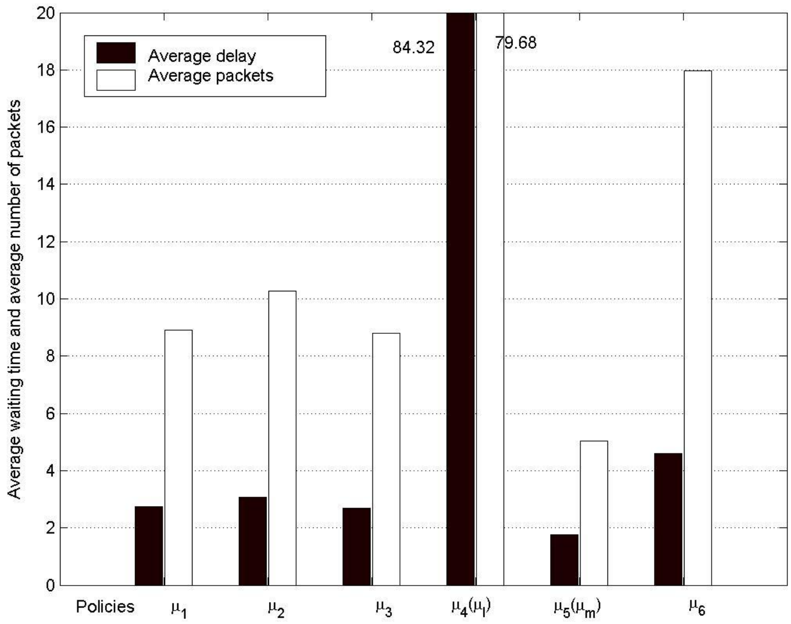

Figure 5.4: Average delay and average buffer occupancy of different policies

which are also represented with Fig. 5.4. The numerical results also verify Proposition 5.3, i.e., the average waiting time is upper and lower bounded by that of policy μ_4 (μ_l) and μ_5 (μ_m), respectively. We see that policy μ_4 (μ_l) has a very large buffer overflow probability. This is because the service rate (0.95 packet/frame) of policy μ_4 (μ_l) is far lower than its arrival rate. We note that policy μ_5 (μ_m) has the smallest $\overline{W}$, $\overline{N}$ and P_o compared with the other policies. However policy μ_5 (μ_l) is not average cost optimal (as shown in Fig. 5.3) since its usage cost is also very high. We also note that policy μ_6 has larger $\overline{W}$, $\overline{N}$ and P_o compared with policies μ_1 and μ_2. These indicate that using variable service rates is better than using a single service rate.

5.4 A Class of Simple Policies

In previous sections, we have shown that the optimal policies have a monotone property. Furthermore, the numerical results suggest that some threshold-based policies with fewer available actions may well approximate the optimal policies with a full set of available actions. Thus in this section, we propose and analyze a class of simple policies with a single threshold.

5.4.1 A Class of Threshold-based Simple Policies

We define a class of simple policies μ_z with one threshold as follows. Denote $\underline{\lambda} = \lceil \frac{\lambda}{f_s} \rceil$ and $\overline{\lambda} = \lceil \frac{\lambda}{f_s} \rceil + 1$. For a given threshold $z \in (0, \underline{\lambda}]$, we partition the buffer into two parts: $(0, z)$ and $[z, \infty)$. Policy μ_z is defined as:

$$\mu_z(s) = \begin{cases} \min\{s, \underline{a}\}, & s < z, \quad 1 \leq \underline{a} \leq \underline{\lambda} \\ \min\{s, \overline{a}\}, & s \geq z, \quad \overline{\lambda} \leq \overline{a} \leq A \end{cases} \tag{5.22}$$

Obviously, we may have many other threshold options and also the available actions in different partitioned sets. However, the threshold used here may (partly) characterize the arrival process and the effect of retransmissions, i.e., $\underline{\lambda}$ (or $\overline{\lambda}$) may characterize the effective average number of packets (plus possible retransmitted packets) that arrive in a frame. The intuition behind the simple policies is as follows. We do not want the buffer occupancy to get too large which may increase the holding cost as well as the probability of overflow. On the other hand, keeping the buffer occupancy too small may also be undesirable as it may need more actions that increase the usage cost. Note that buffer occupancy either stays less than the threshold or stays greater than the threshold. Thus policy μ_z is in essence trying to control the buffer occupancy around the threshold (as a compromise).

Let $\{Q_t\}_{t=0}^{\infty}$ denote the arrival process. As the arrival process is IID, we sometimes uses $\hat{Q}$ to denote the corresponding random variable. Let $U_t^{\mu_z} = \mu_z(s)$ denote the number of packets to transmit in frame t given the buffer occupancy is s at the beginning of frame t and the policy μ_z is applied. Let $\mathbf{S}^{\mu_z} = \{S_t^{\mu_z}\}_{t=0}^{\infty}$ be the buffer occupancy process under policy μ_z. Thus the dynamics of the buffer process under policy μ_z can be written as

$$S_{t+1}^{\mu_z} = \min\{B, Q_t + [S_t^{\mu_z} - f_s U_t^{\mu_z}]^+\}, \quad t \geq 0 \tag{5.23}$$

in which $[x]^+ = \max\{0, x\}$. Since μ_z implies that the buffer occupancy will be controlled around the threshold that is larger than the average effective arrival rate, the queueing system under policy μ_z is stable and buffer overflow may be avoided provided that $B \gg z$. With this in mind, it is reasonable to look instead at the infinite buffer system

($B = \infty$) associated with (5.23). Then we consider the following recursion of buffer occupancy instead:

$$S_{t+1}^{\mu_z} = \max\{S_t^{\mu_z} + Q_t - f_s U_t^{\mu_z}, Q_t\}, \;\; t \geq 0. \tag{5.24}$$

Now we may again apply the stochastic comparison method to estimate the buffer average occupancy and to provide a (more accurate) average delay bound for these simple policies. We discuss it in the next section.

5.4.2 An Upper Bound for Average Delay

Besides the stochastic comparison method, we will also use some theory on *random walks* to establish our analytical results. We have introduced some basics of stochastic comparison before and only provide some preliminaries of random walk here. R. Gallager [26] (1995) provides more discussions on random walks. Let $\{X_i\}_{i=0}^{\infty}$ be a sequence of identical and independent distributed (IID) random variables, i.e, X_i can be considered as copies of a random variable X with mean $\mathbb{E}[X] = \overline{X} < 0$. Let $\psi(r) = ln(\mathbb{E}[e^{Xr}])$ be the semi-invariant moment generating function of X. We assume that $\psi(r)$ is finite in an open interval (r^-, r^+), $r^- < 0 < r^+$ and $\psi(r)$ has a root at $r^* > 0$. Let $\{Y_{t+1}\}_{t=0}^{\infty}$ be a process defined by Y_0 and $Y_{t+1} = \max\{Y_t + X_t, 0\}$. Thus $\{Y_t\}$ is a random walk restricted to the positive axis. Let $\hat{Y}$ be a random variable with the steady state distribution. Assume that $Y_t \rightarrow_t \hat{Y}$ almost surely (a.s.) and $\lim_{t\to\infty} \Pr(Y_t > y) = \Pr(\hat{Y} > y)$. From the theory of random walks, we have the following result:

$$\lim_{t\to\infty} \Pr(Y_t > y) = \Pr(\hat{Y} > y) \leq e^{-r^* y}, \;\; \text{for all} \;\; y \geq 0 \tag{5.25}$$

When Y_t represents the waiting time in a $G/G/1$ queue, this result is known as Kingman's Bound (R. Gallager [26], page 234).

We define a second buffer process used as an auxiliary process in the proof of the lemma. Consider the following control policy:

$$\tilde{\mu}_z(s) = \begin{cases} 0, & s < z, \\ \min\{s - z, \overline{a}\}, & s \geq z, \;\; \overline{\lambda} \leq \overline{a} \leq A \end{cases} \tag{5.26}$$

The above control policy implies that the sender only transmits when the buffer occupancy exceeds the threshold and does not transmit otherwise. The second buffer process $\mathbf{S}^{\tilde{\mu}_z} = \{S_t^{\tilde{\mu}_z}\}_{t=0}^{\infty}$ under policy $\tilde{\mu}_z$ is assumed to experience the same arrival statistics in all frames as the process $\mathbf{S}^{\mu_z} = \{S_t^{\mu_z}\}_{t=0}^{\infty}$. It is given as

$$S_{t+1}^{\tilde{\mu}_z} = \max\{S_t^{\tilde{\mu}_z} + Q_{t+1} - f_s U_t^{\tilde{\mu}_z}, Q_{t+1}\} \tag{5.27}$$

and $S_0^{\tilde{\mu}_z} = \max\{S_0^{\mu_z}, z\}$. Note that the second process $\mathbf{S}^{\tilde{\mu}_z}$ will be restricted to stay in $[z, \infty)$ for all frames. Let $\hat{S}^{\mu_z}$ and $\hat{S}^{\tilde{\mu}_z}$ be the random variables with the same steady state distributions for the processes $\mathbf{S}^{\mu_z}$ and $\mathbf{S}^{\tilde{\mu}_z}$ (if they exist), respectively. We make the following assumption.

Assumption 5.2 *Assume that* $S_t^{\mu_z} \to_t \hat{S}^{\mu_z}$ *almost surely (a.s.) and* $\lim_{t\to\infty} \Pr(S_t^{\mu_z} > s) = \Pr(\hat{S}^{\mu_z} > s)$*, and* $S_t^{\tilde{\mu}_z} \to_t \hat{S}^{\tilde{\mu}_z}$ *almost surely (a.s.) and* $\lim_{t\to\infty} \Pr(S_t^{\tilde{\mu}_z} > s) = \Pr(\hat{S}^{\tilde{\mu}_z} > s)$

It may be shown that the assumption is true when the induced Markov chains are ergodic and irreducible under the stationary policy μ_z and $\tilde{\mu}_z$, respectively. The following lemma provides an upper bound on the average number of packets in the buffer for a simple policy with one threshold. We should note that the bound may not be the tightest one. However, it provides some insights on the buffer occupancy.

Lemma 5.5 *Assumption 5.2 holds. Then the average buffer occupancy under policy* μ_z *satisfies*

$$\mathbb{E}[\hat{S}^{\mu_z}] \leq z + \frac{e^{r^*(\overline{a}) f_s \overline{a}}}{r^*(\overline{a})} \tag{5.28}$$

where $r^*(\overline{a})$ *is the unique positive root of* $\psi(r) = ln(\mathbb{E}_{\hat{Q}}[e^{(\hat{Q}-\overline{a})r}])$ *and* $\mathbb{E}_{\hat{Q}}$ *is the expectation with respect to the arrival process.*

Proof: The proof consists of two steps. We first show that $\mathbf{S}^{\mu_z} \leq_{st} \mathbf{S}^{\tilde{\mu}_z}$. We then show that $\hat{S}^{\tilde{\mu}_z}$ has an exponential bound.

As we assume the steady state distribution exists, then $\mathbb{E}[\hat{S}^{\mu_z}]$ can be written as the complementary distribution function of $\hat{S}^{\mu_z}$ as follows

$$\mathbb{E}[\hat{S}^{\mu_z}] = \int_0^{\infty} \Pr(\hat{S}^{\mu_z} > s) ds \tag{5.29}$$

For $s < z$, we upper bound $\Pr(\hat{S}^{\mu_z} > s)$ by 1, then we have

$$\begin{aligned} \mathbb{E}[\hat{S}^{\mu_z}] &= \int_0^z \Pr(\hat{S}^{\mu_z} > s)ds + \int_z^\infty \Pr(\hat{S}^{\mu_z} > s)ds \\ &\leq z + \int_0^\infty \Pr(\hat{S}^{\mu_z} > s + z)ds \end{aligned} \tag{5.30}$$

Next, we consider how to bound $\Pr(\hat{S}^{\mu_z} > s + z)$. We resort to the auxiliary process. We may show that $\mathbf{S}^{\mu_z}$ is stochastically smaller than $\mathbf{S}^{\tilde{\mu}_z}$ by a constant, i.e., $\mathbf{S}^{\mu_z} \leq_{st} \mathbf{S}^{\tilde{\mu}_z} + f_s\overline{a}$. Indeed the two processes can be related with the following inequality.

$$S_t^{\tilde{\mu}_z} \geq S_t^{\mu_z} - f_s\overline{a}\ , \quad \text{for all} \;\; t = 0, 1, \cdots \tag{5.31}$$

We show (5.31) by induction arguments. By assumption, it is true at frame 0, as $S_0^{\tilde{\mu}_z} \geq S_0^{\mu_z} \geq S_0^{\mu_z} - f_s\overline{a}$. Now assume that at frame $t, t \geq 0$, $S_t^{\tilde{\mu}_z} \geq S_t^{\mu_z} - f_s\overline{a}$, we will show that it also holds for frame $t + 1$. We consider two cases: (1) $S_t^{\mu_z} \geq z$, and (2) $S_t^{\mu_z} < z$.

Case (1): $S_t^{\mu_z} \geq z$. In this case, $U_t^{\tilde{\mu}_z} \leq U_t^{\mu_z}$ and by induction hypothesis $S_t^{\tilde{\mu}_z} \geq S_t^{\mu_z} - f_s\overline{a}$, hence we have

$$\begin{aligned} S_{t+1}^{\tilde{\mu}_z} &= \max\{S_t^{\tilde{\mu}_z} + Q_{t+1} - f_sU_t^{\tilde{\mu}_z}, Q_{t+1}\} \\ &\geq \max\{S_t^{\mu_z} - f_s\overline{a} + Q_{t+1} - f_sU_t^{\mu_z}, Q_{t+1}\} \\ &\geq \max\{S_t^{\mu_z} + Q_{t+1} - f_sU_t^{\mu_z}, Q_{t+1}\} - f_s\overline{a} \\ &= S_{t+1}^{\mu_z} - f_s\overline{a} \end{aligned}$$

Case (2): $S_t^{\mu_z} < z$. In this case, $S_t^{\tilde{\mu}_z} \geq z \geq S_t^{\mu_z}$ and $U_t^{\tilde{\mu}_z} \leq \overline{a} \leq U_t^{\mu_z} + \overline{a}$, hence we have

$$\begin{aligned} S_{t+1}^{\tilde{\mu}_z} &= \max\{S_t^{\tilde{\mu}_z} + Q_{t+1} - f_sU_t^{\tilde{\mu}_z}, Q_{t+1}\} \\ &\geq \max\{S_t^{\mu_z} + Q_{t+1} - f_sU_t^{\mu_z} - f_s\overline{a}, Q_{t+1}\} \\ &\geq \max\{S_t^{\mu_z} + Q_{t+1} - f_sU_t^{\mu_z}, Q_{t+1}\} - f_s\overline{a} \\ &= S_{t+1}^{\mu_z} - f_s\overline{a} \end{aligned}$$

From the induction arguments, we now conclude $S_t^{\tilde{\mu}_z} \geq S_t^{\mu_z} - f_s\overline{a}$ for all $t \geq 0$ and hence $\Pr(S_t^{\mu_z} > s + z) \leq \Pr(S_t^{\tilde{\mu}_z} > s + z - f_s\overline{a})$ almost surely (a.s.) for all s and $t \geq 0$. Under the Assumption 5.2 and by letting $t \to \infty$, we have

$$\Pr(\hat{S}^{\mu_z} > s + z) \leq \Pr(\hat{S}^{\tilde{\mu}_z} > s + z - f_s\overline{a}) \tag{5.32}$$

Note that the process $\{S_t^{\tilde{\mu}_z}\}_{t=0}^{\infty}$ is a random walk restricted to $[z, \infty)$ and $\mathbb{E}[\hat{Q} - f_s\overline{a}] = \lambda - f_s\overline{a} < 0$. Thus using (5.25) we have

$$\Pr(\hat{S}^{\tilde{\mu}_z} > s + z - f_s\overline{a}) \leq e^{-r^*(\overline{a})(s - f_s\overline{a})}$$

and also

$$\Pr(\hat{S}^{\mu_z} > s + z) \leq e^{-r^*(\overline{a})(s - f_s\overline{a})}$$

Substituting this into (5.30), we have

$$\begin{aligned} \mathbb{E}[\hat{S}^{\mu_z}] \leq & \; z + \textstyle\int_0^{\infty} e^{-r^*(\overline{a})(s - f_s\overline{a})} ds \\ = & \; z + \frac{e^{r^*(\overline{a}) f_s \overline{a}}}{r^*(\overline{a})}, \end{aligned} \tag{5.33}$$

which the desired result and this completes the proof. ■

Proposition 5.4 *Assume that the induced Markov chain under policy μ_z is ergodic and the buffer overflow does not occur. The average waiting time $\overline{W}^{\mu_z} \equiv \mathbb{E}[W^{\mu_z}]$ satisfies:*

$$\overline{W}^{\mu_z} \leq \frac{1}{2} + \frac{z}{\lambda} + \frac{e^{r^*(\overline{a}) f_s \overline{a}}}{r^*(\overline{a})\lambda} \tag{5.34}$$

Proof: From the assumption and Little's law, we have:

$$\mathbb{E}[W^{\mu_z}] = \frac{1}{2} + \frac{\mathbb{E}[\hat{S}^{\mu_z}]}{\lambda} \tag{5.35}$$

where the constant $\frac{1}{2}$ in (5.35) accounts for the fact that a batch arrival can occur anywhere within a frame but the packets can only be transmitted after the current frame. The proof follows by applying Lemma 5.5 in (5.35). ■

5.4.3 Numerical Examples

We present some numerical examples to illustrate the delay bound and the performance of the simple policies. We still consider a geometric arrival process with the probability function given by (5.7). Hence the mean arrival rate is $\lambda = \frac{1}{Q} - 1$. According to Lemma 5.5, $r^*(\overline{a})$ should satisfy:

$$\begin{aligned} & \mathbb{E}_{\hat{Q}}\left[e^{(\hat{Q} - \overline{a}) r^*(\overline{a})}\right] = 1 \\ \Rightarrow \; & \textstyle\sum_{i=0}^{\infty} Q(1-Q)^i e^{(i - \overline{a}) r^*(\overline{a})} = 1 \\ \Rightarrow \; & Q e^{-\overline{a} r^*(\overline{a})} + (1 - Q) e^{r^*(\overline{a})} - 1 = 0. \end{aligned} \tag{5.36}$$

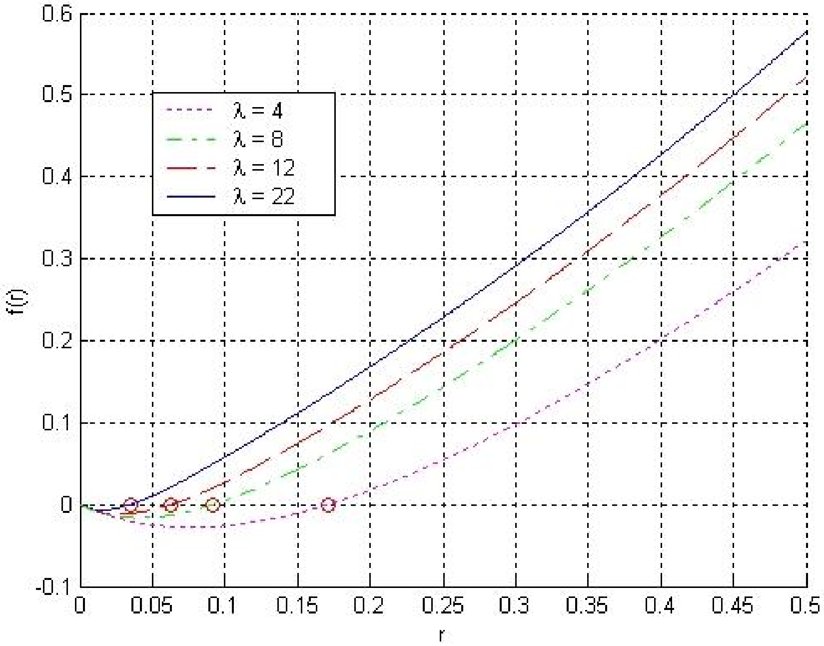

Figure 5.5: Examples of $f(r)$ ($Q = 1/(1+\lambda)$ and $\overline{a} = 2\lambda$)

Let $f(r) = Qe^{-\overline{a}r} + (1-Q)e^{r} - 1$. It is not hard to see that for a given Q, if $\overline{a} > \frac{1}{Q}$, then $f(r)$ has a unique positive root. This is satisfied as we require $\overline{a} > \lceil \frac{\lambda}{f_s} \rceil + 1 \geq \frac{1}{Q}$. We plot some examples of $f(r)$ and the corresponding positive root in Fig 5.5, where we set $Q = 1/(1+\lambda)$ and $\overline{a} = 2\lambda$.

We first investigate the delay bound and the average delay given different choices of the available actions. To compute the upper delay bound given by (5.34), we need to know $\overline{a}$ and z. We set the available actions first and we can determine the (optimal) threshold z by computing an optimal policy. We set $c_0 = 1.0$, $c_1 = 1.6$ and $B = 300$ when computing optimal policies. We evaluate the following three schemes: Scheme A has the available actions $\underline{a} = \lambda$ and $\overline{a} = 2\lambda$; Scheme B has the available actions $\underline{a} = 1$ and $\overline{a} = \lceil 1.5\lambda \rceil$; and Scheme C has the available actions $\underline{a} = 1$ and $\overline{a} = \lceil \frac{\lambda}{f_s} \rceil + 1$. Recall that we use the smaller action $\underline{a}$ and the larger action $\overline{a}$ to control the buffer occupancy close to zero and close to the threshold, respectively. Hence we can expect that the upper delay bound becomes a tight bound for small $\underline{a}$ and $\overline{a}$. Fig. 5.6 illustrates the delay bound and the average delay of different schemes as a function of the arrival rate. As expected, the delay bound for Scheme C is tighter than for Scheme B, and

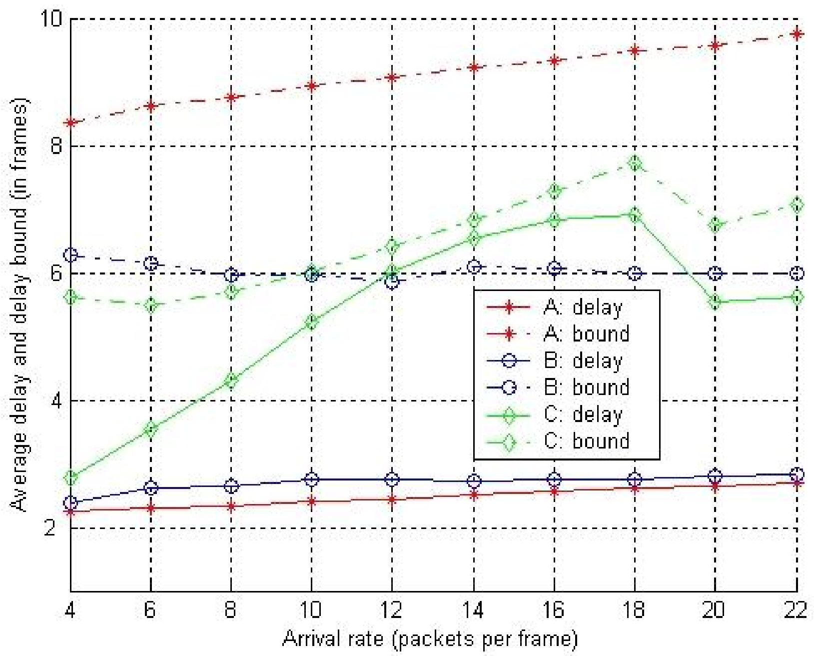

Figure 5.6: Average delay and delay bound of different optimal policies

the bound for Scheme B is tighter than that for Scheme A.

We next compare the performance of the simple optimal policies with only two available actions to those optimal policies with more than two available actions. Let N_a denote the number of available actions. We use the following method to set the available actions:

$$a_i = \lceil \frac{2\lambda}{N_a} \times i \rceil, \quad i = 1, 2, \cdots, N_a. \tag{5.37}$$

For example, when $\lambda = 4$ and $N_a = 3$, the available actions are $\mathcal{A} = \{3, 6, 8\}$. Note that the maximum available action is the same for different N_a, i.e., $a_{N_a} = \lceil 2\lambda \rceil$. The justification behind such a method is for a fair comparison. The range of the available actions is the same but different N_a quantifies the range differently. Indeed, (5.37) is to emulate the common quantization method and the larger N_a, the finer the quantization granularity.

Fig. 5.7 plots the policy value of optimal policies with different number of available actions as a function of the arrival rate. We observe that the policy value of the optimal policy with more available actions, i.e., with large N_a, is smaller than those with fewer available actions. This is due to the effect of a finer quantization with more available

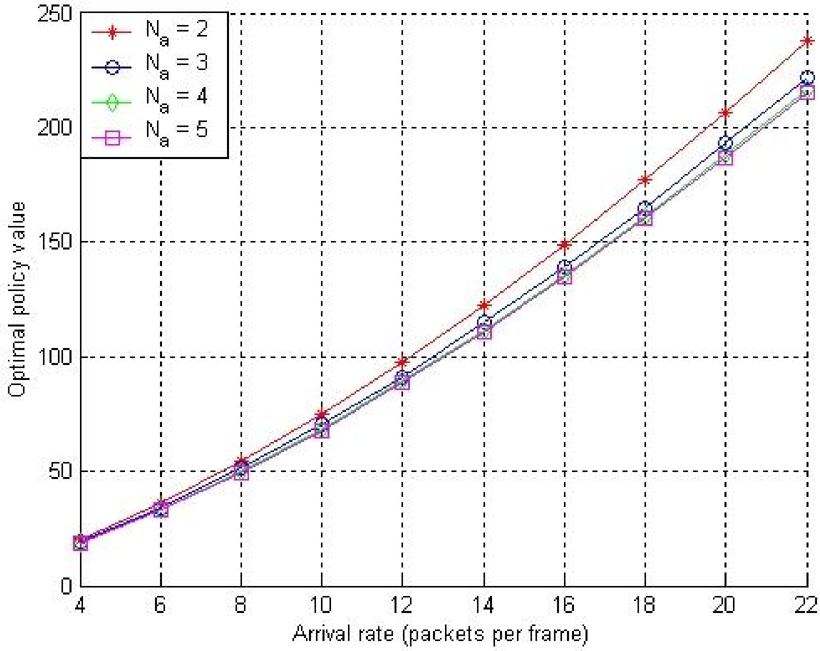

Figure 5.7: Policy value of optimal policies with different number of available actions

actions. However, we also observe that the policy values of the optimal policy with $N_a = 2$ are close to those with large N_a. This indicates that the simple optimal policies could be a good approximation.

Note that the method used in Section 5.4.2 to derive the upper delay bound for optimal policies with only two available actions also applies to those policies with more available actions, where we can simply use the largest buffer threshold and the largest action to derive the delay bound for the policies with more than one buffer threshold [5]. However, we should note that the upper delay bound becomes less tight for policies with more than two actions. Fig. 5.8 compares the delay bound and the average delay of the optimal policies with different number of available actions as a function of the arrival rate. We observe that the delay bound for the optimal policies with $N_a = 2$ is tighter than the others, i.e., the distance between the bound and the average delay is closer. We also observe that the average delay for the optimal policies with $N_a = 2$ is larger than the other policies, however, the differences are small. This again indicates

[5]The proof can be simply developed by upper bounding the probability for the buffer occupancy less than the largest threshold by 1 in (5.30).

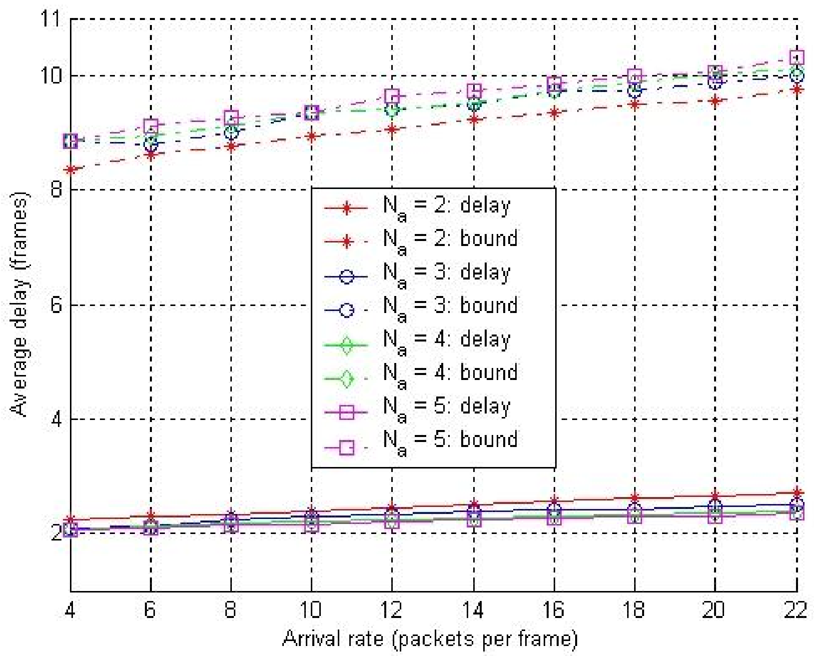

Figure 5.8: Average delay and delay bound of optimal policies with different number of available actions

that the simple policies could be a good approximation.

5.5 Extension to The Existence of Competitions

So far, we have discussed the problem of rate allocation for a single data user and formulated it as a Markov decision problem. In our problem defined in Section 5.1, the rate allocation is considered to be carried out independently and exclusively for a single user without considering any (possible) interactions across users. However, in a CDMA-based network, since multiple users can transmit simultaneously by using different spreading codes [40, 65], a user's transmission may impacts on all the other users' transmissions, e.g., the transmission power of a user also serves as interference to other users and impact the transmission quality of other users. Indeed, if the number of simultaneous transmitting users in a frame increases, a user has to use a higher transmission power to maintain the same transmission quality in general. This phenomenon can be seen as a kind of *competition* across users. Furthermore, there may exist some *competition costs* resulting from different numbers of simultaneously transmitting users, e.g., different transmission powers. Thus in this section, we extend the previous rate allocation problem and put the self-optimization problem for an isolated single user into the existence of competitions across multiple users. We should note that such a competitive and dynamic allocation problem would be better modelled from a *game* theory [24] framework [6], or more precisely, from a *competitive Markov decision processes* (*dynamic games*) [41] theory framework, which are also much more complicated. Nonetheless, we provide an alternative formulation that is still within the Markov decision processes theory framework for such a problem. In this section, we describe the competitions in detail and provide our modelling technique for such an extended problem. Some qualitative analysis are also provided in this section.

5.5.1 Competition Across Users

We explain the competition across users in detail with an example. It is well known that the transmission quality (e.g., BER or FSP, f_s) of a user is determined by its received signal to interference and noise ratio. In a CDMA-based network, the transmissions of other users form a major part of the interference seen by a particular user. However, the interference over the air may be hard to track as the number of active users [7] and their transmission rates can be time-varying (and unpredictable). In general, the more users transmit and the larger their transmission rates, the larger the interference over the air and hence the larger the transmission power a user needs to maintain the same transmission quality. Therefore, there are some costs associated with competing with other users and we call this the *competition cost*. The following example can serve as an explanation of the competition cost.

Consider the uplink transmission power allocation. To guarantee a minimum BER requirement, the following equation can be used to calculate the transmission power for an active user [3, 68]

$$p_i = \frac{g_i \eta_0 W}{h_i(1 - \sum_{j=1}^{n} g_j)}. \tag{5.38}$$

[6] We note that recently many researchers have applied game theory in wireless communications. For example, Mandayam et.al. applied game theory in wireless power control [31, 70, 71].

[7] In this section, the term *active users* means the users transmitting simultaneously in a frame.

We will discuss a bit more on (5.38) in the next chapter. In (5.38), p_i is the transmission power of user i, h_i is its path gain, W is the system spreading bandwidth, η_0 is the background noise spectral density (possibly plus inter-cell interference) at the base station, n is the number of users that transmit simultaneously and g_i is the *power index*, given by

$$g_i = \frac{\gamma_i}{\gamma_i + W/R_i} \tag{5.39}$$

In (5.39), R_i is the instantaneous data rate of user i and γ_i is the target E_b/I_0 (the bit-energy-to-interference-power-spectral-density) of user i with transmission rate R_i. It can be observed from (5.38) that assuming that all active users have the same power index, the transmission power of user i is increasing in the number of active users, i.e., p_i is increasing in n. Another observation from (5.38) and (5.39) is that the larger the transmission rate of a user, the larger the transmission power of all users, provided that all other conditions remain unchanged. Thus a nonlinear resource usage cost function, cf. (5.14), can be used to cover such a situation, i.e., the large transmission rate should be charged much more. On the other hand, we can simply approximate the competition cost by relating it to the number of active users in the system only.

The number of active data users in the system can be modelled either by a stochastic process (called *competition process* hereafter), or as a degenerate case, by an identical independent distributed random variable with a common support, i.e., a memoryless competition process. A data user needs to request (and use) different transmission rates based on the usage charge, its own transmission requirement, and the system competition dynamics. Thus we put a single user self-optimization into the presence of multiple users, i.e., the effect of multiple competitive users is represented by the competition process in our model.

5.5.2 Extended Problem Formulation

We still formulate the rate allocation problem for a single user as a Markov decision problem. However, the competition process is included in our problem formulation. The system model is shown in Fig. 5.9, which is almost the same as the model in Fig. 5.1 except for the information on the number of active users that is taken into

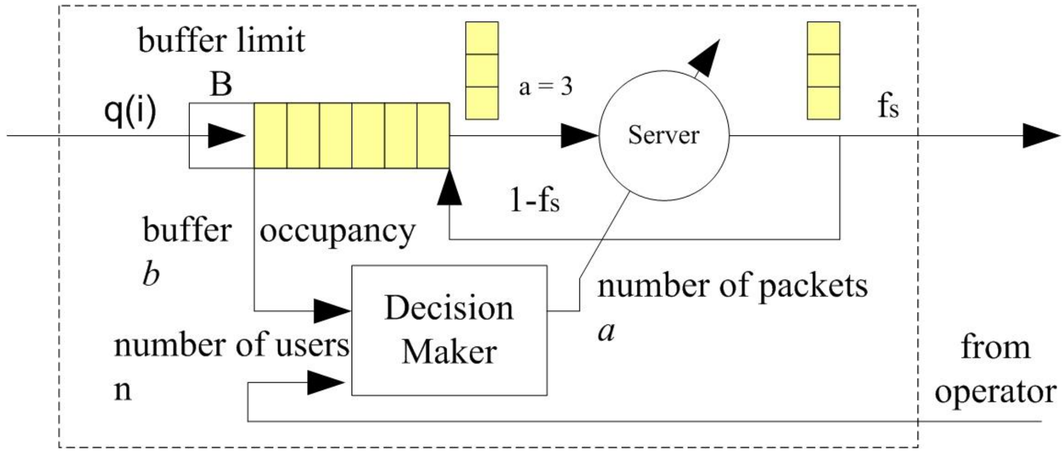

Figure 5.9: System model of the extended problem

account by the decision maker to make decisions. We only provide the description of the competition process here as the other system descriptions are the same as those in Section 5.1. Let $\mathbf{N} = \{N_t\}_{t=0}^{\infty}$ denote the competition process and let n_t denote the number of active data users in the system in frame t. For simplicity, we sometimes use N to denote the population of data users in the system. As the user in consideration remains active for the convenience of our analysis, we thus have $n_t \geq 1$. If $\{N_t\}$ is modelled as a stochastic process, we let $h_{ij} \equiv \Pr\{n_{t+1} = j | n_t = i\}$, $1 \leq i, j \leq N$ denote the transition probability that the number of active data users transits to j in frame $t+1$ when the number of active users is i in frame t. We assume stationary transition probabilities h_{ij} and hence the stochastic process over the number of active users is only dominated by a transition matrix $\mathbf{H} = [h_{ij}]_{N \times N}$. Although a new data user can arrive anywhere within a frame, we assume that a change in the number of active users occurs only at the boundary of a frame. We assume that the Markov chain (given by $\mathbf{H}$) over the number of active users in the system is ergodic and irreducible. If $\{N_t\}$ is modelled as a memoryless process, we let $h(n)$ denote the probability of n active users in a frame, which is assumed to be drawn from a common known probability distribution with support $\{1, 2, \cdots, N\}$.

We now summarize the difference of the MDP formulation between the extended problem and the problem defined in Section 5.1. Now an element in the state space is denoted as $\mathbf{s} = (b, n)$ where b, $0 \leq b \leq B$, is the number of packets queued in the buffer and n, $1 \leq n \leq N$, is the number of active users (including the user in consideration).

Sometimes, we use $(b, \cdot)$ or $(\cdot, n)$ to denote the states with one element being b or n while the other element being arbitrary. The transition probabilities are now given by

$$\mathrm{Tr}((b', n')|(b, n), a) = \begin{cases} q(b')h_{nn'} & b = 0 \\ (1 - f_s)q(b' - b)h_{nn'} + f_s q(b' - b + a)h_{nn'} & b \geq 1, \quad a \leq \min\{A, b\} \\ 0 & otherwise \end{cases} \tag{5.40}$$

Again $q(i < 0) = 0$ and we assume $s' \leq B$. Now the cost structure $C(\mathbf{s}, a)$ (sometimes written as $C(b, n, a)$) is given as $C(\mathbf{s}, a) = C_h(b) + C_c(n) + C_u(a)$, where C_h, C_c and C_u are the holding cost, competition cost and usage cost functions, respectively. Furthermore, C_h, C_c and C_u are assumed to be nonnegative and nondecreasing functions. Again, we may use either the discount optimal criterion or the average optimal criterion to determine the optimal policies. In the following subsections, we provide some qualitative analysis of the extended problem, which are based on the discount optimal criterion. However, as discussed earlier, the analysis is also applicable to the average optimal criterion once we can identify the unichain property.

5.5.3 Characteristic of Value Function

In this subsection, we provide some qualitative analysis which describe some characteristic of the value function $u_\rho(\mathbf{s}) = u_\rho(b, n)$. In particular, we show that $u_\rho(b, n)$ is monotone in each variable when the other is fixed. Note that stationary discount optimal policies for the extended problem exist and further, the value function $u_\rho(\mathbf{s})$ satisfies the optimal equation (2.11) where the states $\mathbf{s} = (b, n)$ and the transition probabilities $\mathrm{Tr}((b', n')|(b, n), a)$ should be used instead.

Proposition 5.5 *$u_\rho(b, n)$ is nondecreasing in b for all n.*

Proof: The proof proceeds by induction.

For $k = 0$, we set $u_\rho^0(b, n) = 0$ for all (b, n) and hence the proposition holds when $k = 0$. For $k = 1$, let $a_{b+1}^{*;1}$ be the optimal action that realizes the minimum part of (2.11) for state $(b + 1, n)$ for all $b \geq 1$. Note that $\mathcal{A}_{(b,n)} \subseteq \mathcal{A}_{(b+1,n)}$ from the action space definition. Consider two cases.

i) $a^{*,1}_{b+1} \in \mathcal{A}_{(b,n)}$. In this case, we have

$$\begin{aligned} u^1_\rho(b+1,n) = &\ C_h(b+1) + C_u(a^{*,1}_{b+1}) + C_c(n) \\ \geq &\ C_h(b) + C_u(a^{*,1}_{b+1}) + C_c(n) \\ \geq &\ \min_{a\in\mathcal{A}_{(b,n)}} \{C_h(b) + C_u(a) + C_c(n)\} = u^1_\rho(b,n) \end{aligned}$$

ii) $a^{*,1}_{b+1} \notin \mathcal{A}_{(b,n)}$. In this case, $a^{*,1}_{b+1} - 1 \in \mathcal{A}_{(b,n)}$ according to our action space definition. Then we have

$$a^{*,1}_{b+1} > a^{*,1}_{b} = \arg\min_{a\in\mathcal{A}_{(b,n)}} \{C_u(a) + C_h(b) + C_c(n)\}$$

and

$$C_h(b+1) + C_u(a^{*,1}_{b+1}) + C_c(n) \geq C_h(b) + C_u(a^{*,1}_{b}) + C_c(n)$$

since C_h and C_u are nondecreasing. Thus we have $u^1_\rho(b+1,n) \geq u^1_\rho(b,n)$ also.

Now assume that the proposition holds for some $k \geq 1$, i.e., $u^k_\rho(b+1,n) \geq u^k_\rho(b,n)$ for all $b \geq 1$, we then show that it also holds for $k+1$. By substituting the general transition probability representation of $\mathrm{Tr}(s'|s,a)$ in (2.14) with the one defined in (5.40), the $(k+1)$th step of the value function can be written as

$$\begin{aligned} &u^{k+1}_\rho(b+1,n) \\ = &\min_{a\in\mathcal{A}_{(b+1,n)}} \Big\{ C_h(b+1) + C_u(a) + C_c(n) + \ \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{b'} q(b'-b-1) u^k_\rho(b',n') \\ &\quad + \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b-1+a) u^k_\rho(b',n') \Big\} \end{aligned} \tag{5.41}$$

for state $(b+1,n)$ and

$$\begin{aligned} &u^{k+1}_\rho(b,n) \\ = &\min_{a\in\mathcal{A}_{(b,n)}} \Big\{ C_h(b) + C_u(a) + C_c(n) + \ \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{b'} q(b'-b) u^k_\rho(b',n') \\ &\quad + \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b+a) u^k_\rho(b',n') \Big\} \end{aligned} \tag{5.42}$$

for state (b,n), respectively. We first compare the fourth term in the minimum part of the two equalities. By the induction hypothesis, we then have

$$\begin{aligned} &\rho(1-f_s)\sum_{n'} h_{nn'} \left[\sum_{b'} q(b'-b-1)u^k_\rho(b',n') - \sum_{b'} q(b'-b)u^k_\rho(b',n')\right] \\ = &\ \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{i=0} q(i) \left[u^k_\rho(b+i+1,n') - u^k_\rho(b+i,n')\right] \\ \geq &\ 0 \end{aligned} \tag{5.43}$$

Let $a_{b+1}^{*,k+1}$ be the optimal action that realizes the minimum part in (5.41). Similar to (5.43), we also have

$$\begin{aligned} & \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b-1+a_{b+1}^{*,k+1}) u_\rho^k(b',n') \\ \geq\ & \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b+a_{b+1}^{*,k+1}) u_\rho^k(b',n') \end{aligned} \tag{5.44}$$

Again, consider two cases.

i) $a_{b+1}^{*,k+1} \in \mathcal{A}_{(b,n)}$. We have

$$\begin{aligned} & u_\rho^{k+1}(b+1,n) \\ = \ & C_h(b+1) + C_u(a_{b+1}^{*,k+1}) + C_c(n) + \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{b'} q(b'-b-1) u_\rho^k(b',n') \\ & + \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b-1+a_{b+1}^{*,k+1}) u_\rho^k(b',n') \\ \geq \ & C_h(b) + C_u(a_{b+1}^{*,k+1}) + C_c(n) + \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{b'} q(b'-b) u_\rho^k(b',n') \\ & + \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b+a_{b+1}^{*,k+1}) u_\rho^k(b',n') \\ \geq \ & \min_{a\in\mathcal{A}_{(b,n)}} \Big\{ C_h(b) + C_u(a) + C_c(n) + \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{b'} q(b'-b) u_\rho^k(b',n') \\ & \qquad + \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b+a) u_\rho^k(b',n') \Big\} \\ = \ & u_\rho^{k+1}(b,n) \end{aligned}$$

The first inequality follows from the nondecreasing of C_h, (5.43) and (5.44).

ii) $a_{b+1}^{*,k+1} \notin \mathcal{A}_{(b,n)}$. Recall that the action space $\mathcal{A}_{(b+1,n)} = \{1,\cdots,\min\{A,b+1\}\}$ and $\mathcal{A}_{(b,n)} = \{1,\cdots,\min\{A,b\}\}$. Thus if $a_{b+1}^{*,k+1} \notin \mathcal{A}_{(b,n)}$, let $\hat{a}_{b+1}^{k+1} = a_{b+1}^{*,k+1} - 1$ and hence $\hat{a}_{b+1}^{k+1} \in \mathcal{A}_{(b,n)}$. We then have

$$\begin{aligned} & u_\rho^{k+1}(b+1,n) \\ = \ & C_h(b+1) + C_u(a_{b+1}^{*,k+1}) + C_c(n) + \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{b'} q(b'-b-1) u_\rho^k(b',n') \\ & + \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b-1+a_{b+1}^{*,k+1}) u_\rho^k(b',n') \\ \geq \ & C_h(b) + C_u(\hat{a}_{b+1}^{k+1}) + C_c(n) + \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{b'} q(b'-b) u_\rho^k(b',n') \\ & + \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b+\hat{a}_{b+1}^{k+1}) u_\rho^k(b',n') \\ \geq \ & \min_{a\in\mathcal{A}_{(b,n)}} \Big\{ C_h(b) + C_u(a) + C_c(n) + \rho(1-f_s)\sum_{n'} h_{nn'} \sum_{b'} q(b'-b) u_\rho^k(b',n') \\ & \qquad + \rho f_s \sum_{n'} h_{nn'} \sum_{b'} q(b'-b+a) u_\rho^k(b',n') \Big\} \\ = \ & u_\rho^{k+1}(b,n) \end{aligned}$$

This first inequality follows from the nondecreasing C_h and C_u and (5.43).

Therefore, we have $u_\rho^k(b+1,n) \geq u_\rho^k(b,n)$ for all $k \geq 0$ by the induction argument. Then the proposition follows by letting $k \to \infty$, i.e., $u(b,n) = \lim_{k\to\infty} u^k(b,n)$ is nondecreasing in b for fixed n. ■

Proposition 5.5 states that a large buffer occupancy is undesirable as this increases the holding cost and then may increase the value of a policy. Thus the sender needs to consider using higher transmission rates that can optimally balance between the holding cost and the usage cost, when the buffer occupancy increases. Given an additional assumption, the following proposition states that a competitive situation is also undesirable as it may increase the competition costs, i.e, it may increase the power consumption when the number of active users increases provided other situations remain unchanged.

We have introduced and applied the concept of *stochastic orders* to compare two stochastic processes (see Section 5.3.3 and Section 5.4.2). Now we consider the stochastic orders for two probability vectors $\mathbf{p} = (p_i)_{i=1}^N$, $\sum_i^N p_i = 1$ and $\mathbf{p}' = (p'_i)_{i=1}^N$, $\sum_i^N p'_i = 1$. $\mathbf{p}$ is said to be stochastically smaller than $\mathbf{p}'$, denoted as $\mathbf{p} \leq_{st} \mathbf{p}'$, if and only if $\sum_{i=k}^N p_k \leq \sum_{i=k}^N p'_k$, $k = 1, 2, \cdots, N$. The vector $\mathbf{p}$ is strictly dominated by $\mathbf{p}'$ if and only if strictly inequality holds for $k = 2, 3, \cdots, N$. For a nondecreasing real-valued sequence $\{v_i\}$, $v_i \leq v_{i+1}$, if $\mathbf{p} \leq_{st} \mathbf{p}'$ then $\sum_{i=1}^N v_i p_i \leq \sum_{i=1}^N v_i p'_i$ (see Lemma 4.7.2 in Puterman [62]). A transition probability matrix can be considered as a column of probability row vectors. A probability transition matrix is said to be *monotone* if each row vector is stochastically smaller than all the row vector(s) below it (see [42] for more discussion). For example, consider the transition matrix $\mathbf{H}$ and let $\mathbf{H}_i = (h_{i1}, \cdots, h_{iN})$, $i = 1, 2, \cdots, N$, denote its row vectors. If $\mathbf{H}$ is a monotone matrix, we have $\mathbf{H}_{i-1} \leq_{st} \mathbf{H}_i$, $i = 2, \cdots, N$. It is not difficult to find many analytical models whose transition matrix $\mathbf{H}$ is monotone. For example, for a $M/M/n$ queueing system with finite population, its transition matrix is monotone. Suppose that the transition matrix $\mathbf{H}$ is monotone, we have the following result, which also states that a highly competitive situation is undesirable as this increases the competition cost.

Proposition 5.6 *Suppose that the transition matrix of the competition process is monotone, i.e.,* $\mathbf{H}$ *is monotone. Then* $u_\rho(b,n)$ *is nondecreasing in* n *for all* b.

Proof: The proof proceeds by induction. Note that for fixed b, the action space is the same for all states $(b,\cdot)$. Hence we omit the action space in the proof.

For $k=0$, we choose $u_\rho^0(b,n)=0$ for all (b,n). We first show that the proposition holds for $k=1$. Let $a_{n+1}^{*,1}$ be the optimal action that realizes the minimum part of (2.11) for state $(b,n+1)$. we have

$$\begin{aligned} u_\rho^1(b,n+1) = &\; C(b)+C_c(n+1)+C_u(a_{n+1}^{*,1}) \\ \geq &\; C(b)+C_c(n)+C_u(a_{n+1}^{*,1}) \\ \geq &\; \min_a\{C_h(b)+C_c(n)+C_u(a)\} = u_\rho^1(b,n) \end{aligned}$$

The first inequality follows from the nondecreasing C_c.

Assume that the proposition holds for $k \geq 1$, i.e., $u_\rho^k(b,n) \leq u_\rho^k(b,n+1)$ for all $1 \leq n \leq N-1$. We now show that it is also true for $k+1$. By substituting the general transition probability representation of $\mathrm{Tr}(s'|s,a)$ in (2.14) with the one defined in (5.40), the $(k+1)$th step of the value function can be written as

$$\begin{aligned} & u_\rho^{k+1}(b,n+1) \\ = & \min_a\Big\{C_h(b)+C_c(n+1)+C_u(a)+\rho(1-f_s)\sum_{b'}q(b'-b)\sum_{n'}h_{n+1,n'}u_\rho^k(b',n') \\ & \quad +\rho f_s\sum_{b'}q(b'-b+a)\sum_{n'}h_{n+1,n'}u_\rho^k(b',n')\Big\} \end{aligned} \tag{5.45}$$

for state $(b,n+1)$. Let $a_{n+1}^{*,k+1}$ be the optimal action that realizes the minimum in (5.45) and we have

$$\begin{aligned} & u_\rho^{k+1}(b,n+1) \\ = \; & C_h(b)+C_c(n+1)+C_u(a_{n+1}^{*,k+1})+\rho(1-f_s)\sum_{b'}q(b'-b)\sum_{n'}h_{n+1,n'}u_\rho^k(b',n') \\ & +\rho f_s\sum_{b'}q(b'-b+a_{n+1}^{*,k+1})\sum_{n'}h_{n+1,n'}u_\rho^k(b',n') \\ \geq \; & C_h(b)+C_c(n)+C_u(a_{n+1}^{*,k+1})+\rho(1-f_s)\sum_{b'}q(b'-b)\sum_{n'}h_{n,n'}u_\rho^k(b',n') \\ & +\rho f_s\sum_{b'}q(b'-b+a_{n+1}^{*,k+1})\sum_{n'}h_{n,n'}u_\rho^k(b',n') \\ \geq \; & \min_a\Big\{C_h(b)+C_c(n)+C_u(a)+\rho(1-f_s)\sum_{b'}q(b'-b)\sum_{n'}h_{n,n'}u_\rho^k(b',n') \\ & \quad +\rho f_s\sum_{b'}q(b'-b+a)\sum_{n'}h_{n,n'}u_\rho^k(b',n')\Big\} \\ = \; & u_\rho^{k+1}(b,n) \end{aligned}$$

The first inequality is from the fact that C_c is nondecreasing in n, $q(\cdot) \geq 0$, $\mathbf{H}_{n-1} \leq_{st} \mathbf{H}_n$ (proposition assumption) and $u_\rho^k(b,n) \leq u_\rho^k(b,n+1)$ (induction hypothesis).

Therefore, we have $u_\rho^{k+1}(b,n+1) \geq u_\rho^{k+1}(b,n)$ for all $k \geq 0$ by the induction argument. Then the proposition follows by letting $k \to \infty$, i.e., $u_\rho(b,n) = \lim_{k\to\infty} u_\rho^k(b,n)$ is nondecreasing in n for fixed b. ■

5.5.4 Property of Optimal Policies

In this subsection, we provide some qualitative analysis which describes the property of the optimal policies. In particular, we prove that the optimal policies are also nondecreasing in the buffer occupancy. Again, the monotonicity is also built on showing that $u_\rho(b,n)$ is a convex function of b based on the induction argument. The following lemma and proposition summarize the monotonicity property of the optimal policies.

Lemma 5.6 *Assumption 5.1 holds and further Assume the buffer is infinite, i.e., $b \in \{0,1,\cdots\}$, the value function $u_\rho(b,n)$ is a convex function of b for all n.*

Proposition 5.7 *Assume that buffer overflow does not occur. Then the stationary discount optimal policy of the extended rate allocation problem is monotonically increasing in the buffer occupancy. Mathematically, this means that*

$$d^*(b+1,n) \geq d^*(b,n), \quad b \geq 1 \tag{5.46}$$

Their proofs are similar to that of Lemma 5.2 and Proposition 5.1, respectively, and hence we omit them here. Again, the explanation behind Proposition 5.5 is that a large transmission rate should be chosen to deplete the buffer (quickly) and to avoid the increasing holding cost, when the buffer occupancy increases.

We have not been able to prove that the same monotone property holds in n direction for the optimal actions. However, when the competition process is degenerated to a memoryless process, i.e., the probability of n active users in a frame is given as $h(n)$ ($\sum_n h(n) = 1$) which is independent of the frame index, we have an interesting result indicating the optimal actions are insensitive to the number of users in the cell. An intuitive explanation could be that since a user cannot predict whether the number of active user would increase in the next frame or not, the best response for a user is not to change its decisions and to use the state-action pair that is determined by the buffer occupancy. The result is summarized in the following proposition.

Proposition 5.8 *Assuming the competition process $\{N_t\}$ is memoryless. Then the stationary discount optimal policy of the extended rate allocation problem is independent of the number of active users. Mathematically, this means $d^*(b, n+1) = d^*(b, n)$ for all b and n if $\{N_t\}$ is memoryless.*

Proof: The proof proceeds by contradiction.

Let a^*_{n+1} and a^*_n be the optimal actions for states $(b, n+1)$ and (b, n), respectively. Assuming $a^*_{n+1} < a^*_n$, we show that the assumption leads to a contradiction. As $\{N_t\}$ is memoryless, the discount optimal equations (2.11) can be written as:

$$\begin{aligned} & u_\rho(b, n) \\ = & \min_{a \in \mathcal{A}_{(b,n)}} \Big\{ C_h(b) + C_c(n) + C_u(a) + \rho(1 - f_s) \sum_{n'} h(n') \sum_i q(i) u_\rho(b + i, n') \\ & \qquad + \rho f_s \sum_{n'} h(n') \sum_i q(i) u_\rho(b + i - a, n') \Big\} \end{aligned} \tag{5.47}$$

As the optimal actions realize the minimum part of the right hand side of (5.47), then for state (b, n), we have:

$$\begin{aligned} & C_u(a^*_n) + \rho f_s \sum_{n'} h(n') \sum_i q(i) u_\rho(b + i - a^*_n, n') \\ < \; & C_u(a^*_{n+1}) + \rho f_s \sum_{n'} h(n') \sum_i q(i) u_\rho(b + i - a^*_{n+1}, n') \end{aligned} \tag{5.48}$$

Note that the strict inequality holds since we always break ties by choosing the smallest action when choosing an optimal action. Similarly, for state $(b, n+1)$, we have

$$\begin{aligned} & C_u(a^*_{n+1}) + \rho f_s \sum_{n'} h(n') \sum_i q(i) u_\rho(b + i - a^*_{n+1}, n') \\ \leq \; & C_u(a^*_n) + \rho f_s \sum_{n'} h(n') \sum_i q(i) u_\rho(b + i - a^*_n, n') \end{aligned} \tag{5.49}$$

Combing (5.48) and (5.49), we have $0 < 0$. Thus the assumption is not true and we have $a^*_{n+1} \leq a^*_n$. Now again we assume that $a^*_{n+1} > a^*_n$. With the similar argument to above, we can show that such assumption is not true also and we have $a^*_{n+1} \geq a^*_n$. Thus the arguments above imply that $a^*_{n+1} = a^*_n$ which is the desired result. ■

5.6 Summary

In this chapter, we have studied the rate allocation for a single data user as a Markov decision problem. The property of optimal policies have been discussed and a case study with numerical examples has been presented. We have also proposed and analyzed a class of simple policies. In the last part of this chapter, we have extended the rate control problem for an isolated single user to one in the presence of multiple users. The competition across users has been represented by a competition process over the number of active users in the extended problem formulation. The characteristic of value function and the property of optimal policies for the extended problem has also been studied.

We note that the optimization problems discussed so far are from a single user's point of view. Another optimization problem from the point of view of the whole system of multiple users is also important to wireless communications. We study a system level optimization problem in the next chapter.

Chapter 6

Fair-effort Based Resource Allocation

In the previous chapters, we have studied some resource management mechanisms, i.e., transmission scheduling, power control and rate allocation, from a single user's point of view. In this chapter, we consider the resource allocation problem from the viewpoint of an operator who allocates the system resources among multiple users. The operator should take both fairness among users and system utilization efficiency into account when designing an allocation policy. In this chapter, we first propose a new fairness criterion and a novel fair-effort fairness model. We then present a simple credit based algorithm to approximate the proposed fairness model. Furthermore, based on our fairness model, we provide a detailed packet level resource allocation scheme for a CDMA based network which supports variable instantaneous data rate allocation. Finally, some numerical results are provided at the end of the chapter.

6.1 Problem Description

We consider a single cell in a hybrid CDMA/TDMA network in this chapter. We focus on the resource allocation problem for only data services in a synchronized uplink channel but similar arguments apply to the downlink. In such an uplink channel, mobile users send their transmission requests to a centralized controller (the base station). The controller schedules which user(s) should transmit in the next frame, and it then decides

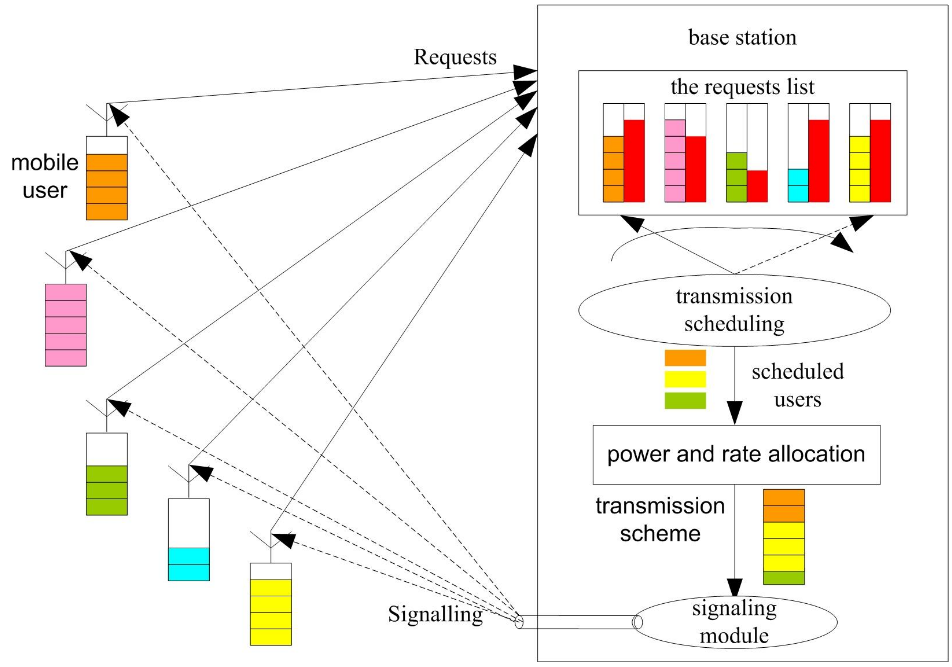

Figure 6.1: System model

the transmission power and the transmission rate for those scheduled user(s). We focus on the policy used by the controller, which consists of the above series of actions. The detailed protocols and the formats for the requests and related signallings are beyond the scope of this thesis (some related works on protocols design can be found in [2, 45]). Fig. 6.1 illustrates the system model of our resource allocation problem. Mobile users send transmission requests and report some related information such as their channel conditions and buffer occupancies. The base station schedules a transmission scheme consisting of the scheduled users and their allocated transmission powers and rates in each frame to all users and the scheduled users transmit according to such a scheme.

In the time domain, transmissions and resource allocations are assumed to be synchronized on a frame by frame basis. In the code domain, a user may use different codes for spreading and thus may have difference instantaneous data rates. Fig. 6.2 illustrates the transmission model. The size of a pipe roughly represents the instantaneous data rate of a flow in a frame. In the time domain, transmission scheduling can be used for time multiplexing and hence a flow may not necessarily transmit in every frame. This is represented with the non-continuous pipe in Fig. 6.2. In the code domain, the number of users that can transmit simultaneously in a frame is not fixed

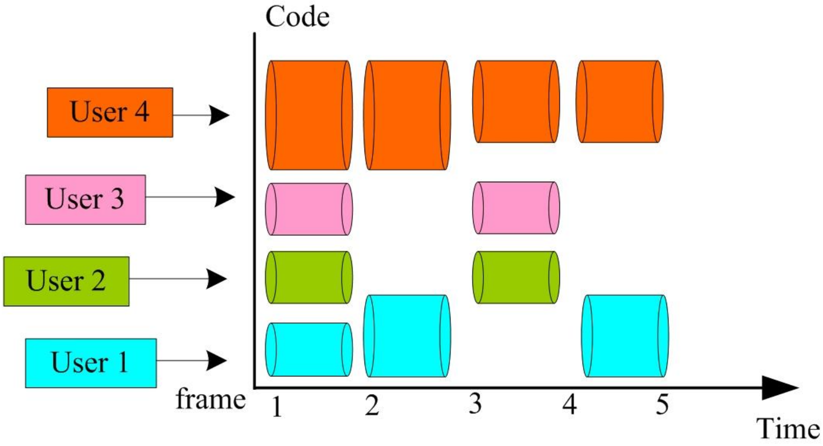

Figure 6.2: A transmission example in a synchronized CDMA channel

as this is influenced by the interference limited system capacity.

An optimal (and ideal) power allocation scheme can guarantee that all scheduled users experience the same frame success probability, which will be introduced later. The transmission scheduling and the rate allocation together determine the (effective) average throughput of a backlogged user. We can use the average total instantaneous throughput that is summed over all scheduled flows to measure the system utilization efficiency. One calibration for the fairness among users is in terms of the difference between the *weighted* average throughput of backlogged flows, which is also the widely referred fairness measure. S. Oh and K. Wasserman [59] have shown that the average total instantaneous throughput can be maximized by allocating larger instantaneous data rates to those users with relatively better channel conditions and allocating smaller instantaneous data rates to those users with relatively poorer channel conditions. However, this allocation scheme may introduce unfairness among users and some users with relatively poorer channel conditions may be starved of service for a long time. This is also undesirable. Such a tradeoff between fairness and efficiency motivates us to propose a new fairness model which incorporates the time varying channel conditions as a factor impacting on the quota of resources allocated to a user. We discuss our fairness model and compare it to the classical generalized processor sharing (GPS) fairness model in the next section.

6.2 Fair-effort Resource Sharing

6.2.1 The Fair-effort Resource Sharing Model

Consider a single node with multiple backlogged flows waiting for transmission through a common link of capacity C. In the GPS model, all flows are assumed as fluid flows with fixed nominal weights and the link capacity C is often considered as a constant. During any infinitesimally small time interval Δt, a backlogged flow i with a fixed weight ω_i should be allocated a link capacity of $C \cdot \Delta t \cdot (\omega_i / \sum_{j \in \mathcal{B}(\Delta t)} \omega_j)$, where $\mathcal{B}(\Delta t)$ is the set of backlogged flows during Δt. If a flow is scheduled to transmit, its head of line packet will be transmitted with the (fixed) link transmission rate. Hence the link capacity is shared among backlogged flows in a time-multiplexed way via the packet scheduling. Another mathematical formulation of the GPS is as follows:

$$\left| \frac{W_i(t_1, t_2)}{\omega_i} - \frac{W_j(t_1, t_2)}{\omega_j} \right| = 0. \tag{6.1}$$

Eq. (6.1) allows that the resource granted to any two flows $W(t_1, t_2)$ in a time interval $[t_1, t_2)$ during which they are continuously backlogged are proportional to their nominal weights ω. The GPS model makes two implicit assumptions when it is used as the fairness reference in communications. First, the resource is a *public resource* that is allocated to some competing users through an allocation process. Second, users participate *passively* in the resource allocation process in that they cannot impact on the total amount of public resources they receive, and they cannot change their resource shares without changing their nominal weights. We contend that the second implicit assumption is not universally appropriate, especially for wireless systems where channel interference is not uncommon. For example, consider the uplink in a CDMA-based network. More than one user may transmit simultaneously with the use of the spreading technique. Suppose two users have the same instantaneous data rates and BER requirements. Then their received powers at the base station should be kept equal with optimal power allocation [3, 68]. That is, if p_i and h_i are the transmission power and path gain of user $i = 1, 2$ respectively, $p_1 h_1 = p_2 h_2$. However these users may have different transmission powers due to their different path gains. For example, $p_1 > p_2$ if $h_1 < h_2$. Since the system capacity of a CDMA network is interference limited, these users can influence the capacity of the system by varying their transmission powers. Thus it is possible for the users to interact *actively* in the resource allocation process in that their transmission powers affect the amount of available public resources that can be allocated. Note that in this example, the two users get their fair share of public resources under the GPS model in that they have the same transmission rate.

The proposed Fair-Effort Sharing (FES) model is defined as:

$$\left| \frac{W_i(t_1, t_2)}{e_i(t_1, t_2)} - \frac{W_j(t_1, t_2)}{e_j(t_1, t_2)} \right| = 0 \tag{6.2}$$

In (6.2), $e(t_1, t_2)$ quantifies the effort a user puts in to get $W(t_1, t_2)$ amount of the public resource during the time interval $[t_1, t_2)$. We require $e(t_1, t_2) > 0$ for any backlogged flow during the time interval $[t_1, t_2)$. Compared with the GPS model, the main difference is the use of a time-dependent measure of effort instead of a fixed nominal weight. This simple modification allows the incorporation of the (possible) interactions between the competing users and the resource process. Thus, if the nominal weight of a user can be dynamically changed during the resource allocation process, the GPS model evolves to the FES model.

Recently, F. Kelly [43, 44] has proposed a new fairness concept, *proportional fairness*, for elastic traffic. Each data user is assigned a *utility* which is a function of the allocated resources, e.g., the transmission rate. Furthermore, each user bids its resource quota through its *willingness-to-pay* rate. An allocated resource vector $\vec{x} = (x_1, x_2, \cdots)$ is said to be *proportional fair* if it is feasible (that is $x_i \geq 0$ and $\sum_i x_i \leq C$) and if for any other feasible allocation vector $\vec{x}'$, the aggregate proportional changes is zero or negative:

$$\sum_i \frac{x_i' - x_i}{x_i} \leq 0. \tag{6.3}$$

Kelly has proved that there exists a willingness-to-pay vector and a resource allocation vector to maximize both the system total utility that is summed up over all users and the *net utility* of each user which is the utility minus its willingness-to-pay, when the system is in a long-term equilibrium (see Theorem 1 in [43]). Furthermore, Kelly has shown that at equilibrium, the resource vector is proportional fair if the utility function is a logarithmic function. Though Kelly has considered that users may actively participate in the resource allocation process (through their willingness-to-pay), he still assumes that the available resources are constant. We note that our FES model can also achieve the proportional fairness in the long-term equilibrium if each user adjusts its effort (its willingness-to-pay for example) to maximize its net utility when receiving resources. Furthermore, Kelly does not address the procedure of how to achieve proportional fairness. However, we provide not only the abstract fairness model but also the implementation reference.

As with the GPS model, the FES model only serves as an abstract reference scheduler. The measure of effort can be defined differently for different applications. Also, the FES model is based on the fluid flow assumption. Let $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$ be a series of partitions over a time interval $[0, T)$. We can then use the following discrete version to approximate the fluid version of the FES model.

$$\left| \sum_{n=0}^{N-1} \frac{W_i(t_n, t_{n+1})}{e_i(t_n)} - \sum_{n=0}^{N-1} \frac{W_j(t_n, t_{n+1})}{e_j(t_n)} \right| = 0 \tag{6.4}$$

In (6.4), $e_i(t_n)$ quantifies the effort at the time instant t_n which is used to approximate $e_i(t_n, t_{n+1})$ during the whole time interval $[t_n, t_{n+1})$.

6.2.2 A Fair-Effort Crediting Algorithm

We now present a general algorithm, called the fair-effort crediting algorithm (FECA), to implement the FES model. Without loss of generality, we consider a wireless system where (1) a central base station allocates resources only among known backlogged flows, (2) all flows are kept in separate queues and (3) transmissions are synchronous and time-aligned on a frame-by-frame basis. Note that these are according with our problem context, i.e., the system model and the transmission model in Section 6.1.

The base station maintains a sorted list $\mathcal{L} = \{\vec{l_1}, \vec{l_2}, \cdots\}$ where $\vec{l}$ is a tuple (i, σ_i, b_i, e_i) of real values. i is the unique identification of a flow, σ_i is a counter which stores the credits of flow i, b_i is the record of the flow length (i.e., amount of data to transmit) and e_i is the record of the recently updated measure of effort of flow i. Suppose each flow has a virtual *reference flow* which always gets its fair-effort share of resources. σ_i indicates the difference in flow i's credits relative to its virtual reference flow, i.e., $\sigma_i > 0$ ($\sigma_i < 0$) means that flow i is *lagging* (*leading*) its reference flow. σ_i is initialized/reset to zero when the flow i buffer becomes backlogged/empty. The measure of effort e_i can be computed from some effort-labelling functions. In a frame t, FECA operates as follows.

The fair-effort crediting algorithm (FECA):

FECA-1: (Sort the list) The list $\mathcal{L}$ is sorted according to the credits of the entries, from high to low. If two entries have the same credits, they are sorted according to their efforts, from high to low. If a flow i has a zero b_i (i.e., empty flow), its corresponding entry is removed from the list. That is, $\mathcal{L} = \mathcal{L} \setminus \{\vec{l_i}\}$ if $b_i(t) = 0$.

FECA-2: (Calculate the total effort, $E(t)$) The total effort is the sum of the effort records of all backlogged flows, i.e., $E(t) = \sum_{i \in \mathcal{L}} e_i(t)$.

FECA-3: (Schedule and allocate resources) Let $\mathcal{M}(t) = \{m_1, m_2, \cdots, m_t\}$ denote the set of scheduled flows. We allocate resources as denoted by $W_{m_1}(t, t+1)$, $W_{m_2}(t, t+1)$, $\cdots$, $W_{m_t}(t, t+1)$ to these flows. We require that $\sigma_{m_1} \geq \sigma_{m_2} \geq \cdots \geq \sigma_{m_t}$. This means that the lagging most flow is considered first.

FECA-4: (Compute the total allocated resources, $W(t, t+1)$) The total allocated resources is the sum from all allocated flows, i.e., $W(t, t+1) = \sum_{i \in \mathcal{M}} W_i(t, t+1)$.

FECA-5: (Update the credits) All entries in the list have their credits updated as follows:

$$\sigma_i(t+1) = \begin{cases} \sigma_i(t) + \frac{e_i(t)}{E(t)} W(t, t+1) & i \notin \mathcal{M}(t) \\ \sigma_i(t) + \frac{e_i(t)}{E(t)} W(t, t+1) - W_i(t, t+1) & i \in \mathcal{M}(t) \end{cases} \tag{6.5}$$

In step FECA-5, $\frac{e_i(t)}{E(t)}$ $W(t, t+1)$ is the fair-effort share quota of flow i which is proportional to its effort, and $W_i(t, t+1)$ is its actual allocated resource. In each frame, if the actual allocated resources for each backlogged flow equals its fair effort share, i.e., $W_i(t, t+1) = \frac{e_i(t)}{E(t)}$ $W(t, t+1)$, then such an allocation can be easily proved to satisfy the discrete version of the FES model. We define the following convergence measure to measure how fast (on average) a flow is leading/lagging its fair effort share. Suppose flow i is continuously backlogged during the time interval $[0, T)$. Let $S_i(T)$ be the total amount of resources it receives during $[0, T)$:

$$S_i(T) = \sum_{t=0}^{T-1} W_i(t, t+1) \tag{6.6}$$

Then $S_i(T)/T$ is the average throughput of flow i. Let $F_i(T)$ be the system resources that should be allocated to flow i according to the FES model during $[0, T)$. It is given by:

$$F_i(T) = \sum_{t=0}^{T-1} \frac{e_i(t)}{E(t)} W(t, t+1) \tag{6.7}$$

The convergence rate $\delta_i(T)$ of flow i which is continuously backlogged in $[0, T)$ is then:

$$\delta_i(T) = |F_i(T) - S_i(T)|/F_i(T) \tag{6.8}$$

6.3 A Resource Allocation Scheme

In this section, we present a detailed packet level resource allocation scheme which consists of transmission scheduling, power allocation and rate allocation for a CDMA uplink channel supporting multiple instantaneous data rates. Power allocation determines the transmission quality while transmission scheduling together with rate allocation determine both the system achievable average total throughput and the average throughput of each flow. We use the system achievable average total throughput as the system resources to be allocated to backlogged users, however, the resource allocation policy is based on our fair-effort model.

6.3.1 Optimal Power Allocation

A centralized optimal power allocation scheme has been proposed in [3, 68]. In a CDMA-based mobile network, the transmission power determines the received *bit-energy-to-interference-power-spectral-density*, E_b/I_0, and further determines the experienced bit error rate (BER) at the receiver. Consider an uplink channel in a CDMA-based network. The received E_b/I_0 can be computed as:

$$\left(\frac{E_b}{I_0}\right)_i = \frac{W}{R_i} \frac{P_i}{\sum_{j=1, j\neq i}^{N(t)} P_j + \eta_0} \tag{6.9}$$

where W is the system spreading bandwidth, R_i is the data rate of flow i, P_i is the received power of flow i at the base station, $N(t)$ is the number of flows that transmit in frame t and η_0 is the background noise spectral density at the base station. To guarantee a minimum BER requirement, the received E_b/I_0 should satisfy

$$\left(\frac{E_b}{I_0}\right)_i \geq \gamma_i^*, \tag{6.10}$$

where γ_i^* is the target E_b/I_0 corresponding to the minimum BER requirement given a particular allocated transmission rate R_i for flow i. When (6.10) holds with equality and from (6.9), the base station can allocate the received power level to each scheduled flow according to [3, 68]

Optimal Power Allocation:

$$P_i = \frac{g_i \eta_0 W}{1 - \Sigma_{j=1}^{N(t)} g_j}, \tag{6.11}$$

where g_i is the *power index*

$$g_i = \frac{\gamma_i^*}{\gamma_i^* + W/R_i}. \tag{6.12}$$

Eq. (6.11) is called optimal power allocation in the sense that the BER requirements of all scheduled flows are met with equality. Such a power allocation also minimizes the total received power, i.e., $\sum_i P_i$ is minimized. Also note that (6.11) is an idealized or perfect power allocation scheme. It assumes that the base station has the up-to-date information of each mobile user and it also assumes that the path gain of each mobile remains constant in a frame. The effect of imperfect power allocation for voice services has been investigated by D. Zhao [100] in a multi-code CDMA network. The effect of imperfect power allocation for data services is beyond the scope of this thesis and is suggested as a possible extension. We assume that path gain information of all mobile users are known to the base station instantaneously. To guarantee feasible solutions, we require $\sum_{i=1}^{N(t)} g_i < 1$. Since the uplink transmission power is provided by a mobile's battery, another limitation related to power allocation is that the allocated power should not be greater than the peak (maximum) transmission power limit p_i^{max}, that is,

Peak Power Limit:

$$P_i \leq h_i p_i^{max} \tag{6.13}$$

where h_i is the path gain of flow i. Thus to achieve a feasible solution, a conservative constraint is needed and the aggregate power index in each frame needs to be kept less than a threshold $1 - \Delta$, i.e.,

Power Capacity Limit:

$$\sum_{j=1}^{N(t)} g_j < 1 - \Delta \tag{6.14}$$

In (6.14), Δ is the reserved power index used to limit the assigned power levels. An example for setting the value of Δ is given by M. Arad [2] where $\Delta = \eta_0 W \max_i\{\frac{g_i}{h_i p_i^{max}}\}$. Another way to avoid infeasible power allocation is by using a more conservative admission control algorithm. (6.14) can also be viewed as the power capacity of each frame.

We use the average frame success probability f_s to set the target E_b/I_0 in each frame since the minimum BER requirement has a one-to-one mapping to the average frame success probability if the modulation, channel coding and other physical layer techniques are the same in all frames. However, as the instantaneous rates are different, the same minimum BER requirement for a flow may correspond to different f_s values. For example, assuming no coding scheme, perfect bit error detection and no error correction, f_s can be computed as:

$$f_s = (1 - BER)^{L_{R_i}} \tag{6.15}$$

where L_{R_i} is the number of bits per frame given rate R_i is allocated. The target E_b/I_0 can be set fixed regardless of the instantaneous data rate, which will result in different f_s values when the transmission rates are different. The target E_b/I_0 can also be set based on a fixed f_s strategy and different target E_b/I_0 can be derived for different instantaneous data rates accordingly. We use the fixed f_s strategy to set the target E_b/I_0 for all flows.

6.3.2 Transmission Scheduling and Rate Allocation

Transmission scheduling determines which flow(s) to transmit in the next frame and rate allocation prescribes the instantaneous rate (the spreading code) for each scheduled flow. Transmission scheduling together with rate allocation determine the achievable total instantaneous data rate in a frame. They also determine the average throughput of each backlogged flow. As our resource allocation scheme uses the FES model as our fairness reference, we use the FECA algorithm in Section 6.2.2 for transmission scheduling (including the queueing strategy), i.e., in each frame, the flows are scheduled based on their effort credits. To complete the resource allocation scheme, we still need to clarify two other aspects. One is how to compute the effort and the other is the instantaneous rate allocation algorithm.

We use path gain as a measure of the effort expended by a flow in order to receive its resource quota since the achievable system total throughput may be dependent on the users' channel conditions. However, we do not need to use the path gain values directly but can partition the range of path gains into a finite set of intervals. We then use a numeric score to represent a path gain interval and the better the path gain the higher the score. Fig. 6.3 plots some example scoring functions.

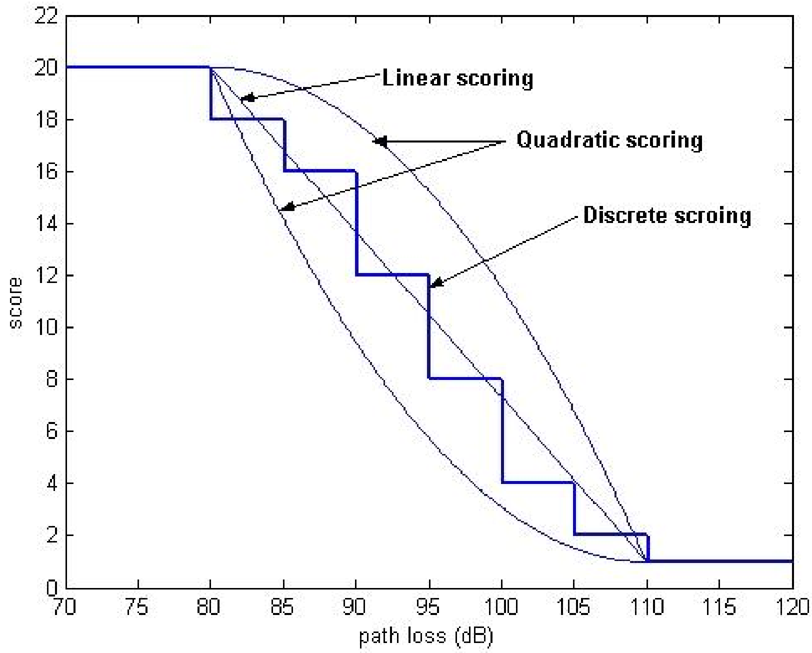

Figure 6.3: Different path gain scoring functions

We use an exhaustive instantaneous rate allocation algorithm as it helps to improve resource utilization efficiency (in a saturated network). Note that different instantaneous data rates can be easily realized by using different spreading codes. In real systems, the available spreading codes are finite and hence the available instantaneous data rates are also finite. Let the set of available instantaneous data rates be denoted as $\mathcal{R} = \{R^1, R^2, \cdots, R^K\}$ with R^1 being the minimum and R^K being the maximum rate, respectively, and $\gamma = \{\gamma^1, \gamma^2, \cdots, \gamma^K\}$ being the corresponding target E_b/I_0. The algorithm is briefly described as follows.

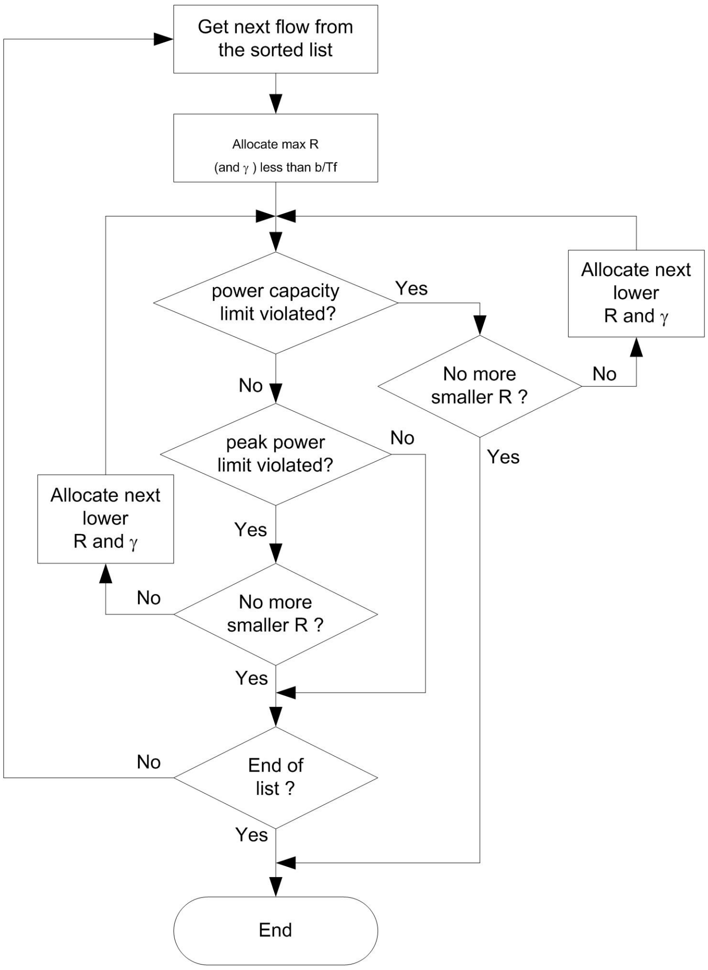

Figure 6.4: Flow chart of the instantaneous data rate allocation algorithm

All backlogged flows are sorted before the rate allocation process according to their effort credits, cf. Section 6.2.2. Let $\mathcal{L}$ denote such a sorted list and let L denote the number of backlogged flows. The allocation of instantaneous data rates to backlogged flows is performed one by one starting from the head flow in the sorted list as follows. The flow is first allocated the highest instantaneous data rate R (and corresponding γ) less than b/T_f, where b is the length of the flow and T_f is the frame length. Conformance to the power capacity limit (6.14) is then checked. If this is satisfied, conformance to the peak power limit (6.13) is then checked. If this is also satisfied, the next flow in the list is scheduled and allocated its instantaneous data rate. On the other hand, if either of the above checks fails, the next lower instantaneous data rate and corresponding E_b/I_0 are allocated and the above checks for conformance of (6.14) and (6.13) are repeated. If the flow in consideration has been allocated the lowest instantaneous data rate but the conformance to the power capacity (6.14) fails, the allocation procedure stops. If the flow in consideration has been allocated the lowest instantaneous data rate but the conformance to the peak power limit (6.13) fails, the flow in consideration is then skipped. These steps are repeated until all flows in the list have been considered or until it is not possible to allocate a flow any rate in the set of available instantaneous data rates. The flow chart in Fig. 6.4 shows this allocation procedure.

The algorithm is called exhaustive in that it tries to schedule as many flows as possible to transmit in each frame until the power capacity limitation is violated or until all backlogged flows have been considered. When the limitations are violated, the algorithm tries to reduce the instantaneous data rate to be allocated to a flow before skipping the flow.

6.4 Numerical Examples

In this section, we provide some simulation results to compare the performance of our resource allocation scheme with that of some other schemes. The simulations are to explore the relationship between the system utilization efficiency and the fairness among users and to investigate the average delay performance of different schemes.

We use *Scheme-A* to represent our allocation scheme described in Section 6.3. When we only change the transmission scheduling strategy while keeping the optimal power allocation and the exhaustive instantaneous data rate allocation, we readily have two other schemes, labelled as *Scheme-B* and *Scheme-C*. Scheme-B uses a simple round robin transmission scheduling strategy which is independent of the path gains of backlogged flows and provides the GPS model based fairness among users. Scheme-C uses a *biased* transmission scheduling strategy which schedules the flows in the order of their path gains, from high to low, in all frames. Scheme-C does not provide fairness guarantees but it can maximize the system utilization efficiency. We also can use a simple round robin transmission scheduling strategy but with fixed instantaneous data rate (fixed spreading code) strategy for rate allocation to form an allocation scheme and two such schemes, labelled as *Scheme-D* and *Scheme-E*, use fixed rate of 120kbps and 240kbps for all scheduled flows in all frames, respectively.

Table 6.1: Some simulation parameters and their values

Parameter	Value
spreading bandwidth	3.84MHz
frame length	10ms
total background noise	1.0×10^{-12} Watt
peak transmission power	0.1 Watt
f_s	0.95
minimum/maximum velocity (if move)	3/80 km/h

We consider a single cell of 1km radius with the base station located at the center of the cell. An uplink transmission channel is shared among several data users. The available instantaneous data rates are chosen from the set $\mathcal{R} = \{15, 30, 60, 120, 240, 480, 960\}$ (in kbps) [1] and the corresponding target E_b/I_0 are set as $\{5.8, 6.5, 7.1, 7.8, 8.4, 9.1, 9.7\}$ (in dB). In our simulations, the length of a frame is set the same as a radio transmission frame of 10ms. Two scenarios are considered: one is that all users stay stationary and the other is that users either move within the cell or stay stationary randomly. The path gains are simulated as mutually independent random processes determined by distance path loss and slow fading (shadowing). Lee's model [81] is used to compute the path loss and a correlation model [33] is used to compute shadowing when users are moving. The path gain (in dB) is given as

$$-h = L + 10n \log_{10}(d) + X \tag{6.16}$$

where L is a constant, n is the path loss exponential and X is the fading process. In our simulations, we set $L = 71$ dB and $n = 4$. The slow fading process X is modelled as an autoregressive process. Let $X^t = (1-\beta)X^{t-1} + \beta Y^t$ be the fading level at frame t, where Y is an independent sequence of log normal random variables with log standard deviation of 10 (dB) and β is a weighting factor which is determined by the ratio of the velocity of a user to the maximum velocity in our simulations. If a user is stationary, the slow fading level is constant and a user with high velocity has less correlated successive fading levels. We assume that path gains remain constant during a frame. To relate

[1] These transmission formats are according to the UMTS standard [15]

the path gain experienced by a flow, we use a discrete scoring function, as shown in Fig 6.3. Other simulation parameters and their values are summarized in Table 6.1.

We first consider a stationary scenario. 8 data users are uniformly distributed in the cell and the distances between users $1, 2, \cdots, 8$ and the base station are set at $200, 300, \cdots, 900$(m), respectively. Each user is assumed to have infinite data backlogged for transmission. Fig. 6.5 plots the achievable system capacity (average total system throughput in total error-free bits received) for different number of active data users. Not unexpectedly, exhaustive instantaneous data rate allocation (Schemes A, B and C) improves the achievable system capacity compared with fixed instantaneous data rate allocation (Scheme D and E) since the former tries to pack as many flows as possible in a frame until the power capacity is achieved. The achievable system capacity of Scheme-A and that of Scheme-B are seen to decrease with increasing number of active users. This is due to the increased likelihood of users with poor path gains transmitting, thereby increasing the level of interference in the cell. Nevertheless, note that Scheme-A results in a higher system capacity than Scheme-B. Since Scheme-C always selects users in the order of their channel conditions, its resulting system capacity is the best. However, Scheme-C may cause starvation to some users as shown in Fig. 6.6, which plots the individual user throughput when there are 8 active users. Scheme-B results in all users having the same throughput. This is obtained at the expense of a reduced system capacity as shown in Fig. 6.5. Scheme-A is seen as a compromise between Scheme-B and Scheme-C, yielding improved system capacity over Scheme-B without starving any user of transmission compared with Scheme-C.

Next, when users are not stationary, resulting in time-varying measures of effort, we use the convergence rate (6.8) to illustrate the effectiveness of the FECA algorithm. We model user mobility as follows. A user moves randomly in a cell. The velocities of users are independent random variables uniformly distributed between 3km/h (minimum) and 80km/h (maximum). The directions of users are independent random variables uniformly distributed between 0 and 2π. We also assume that a mobile user randomly changes its speed and/or direction every few seconds. If a user moves out of the border of a cell, we assume that it reappears at a point that is symmetric to the exiting point. We measure the convergence rate, $\delta_i(T)$, and use $\Delta(T) = max_i\{\delta_i(T)\}$ for different

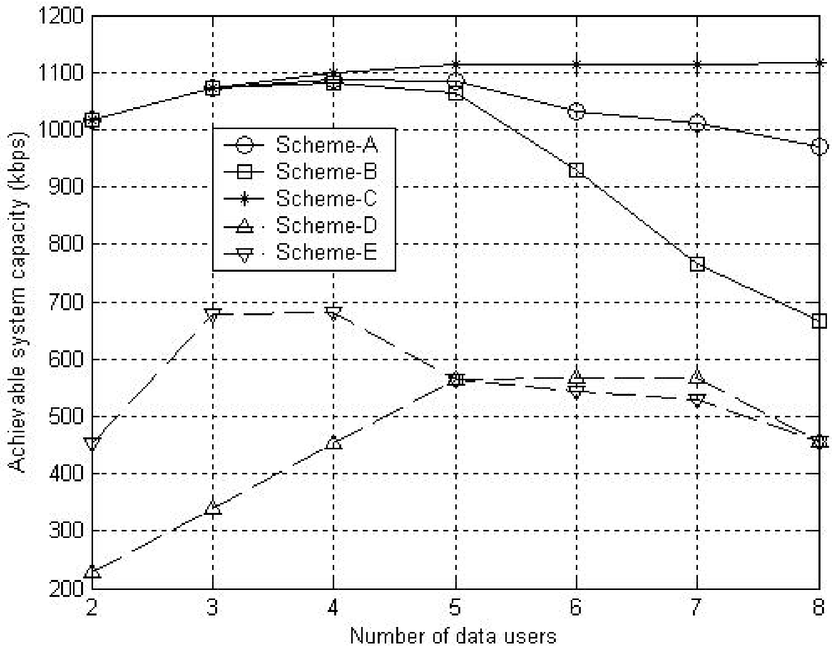

Figure 6.5: Achievable system capacity vs the number of active users. The system load increases with the increase of the number of active users

T as the worst case analysis for verifying the effectiveness of the FECA algorithm. Fig. 6.7 plots the worst case convergence rate for different numbers of data users. We see that the FECA algorithm approximates the discrete FES model well: the worst case normalized leading/lagging percentage reduces quickly when the number of frame increases. When the number of backlogged users increases, a lagging user takes a longer time to obtain its fair share of system resources. This is shown in the figure where it converges faster when there are fewer competing users.

We then evaluate the delay performance. We use the UMTS web surfing traffic model [14] as the data model for each mobile. According to [14], a web surfing session consists of several packet bursts and between two consecutive packet bursts is the reading time modelled as geometrically distributed. We assume that each mobile has a web surfing session consisting of an infinite number of packet bursts. The length of network layer packets is modelled as Pareto distributed. The network layer packets are segmented into equal length (150bits) RLC PDUs (Radio Link Control Protocol Data Units) and the maximum number of retransmission is 5 for each RLC PDU. The

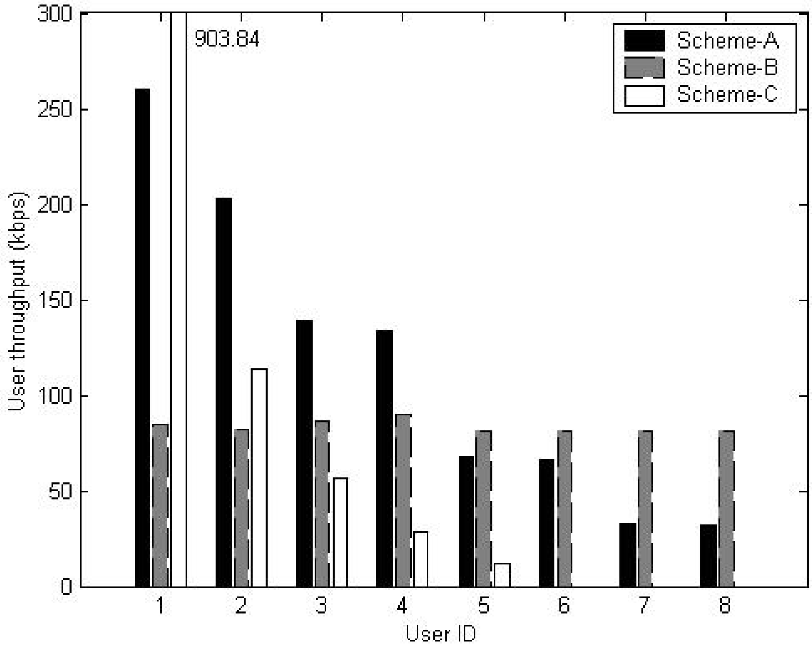

Figure 6.6: Individual user's throughput

buffer length of each mobile is set as 5000 RLC PDUs. We use the UDD 144kbps web surfing model in simulations and other detailed parameters can be found in [14]. The delay of an arbitrary network layer packet is computed as the time of its service completion minus its arrive time. We only compare the delay performance of Schemes A, B and C. In each frame, we run all schemes consecutively and record the delay statistics for different schemes separately. This is done by copying an arriving packet of a flow and enqueueing/dequeueing it to/from the separate queues for different schemes accordingly. All flows initiate a packet bursts at the beginning of a simulation run. A simulation run has 100000 frames and all results are averaged over 100 runs. In the mobility scenario, the initial position and the mobility pattern of a user is set/reset randomly at the beginning of a simulation run.

We first present simulation results in a stationary scenario. 80 stationary data users are uniformly distributed in a cell and the distance between users 10, 20, $\cdots$, 80 and the base stationary are set at 200, 300, $\cdots$,900 (m), respectively. We use a large number of users in order to simulate a loaded system situation. Fig. 6.8 plots the system average delay (averaged over all active users) for different number of active users. Fig. 6.9 plots

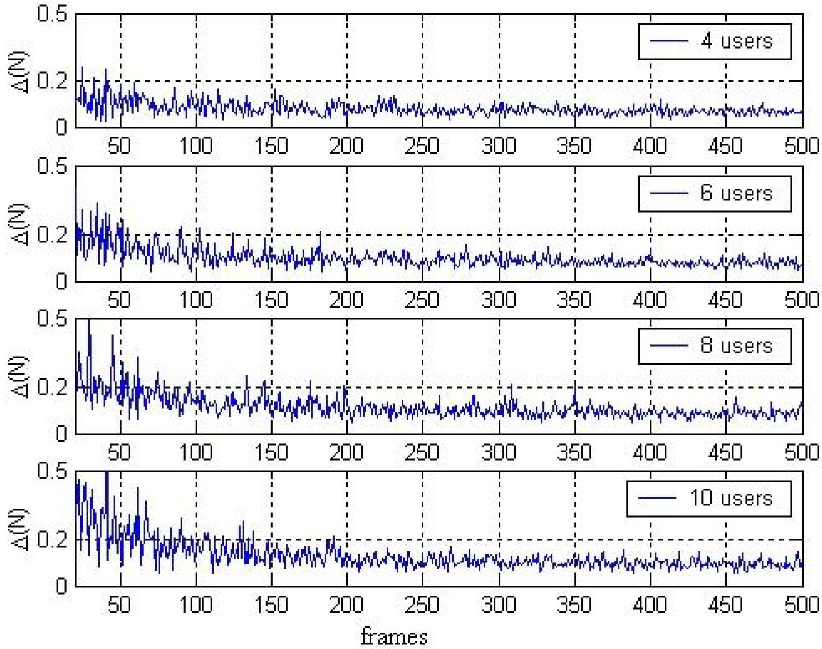

Figure 6.7: Worst case analysis of convergence rate

the delays of users with different locations when there are different active users in the system. Note that in Scheme C, the users with poor channel conditions still have the chance to transmit when the users with good channel conditions have depleted their buffers. As Scheme B does not discriminate users and schedules users on a round robin basis, it results in similar delays of all active users. Schemes A and C discriminate users according to their distances to the base station, thus users far away from the base station experience larger delays compared with users closer to the base station. Both are shown in Fig. 6.9. As Scheme B does not suffer from reduced system capacity in an unloaded or lightly-loaded system, it also results in a smaller system average delay due to its round robin scheduling, as shown in Fig. 6.8. However, when the system becomes loaded, Scheme B results in some reduction of system capacity which in turn increases the average delays of all active users.

We next present simulation results under the mobility scenario. Fig. 6.10 plots the system average delay (averaged over all active users) for different numbers of data users. We see from Fig. 6.10 that Scheme C has the smallest average delay. We explain it as follows. In the mobility scenario, a user may experience time-varying

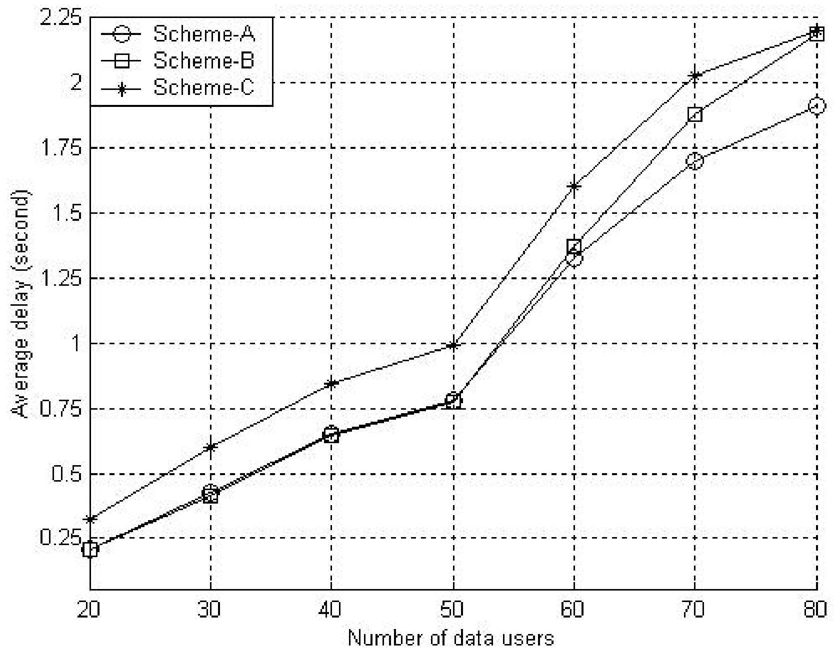

Figure 6.8: System average delay vs the number of active users. The system load increases with the increase of the number of active users.

and location-dependent path gains resulting from its random movements. Recall that Schemes C and A favor users with high path gains by prioritizing their transmissions ahead of users with low path gains. By such transmission prioritization, Schemes A and C can exploit the time-varying channel conditions to some extent in that some users might improve their channel conditions some time later. Also such exploitation of time varying channel conditions may lead to some gains of system capacity improvement (cf. Fig. 6.5), and such gains then can be translated into the decrease of average delay compared with Scheme B in simulations. On the other hand, Scheme B may fail to maintain a high system capacity due to its GPS model based fairness constraint in case of loaded and over-loaded situations (cf. Fig. 6.5). Thus the loss of system capacity, which is at the expense of its fairness enforcement, can be translated into the increase of the system average delay of Scheme B. As Scheme A still has to take care of users with poor channel conditions by not starving them too long, it may experience some losses of system capacity. Thus the average delay of Scheme A is in the middle of Scheme B and Scheme C. However, we should mention that the improvements of Scheme C in

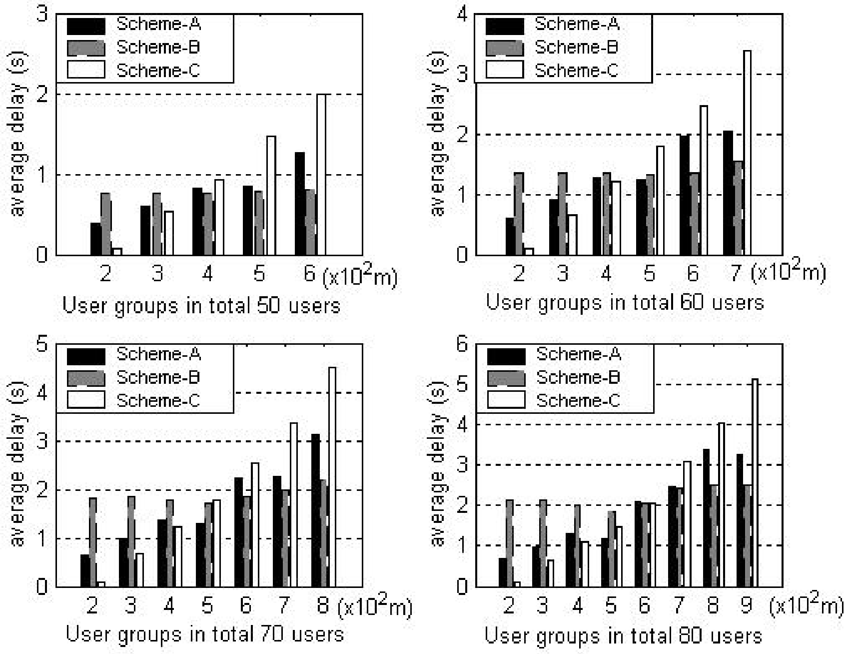

Figure 6.9: Average delay of data users with different locations in stationary scenario (the number of simulated data users is labelled below each sub-figure)

Fig. 6.10 (compared with Scheme B and Scheme A) need to be understood only from the statistical average point of view. This is because although the exploitation of time-varying channel conditions can capture some gains in terms of the increased system capacity (or in terms of the decreased average delay in our simulations), such gains may be at the expense of additional delay of some particular packets of a particular user (cf. Fig 6.11). Fig. 6.11 plots the average delay of some selected data users and the average delay of all simulated data users from only one run of simulation. Note that the arriving packets of a user have the same arrival time for all schemes but they may have different service completion times in different schemes. We see from Fig. 6.11 that some users may experience very large delays in Scheme C due to its biased scheduling scheme.

We summarize our simulations as follows. We first show that the GPS model based fairness criterion may lead to significant decrease in system capacity when the system becomes loaded, while the FES model based fairness criterion can avoid this by allocating resource shares to users according to their efforts, i.e., their path gains in

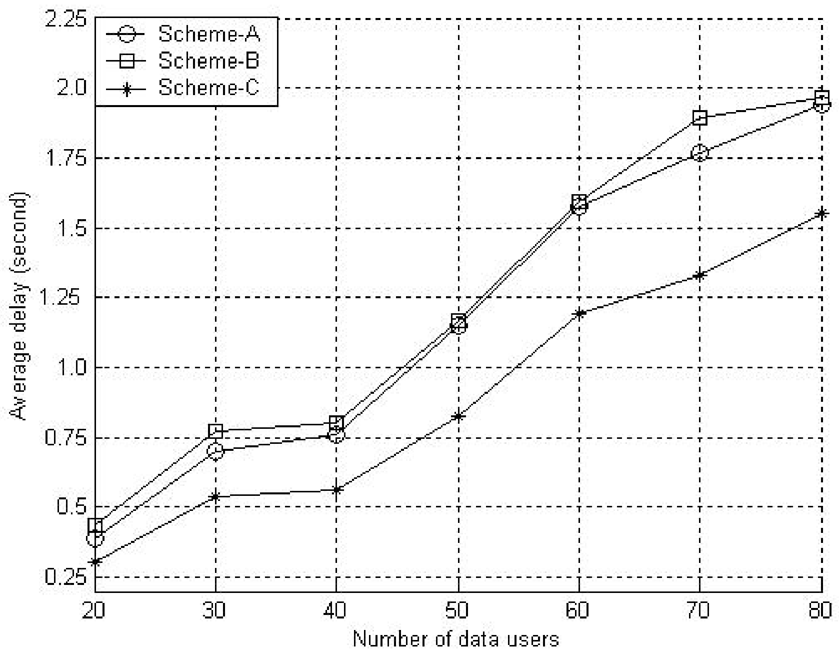

Figure 6.10: System average delay vs the number of data users. The system load increases with the increase of the number of users.

our simulations. Our simulations also clearly illustrate the (possible) tradeoff between system utilization efficiency and fairness among users. We then verify the effectiveness of the FECA algorithm. Although individual users may suffer from the transmission prioritization, they may also benefit from high system utilization efficiency by such transmission prioritization, especially when users have time-varying channel conditions due to their random mobility. Our final simulations provide such examples.

6.5 Summary

In this chapter, we have investigated packet level resource allocation policies for data users from a system operator viewpoint. We focus on two policy design objectives: fairness among users and system utilization efficiency. Unlike the commonly used GPS fairness model, we propose a new FES fairness model. Given any infinitesimally small time interval, the FES has the same objective as that of the GPS model, i.e., to minimize the difference in weighted allocated resources between any two backlogged

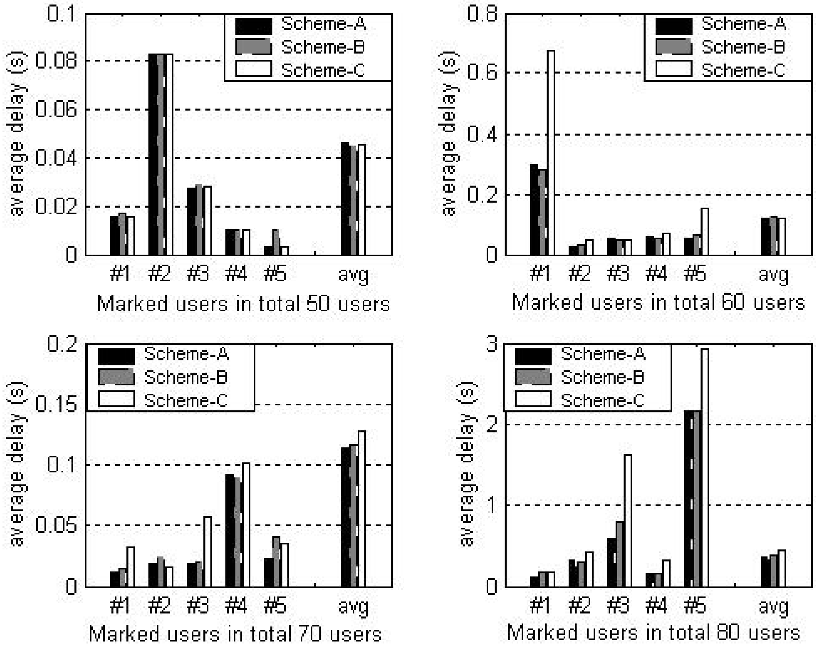

Figure 6.11: Average delay of selected individual data users and average delay of all simulated data users in one run of simulation under mobility scenario (the number of simulated data users is labelled below each sub-figure)

users. However, the nominal weight of a flow is considered time-dependent in the FES model while it is fixed in the GPS model. By this simple modification, we can incorporate the (possible) interaction between users and the resource allocation process. We have also proposed a simple credit-based FECA algorithm to approximate the FES model at the packet level. Based on the FES model and the FECA algorithm, we have also presented a detailed packet level resource allocation scheme for CDMA-based wireless networks. The scheme consists of mechanisms for resource share assignment, transmission scheduling, rate and power allocation. We use exhaustive instantaneous data rate allocation in order to fully utilize the system capacity, while optimal power allocation can provide required guarantees of the transmission quality. We evaluate our proposals via simulations. The simulation results show the advantages of using the FES model as the fairness reference in terms of the system utilization efficiency and verify the effectiveness of the FECA algorithm.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis, we have studied several important radio resource management issues in a cellular mobile network for data services at the packet level. Radio resource management is very important in that it improves the resource utilization efficiency while meeting QoS requirements. With the proliferation of the Internet and its applications, data services will form a large part of the traffic in next generation wireless networks. Many current literature mainly focus on realtime services and few are dedicated to data services. When designing a control policy, we often have to face different costs and an ideal policy should optimally balance these costs. This thesis is devoted to data services and further devoted to studying how to balance different costs. In this thesis, we have studied the following resource management issues, namely, power control, transmission scheduling and rate allocation. We first study these issues separately from a single user's point of view and then jointly from an operator's viewpoint.

The first set of problems is modelled from the stochastic decision theoretic framework and solved by using the MDP mathematical tool. In Chapter 3, a power control policy is required to save transmission energy while meeting the file transfer delay requirement. We have shown how to convert such a constrained stochastic optimization problem to a standard Markov decision problem via the Lagrangian approach. The resulting optimal power control policy is independent of time with the average delay constraint while is time dependent with the strict delay constraint. Numerical examples have shown that besides meeting the delay constraint, the optimal policy greatly reduces transmission energy compared to a fixed power persistent transmission policy. This happens because the channel variations have been opportunistically exploited by the optimal policy.

In Chapter 4, a transmission control policy is required to optimally balance between the transmission cost, the delay cost and the throughput cost. We directly model the problem as an average cost optimal Markov decision problem. We prove the existence of stationary average optimal policies for our problem and explore the property of the optimal policies. The resulting optimal policy is proven to have a structural property: when the buffer occupancy is low, the sender can suspend transmission in some bad channel states to save transmission power; however, when the buffer occupancy exceeds some thresholds, the sender has to transmit in some bad channel states to avoid increasing the delay cost.

In Chapter 5, a rate control policy is designed to minimize the resource usage cost and the delay cost. The resulting optimal policy is shown to have a monotone property, i.e., the optimal action is nondecreasing with the system state. We have also analyzed two extreme policies that give the upper and lower delay bounds among all allowable policies. The analysis is based on the stochastic processes comparison technique. We then propose a class of one-threshold based simple policies to approximate the optimal policy and analyze the upper delay bound for such a simple policy. We also extend the rate control problem against the existence of competitions among multiple users. We then identify the characteristic of value functions and the property of optimal policies for such an extended problem.

We have also studied resource allocation from the viewpoint of an operator. In Chapter 6, we present an integrated resource allocation scheme covering power control, transmission scheduling and rate allocation mechanisms. We propose a new fairness model, the *fair-effort resource sharing* model, and a simple credit based algorithm to implement the proposed fairness model. According to our fairness model, the resource share (quota) allocated to a user is proportional to the user's effort which is considered as time dependent rather than as fixed. Based on our fairness model, we provide a detailed packet level resource scheme which consists of optimal power allocation, exhaustive instantaneous data rate allocation and fair-effort resource sharing. Numerical results are also provided to show the advantages of using our fairness model in terms of the increased system utilization efficiency compared to that of the generalized processor sharing model.

There are still some possible extensions to this thesis which deserve further research. We list some of these issues in the next section.

7.2 Some Future Research Directions

We have successfully applied the MDP theory to model and to solve some resource allocation problems. However, their real applications have some restrictions as we may not always know all quantities beforehand. In this case, adaptive control techniques can be used for online control. When multiple users are considered, competitions arise across users. Another important technique, game theory, can be used to model such a situation. When the topology of a wireless network changes, the available resources may also change accordingly. In such a case, we may need to reconsider the fairness definition and investigate its impact to resource allocation. We briefly discuss these issues and some very recent related works.

Adaptive Control

The theory of Markov decision processes provides a solid mathematical basis for finding an optimal policy, while *reinforcement learning* (see [82] for introduction) provides implementable method to approximate an optimal policy. Some classical reinforcement learning methods include Temporal Difference learning, Q-learning and R-learning [7, 9, 82]. For example, the simplest one-step Q-learning is defined by

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left\{ C(s,a) + \min_{a' \in \mathcal{A}_{s'}} Q(s',a') - Q(s,a) \right\}, \tag{7.1}$$

where $Q(s,a)$ is the Q-factor and $\alpha \in (0,1]$ is a step size parameter. From 7.1, we see that we may not need to know the transition probabilities beforehand when finding an optimal policy. Recently, reinforcement learning has been used for admission control [11, 51] and rate control [52, 66] in a wireless network. We believe that if the property of an optimal policy has been identified and exploited, the efficiency and the convergence rate of the learning algorithm can be greatly improved.

Game Theory

Game theory (see [24, 25] for introduction) provides the theoretical foundation to model how a user adjusts its strategy to maximize its return from competing with other users. The outcome of a game is an equilibrium, the *Nash Equilibrium* [55], that no participant can benefit more by changing its strategy separately. Recently, C. Saraydar et al. have applied game theory to solve the power control problem in a CDMA network [31, 69, 70, 71]. Each user is assigned a utility function of its own power and the others, $U_i(p_i, \mathbf{P}_{-i})$ and the game is defined as:

$$\max_{p_i} U_i(p_i, \mathbf{P}_{-i}) \text{ for all } i. \tag{7.2}$$

This game is a static game while the characteristics of a *mobile* user are not included, e.g., the time-varying channel. We believe that the theory of competitive Markov decision processes, the combination of MDP theory and game theory, could be a more appropriate mathematical tool to model and to solve the competitive and dynamic resource allocation problems.

Fairness Reconsideration

Fairness has always been an important and hot topic in resource allocations. However, the definition of fairness may need to be reconsidered in wireless networks. An example is the fairness in a mobile ad-hoc network. The topology of a mobile ad-hoc network may change in a small time scale and the time-varying connectivity may impact the available resources for allocation [83]. Recently, many new fairness definitions have been proposed, for example, *relative fairness* in [72], *statistical fairness* in [47], for wireless networks. Statistical fairness is defined by

$$\Pr\left(\left|\frac{W_i(t_1,t_2)}{\omega_i} - \frac{W_j(t_1,t_2)}{\omega_j}\right| \geq x\right) \leq f(i,j,x). \tag{7.3}$$

Another issue is the fairness measurement. Throughput is often chosen as the fairness measurement but in wireless networks, the power used to achieve the same throughput may not be the same for different mobile users due to their different geographical locations. To address such a problem, some researchers have proposed to use utility functions as the fairness metric [46, 95]. We believe that there should be much research work to do with the definition, measurement, and evaluation of fairness in wireless networks.

Bibliography

[1] E. Altman, *Applications of Markov Decision Processes in Communication Networks*, chapter 13, pages 537–538, Kluwer Academic Publishers, 2000.

[2] M. A. Arad, *Resource Allocation and Channel Access Control in Multimedia Wireless CDMA Networks*, Ph.D. Thesis, Dept. Electrical and Computer Engineering, Univ. of Toronto, 2000.

[3] M. A. Arad and A. Leon-Garcia, "Scheduled CDMA: A Hybrid Multiple Access for Wireless ATM Networks", *IEEE PIMRC*, pages 913–917, 1996.

[4] M. A. Arad and A. Leon-Garcia, "A Generalized Processor Sharing Approach to Time Scheduling in Hybrid CDMA/TDMA", *IEEE INFOCOM*, pages 1164–1171, 1998.

[5] J. C. Bennett and H. Zhang, "WF^2Q: Worst-case fair weighted fair queueing", *IEEE INFOCOM*, pages 120–128, 1996.

[6] R. A. Berry and R. Gallager, "Communication over Fading Channels with Delay Constraints", *IEEE Trans. on Information Theory*, 48(5):1135–1149, May 2002.

[7] D. P. Bertsekas, *Dynamic Programming and Optimal Control (vol II)*, Athena Scientific, 1995.

[8] D. P. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, 1992.

[9] D. P. Bertsekas and J. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.

[10] A. Demers, S. Keshav, and S. Shenker, "Analysis and Simulation of A Fair Queueing Algorithm", *IEEE SIGCOM*, pages 1–12, 1989.

[11] E.-S. El-Alfy, Y.-D. Yao, and H. Heffes, "A Model-Based Q-Learning Scheme for Wireless Channel Allocation with Prioritized Handoff", *IEEE Globecom*, pages 3668–3672, 2001.

[12] M. Elaoud and P. Ramanathan, "Adaptive Allocation of CDMA Resources for Network-level QoS Assurances", *IEEE Mobicom*, pages 191–199, 2000.

[13] M. H. et.al., "The TD-CDMA Based UTRA TDD Mode", *IEEE Journal on Select. Areas Comm.*, 18(8):1375–1385, Aug 2000.

[14] ETSI, "UMTS Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS", *ETSI TR 101 112 v3.2.0*, April 1998.

[15] ETSI, "UMTS Physical channels and mapping of transport channels onto physical channels (FDD)", *ETSI TS 125 211 v3.4.0*, Sept. 2000.

[16] ETSI, "UMTS Physical channels and mapping of transport channels onto physical channels (TDD)", *ETSI TS 125 221 v3.3.0*, June 2000.

[17] ETSI, "UMTS Physical layer measurements (FDD)", *ETSI TS 125 215 v3.3.0*, June 2000.

[18] ETSI, "UMTS RLC Protocol Specification", *ETSI TS 125 322 v3.3.0*, June 2000.

[19] ETSI, "UMTS RRC Protocol Specification", *ETSI TS 125 331 v3.3.0*, June 2000.

[20] ETSI, "UMTS Service Aspects; Services and Services Capabilities", *ETSI TS 122 105 v3.9.0*, June 2000.

[21] ETSI, "*UMTS QoS Concept and Architecture*", *ETSI TS 123 107 v3.4.0*, Oct. 2000.

[22] ETSI, "*UMTS Radio Resource Management Strategies*", *ETSI TR 125 922 v3.2.0*, June 2000.

[23] ETSI, "UMTS Physical layer measurements (TDD)", *ETSI TS 125 225 v3.9.0*, Dec 2001.

[24] F. Forgó, J. Szép, and F. Szidarovszky, *Introduction to the Theory of Games-Concepts, Methods, Applications*, Kluwer Academic Publishers, 1999.

[25] D. Fudenberg and J. Tirole, *Game Theory*, The MIT Press, 1991.

[26] R. G. Gallager, *Discrete Stochastic Processes*, Kluwer Academic Publishers, 1995.

[27] A. Genz, "Numerical Computation of Multivariate Normal Probabilities", *Journal of Comp. Graph Stat.*, pages 141–149, 1992.

[28] E. N. Gilbert, "Capacity of a burst-noise channel", *Bell Systems Tech. J.*, 39:1253–1266, 1960.

[29] A. J. Goldsmith, "The Capacity of Downlink Fading Channels with Variable Rata and Power", *IEEE Trans. on Vehicular Technology*, 46(3):569–580, August 1997.

[30] A. J. Goldsmith and P. P. Varaiya, "Capacity of Fading Channels with Channel Side Information", *IEEE Trans. on Information Theory*, 43(6):1986–1992, November 1997.

[31] D. Goodman and N. Mandayam, "Power Control for Wireless Data", *IEEE Personal Communications Magazine*, pages 48–54, April 2000.

[32] M. Goyal, A. Kumar, and V. Sharma, "Power Constrained and Delay Optimal Policies for Scheduling Transmission over a Fading Channel", *IEEE INFOCOM*, 2003.

[33] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems", *IEE Electronics Letters*, 23:2145–2146, Nov 1991.

[34] Ö. Gürbüz and H. Owen, "Dynamic Resource Scheduling Strategies for QoS in W-CDMA", *IEEE GLOBECOM*, pages 183–187, 1999.

[35] Ö. Gürbüz and H. Owen, "Dynamic Resource Scheduling Schemes for W-CDMA Systems", *IEEE Communications Magazine*, pages 80–84, October 2000.

[36] J. Herdtner and E. Chong, "Analysis of a Class of Asynchronous Power Control Algorithms for Cellular Wireless Systems", *IEEE Jounal on Select. Areas Comm.*, 18:436–446, May 2000.

[37] T. Holliday and A. Goldsmith, "Optimal Power Control and Source-Channel Coding for Delay Constrained Traffic over Wirelees Channels", *IEEE ICC*, pages 831–835, 2002.

[38] T. Holliday, A. Goldsmith, and P. Glynn, "Optimal Link Adaptation in Wide-band CDMA Systems", *IEEE Global Telecomm. Conf. (Globecom)*, 2002.

[39] T. Holliday, A. Goldsmith, and P. Glynn, "Wireless Link Adaption Policies: QoS for Deadline Constrained Traffic with Imperfect Channel Estimates", *IEEE ICC*, pages 3366–3371, 2002.

[40] H. Holma and E. Antti Toskala, *WCDMA for UMTS-Radio Access for Third Generation Mobile Communications*, John Wiley and Sons Ltd, 2000.

[41] F. Jerzy and K. Vrieze, *Competitive Markov Decision Processes*, Springer, Inc., 1997.

[42] J. Keilson and A. Kester, "Monotone Matrics and Monotone Markov Processes", *Stochastic Processes and their Applications*, 5:231–241, 1997.

[43] F. Kelly, "Charging and Rate Control for Elastic Traffic", *European Trans. on Telecommunications*, 8(1):33–37, January 1997.

[44] F. Kelly, A. Maulloo, and D. Tan, "Rate Control in Communication Networks: Shadow Priceis, Proportional Fairness and Stability", *Journal Oper. Res.*, 49:237–252, 1998.

[45] P. Y. Kong, *Medium Access Control and Transmission Scheduling in Multicode-CDMA Networks*, Ph.D. Thesis, Dept. Electrical and Computer Engineering, The National University of Singapore, 2001.

[46] P. Liu, R. Berry, M. L. Honig, and S. Jordan, "Slow-Rate Utility-Based Resource Allocation in Wireless Networks", *IEEE Globecom*, 2003.

[47] Y. Liu, S. Gruhl, and E. W. Knightly, "WCFQ: An Opportunistic Wireless Scheduler with Statistical Fairness Bounds", *IEEE Transactions on Wireless Communications*, 2(5):1017–1028, September 2003.

[48] F. Lu and R. Serfozo, "M/M/1 Queueing Decision Processes with Monotone Hysteretic Optimal Policies", *Oper. Res.*, 32(5):1116–1132, Sep.-Oct. 1984.

[49] S. Lu, V. Bharghavan, and R. Srikant, "Fair Scheduling in Wireless Packet Networks", *IEEE/ACM Trans. on Networking*, 7(4):473–489, 1999.

[50] D. Ma, A. Makowski, and A. Schwartz, "Estimation and Optimal Control for Constrained Markov Chains", *IEEE Conf. on Decision and Control*, 1986.

[51] B. Makarevitch, "Application of Reinforcement Learning to Admission Control in CDMA Network", *IEEE, PIMRC*, 2:1353–1357, 2000.

[52] B. Makarevitch, "Learning Rate Control for Downlink Shared Channel in WCDMA", *IEEE, PIMRC*, 2003.

[53] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels", *IEEE Trans. Inform. Theory*, 35(6):1277–1290, 1989.

[54] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation Technqieus in Wireless Packet Data Services", *IEEE Communications Magazine*, 38(1):54–64, Januray 2000.

[55] J. F. Nash, "Non-cooperative games", *Annals of Mathematics*, 54:289–295, 1951.

[56] T. E. Ng, I. Stoica, and H. Zhang, "Packet Fair Queueing Algorithms for Wireless Networks with Location-dependent Errors", *IEEE INFOCOM*, pages 1133–1140, 1998.

[57] S.-J. Oh, T. L. Olsen, and K. M. Wasserman, "Distributed Power Control and Spreading Gain Allocation in CDMA Data Networks", *IEEE INFOCOM*, pages 379–385, 2000.

[58] S.-J. Oh and K. M. Wasserman, "Dynamic Spreading Gain Control in Multiservice CDMA Networks", *IEEE Journal on Select. Areas Comm.*, 17(5):918–927, May 1999.

[59] S.-J. Oh and K. M. Wasserman, "Optimality of Greedy Power Control and Variable Spreading Gain in Multi-class CDMA Mobile Networks", *IEEE MOBICOM*, pages 102–112, 1999.

[60] A. Parekh and R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single-node case", *IEEE/ACM Trans. Networking*, 1:344–357, June 1993.

[61] J. Proakis, *Digital Communications*, McGraw-Hill, 1989.

[62] M. L. Puterman, *Markov Decision Processes - Discrete Stochastic Dynamic Programming*, John Wiley and Sons. Inc., 1994.

[63] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay and Rate Constrained Transmission Policies over Wireless Channels", *IEEE Proc. of Global Telecommunications Conference (Globecom)*, 2:806–810, 2001.

[64] D. Rajan, A. Sabharwal, and B. Aazhang, "Transmission Policies for Bursty Traffic Sources on Wireless Channels", *Conference on Information, Science, and Systems (CISS)*, pages 429–433, March 2001.

[65] T. Rapport, *Wireless Communications Principles and Practice*, Prentice Hall, 1996.

[66] F.-C. Ren, C.-J. Chang, and Y.-S. Chen, "A Q-Learning-Based Multi-Rate Transmission Control Scheme for RRC in WCDMA Systems", *IEEE, PIMRC*, 2002.

[67] O. Sallent, J. P. Romero, F. J. Casadevall, and R. Agusti, "An Emulator Framework for a New Radio Resource Management for QoS Guaranteed Services in

W-CDMA Systems", *IEEE Journal on Select. Areas Comm.*, 19(10):1893–1904, Oct 2001.

[68] A. Sampath, P. Kumar, and J. Holtzman, "Power Control and Resource Management for a Multimedia CDMA Wireless System", *Proc. of IEEE PIMRC*, pages 21–25, 1995.

[69] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pareto Efficiency of Pricing Based Power Control in Wireless Data Networks", *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, pages 21–24, 1999.

[70] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pricing and Power Control in a Multicell Wireless Data Network", *IEEE Journal on Select. Areas Comm.*, 19(10):1883–1892, October 2001.

[71] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient Power Control via Pricing in Wireless Data Networks", *IEEE Trans. on Comm.*, 50(2):291–303, Februray 2002.

[72] S. Sarkar and K. N. Sivarajan, "Fairness in Cellular Mobile Networks", *IEEE Trans. Information Theory*, 48(8):2418–2426, August 2002.

[73] L. I. Sennott, *Stochastic Dynamic Programming and the Control of Queueing Systems*, John Wiley and Sons, New York, 1999.

[74] M. Soleimanipour, W. Zhuang, and G. H. Freeman, "Optimal Resource Management in Wireless Multimedia Wideband CDMA Systems", *IEEE Trans. Mobile Computing*, 1(2):143–160, April-June 2002.

[75] L. Song, N. B. Mandayam, and Z. Gajic, "Analysis of an Up/Down Power Control Algorithm for CDMA Reverse Link: A Nonlinear Control System Approach", *IEEE Journal on Select. Areas Comm.*, 19:277–286, Feb. 2001.

[76] S. Stidham, "A survey of Markov decision models for control of networks of queues", *Queueing Systems*, 13:291–314, 1993.

[77] S. Stidham, "Analysis, Design and Control of Queueing Systems", *Oper. Res.*, 50(1):197–216, 2002.

[78] S. Stidham and R. R. Weber, "Monotonic and Insensitive Optimal Policies for Control of Queues with Undiscounted Costs", *Oper. Res.*, 37:611–625, Jul.-Aug. 1989.

[79] S. Stidham and R. R. Weber, *Monotone Optimal Policies for Left-Skip-Free Markov Decision Processes*, chapter 13, pages 191–202, Kluwer Academic Publishers, 1999.

[80] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, John Wiley and Sons, 1983.

[81] G. L. Stüber, *Principles of Mobile Communication*, Kluwer Academic Publishers, 1996.

[82] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, USA, 1998.

[83] L. Tassiulas and S. Sarkar, "Maxmin Fairness Scheduling in Wireless Networks", *IEEE INFOCOM*, pages 763–772, 2002.

[84] B. Wang, K. C. Chua, and H.-P. Schwefel, "Resource Allocation Strategies for Non-realtime Data Users in Uplink UTRA TDD", *Proc. of IEEE WCNC*, pages 1055–1059, 2003.

[85] B. Wang and K. C. Chua, "Adaptive Rate Allocation in a CDMA-based Wireless Network", *a journal version paper draft, in preparation*, 2003.

[86] B. Wang and K. C. Chua, "Fair-Effort Resoure Allocation in a CDMA-based Wireless Network", *a journal version paper draft ready for submission*, 2003.

[87] B. Wang and K. C. Chua, "Optimal File Transfer over a Fading Channel using Adaptive Power Allocation with Delay Constraints", *a jounal version paper draft ready for submission*, 2003.

[88] B. Wang and K. C. Chua, "Resource Allocation Policies for Data Services in a CDMA-based Wireless Network", *Proc. of IEEE PIMRC*, pages 627–631, 2003.

[89] B. Wang and K. C. Chua, "Resource Request Policy for Data Services in a CDMA Uplink", *Proc. of IEEE PIMRC*, pages 631–637, 2003.

[90] B. Wang and K. C. Chua, "Transmission Control Policies for Packet Data over a Time-varying Channel", *a jounal version paper draft ready for submission*, 2003.

[91] B. Wang and K. C. Chua, "Transmission Control Policy for Data Services over a Time-varying Wireless Channel", *Proc. of IEEE ICON*, pages 617–621, 2003.

[92] H. Wang and N. Mandayam, "Energy and Delay Constrained Power Control for File Transfers over Fading Channels", *Proc. of Asilomar Conference*, Nov 2001.

[93] H. Wang and N. Mandayam, "Opportunistic File Transfer under Energy and Delay Constraints", *Proceedings of International Symposium on Wireless Communication*, Sept 2002.

[94] H. S. Wang and N. Moayeri, "Finite-State Markov Channel–A Useful Model for Radio Communication Channels", *IEEE Trans. Vehicular Tech.*, 44(1):163–171, February 1995.

[95] M. Xiao, N. B. Shroff, and E. K. Chong, "A Utility-based Power-Control Scheme in Wireless Cellular Systems", *IEEE/ACM Trans. Networking*, 11(2):210–221, April 2003.

[96] J. Zander and S.-L. Kim, *Radio Resource Management for Wireless Networks*, Artech Hourse, Incorporated, 2001.

[97] D. Zhang and K. M. Wasserman, "Energy Efficient Data Communication Over Fading Channels", *Proc. of IEEE WCNC*, pages 986–991, 2000.

[98] D. Zhang and K. M. Wasserman, "Transmission Schemes for Time-Varying Wireless Channels with Partial State Observations", *IEEE INFOCOM*, pages 467–476, 2002.

[99] Q. Zhang and S. A. Kassam, "Finite-State Markov Models for Rayleigh Fading Channels", *IEEE Trans. Communications*, 47(11):1688–1692, 1999.

[100] D. M. Zhao, *Radio Resource Management in Cellular CDMA Systems Supporting Heterogeneous Services*, Ph.D. Thesis, Electrical and Computer Engineering Department, University of Waterloo, Canada, 2002.

[101] M. Zorzi, R. Rao, and L. Milstein, "ARQ Error Control for Fading Mobile Radio Channels", *IEEE Trans. on Vehicular Tech.*, 46(2):445–455, May 1997.