Impact of User Mobility on Resource Allocation Schemes in Cellular Radio Systems

Thesis by

Neelesh B. Mehta

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2001

(Submitted January 5, 2001)

Contents

List of Figures

List of Tables

Acknowledgements

I am indebted to my advisor Prof. Andrea Goldsmith for her faith in me and her constant encouragement and guidance. I thank her for giving me the freedom to explore a variety of topics, for always being easily accessible, and for her patience. I also wish to acknowledge her valuable advice that helped me in my job search. Her energy and enthusiasm for anything related to wireless will remain a source of inspiration for me in the future.

I thank Prof. R. J. McEliece, Prof. P. P. Vaidyanathan, Dr. D. Divsalar, Dr. A. Kiely, and Prof. M. Chandy for serving on my candidacy and defense committees.

It has been a pleasure interacting with many of the members of Prof. Goldsmith's group – Mohamed-Slim Alouini, SeongTaek Chung, Kevin Yu, Syed Ali Jafar, Xiangheng Liu, Stavros Toumpis, Hrishikesh Mandyam, and Sriram Vishwanath. A special thanks to Slim and Kevin for their enthusiasm, help, and advice. My thanks are also due to Prof. D. Cox and his group at Stanford and Prof. R. McEliece and his group at Caltech for their useful advice whenever I was stuck on a problem.

My stay at Caltech was made memorable thanks to my friends Murat Mese, Meina Xu, Braun graduate house inmates – Aamod Khandekar, Sony Akkarakaran, M. A. Srinivas, Shachi Gosavi, Siddharth Gadgil, and others. I am grateful to the Stanford Outing club, Mayank Sharma, Marc Warsowe and Renate Albrecht, Vijay S. R. Somandepalli, and Muruganandam M. Thiruchengode for making my stay at Stanford, during the final year of my studies, a fruitful one. The seamless transition to Stanford would not have been possible without the three secretaries at Caltech and Stanford – Linda Dozsa, Lilian Porter, and Joice DeBolt.

I conclude with an acknowledgment to my parents, which I have saved for the very last. I thank them for being around whenever I needed a dose of encouragement. It is difficult for me to put down in words my gratitude to them for their affection, sacrifice, and perseverance. Nothing I ever do can possibly repay the debt I owe them.

Abstract

Next generation wireless cellular radio systems are being designed to provide anytime, anywhere communication capabilities to serve a range of applications. The ability to support mobility is a key reason for the increasing demand for such systems. To accommodate this demand, efficient resource allocation schemes that can operate over the harsh wireless channel environment need to be devised. User mobility has a significant influence on the design and performance of these schemes. The focus of this dissertation is the analysis of the impact of mobility on such resource allocation schemes.

What impact mobility has depends on the scheme under consideration. We first analyze the impact of user mobility on the performance of a link adaptation scheme that employs the recently proposed no-transmission mode. In this scheme, users adapt their modulation and coding for transmitting data packets based on their estimates of the link condition and suspend transmissions when link quality is very poor. Based on a simplified system model, we derive expressions for the system performance as a function of the basic, system-defining parameters. We show that for a stable system, the channel correlation, a function of user speed and feed-back delay of estimates, is an important factor that determines the optimal link adaptation thresholds. We then study a packet based multiple access scheme called Packet Reservation Multiple Access (PRMA), which can simultaneously handle the different traffic requirements of periodic, delay intolerant (voice) and bursty, delay tolerant (data) users. An approximate technique is developed to analyze the impact of user mobility as well as channel fading and interference-induced packet errors on PRMA. Both these effects lead to a premature loss of reservation and, consequently, more dropped packets for voice users. Finally, we look at dedicated channel assignment schemes that assign an entire channel to a user for the duration of his conversation. We investigate heuristic prediction based techniques that take into account mobility traffic statistics to modify the new call access criteria. This is done so as to introduce prioritization for hand-off requests in hitherto unprioritized channel assignment schemes.

Chapter 1 Introduction

Wireless communication systems hold out the promise of providing anytime and anywhere tetherless communication capabilities. Cordless telephones, pagers, cellular phones, wireless personal digital assistants, and other wireless products are fast becoming an integral part of our life. The high penetration level of cellular telephones, as high as 70% [1] in some countries, is an oft quoted example of their popularity. A variety of applications and services like wireless multimedia, home wireless networks, wireless Internet access, wireless local area networks, wireless sensor networks, etc., are being developed or proposed to tap the immense potential of wireless communications.

One of the prime reasons for the popularity of wireless systems and their promising tomorrow is the ability to provide tetherless access. Transmitting information over a wireless channel frees the users from having to worry about connecting cables and from being forced to work close to ports where communication services can be accessed. However, providing such a facility is, technologically speaking, a task easier said than done. One can see this in the wide disparity between the data rates achievable today in wireless and wireline networks. Many reasons are responsible for this. We take a look at a few of them below.

The wireless channel is a harsh and unpredictable environment for transmitting information. The transmitted signal undergoes random fluctuations and attenuation while passing through or getting reflected/scattered from natural and man-made objects. The received signal may also need to be separated from interfering signals emanating from transmitters nearby. Furthermore, the desirability of having lightweight wireless hand-held/portable devices, like pagers and cellular phones, imposes severe constraints on the size of the batteries they can carry, and consequently, the energy that can be stored in them. In addition to this, the requirements of the different kinds of multimedia traffic that wireless systems need to handle can differ significantly. A classic example is the contrast between voice and data traffic. While voice traffic is delay intolerant, error tolerant, and periodic, data traffic is

typically delay tolerant, error intolerant, and bursty. Finally, the radio spectrum, in which these channels are provided, is extremely scarce. Therefore, extremely efficient resource allocation techniques that can best handle the scarce resources like spectrum (bandwidth) and energy to accommodate a large number of users, and which, at the same time, can meet the different quality of service requirements of the various services being envisaged, are needed.

Mobility or tetherless access – the primary reason for the popularity of wireless systems – itself engenders many challenges that need to tackled. First of all, the characteristics of the wireless channel between the transmitter and receiver change as either of them moves from one location to another. How fast these changes occur depends on how fast the transmitter or receiver are moving. Second, the resources, like the wireless channel bandwidth and transmit power, allocated to the user to enable communication, need to be dynamically adapted as he moves. Failure to do so can lead to a significant degradation in throughput, or even connection termination. Third, the information destined for the user needs to be rerouted on the fly as he moves from one access port to another. We explore the implications of user mobility on resource allocation in greater detail in later sections.

The significance of the issues raised by user mobility is illustrated by the vastly different levels of support provided for mobility by the different wireless systems prevalent today. We can, in fact, classify these systems based on the extent to which they support mobility [2]. In the process, we also describe in more detail the various wireless systems [3, 2] we have previously alluded to. The trade-offs implemented in designing these systems determine the range and nature of the applications they can support.

1.1 Classification of Wireless Systems

Cordless telephones Cordless telephones provide a short wireless link to replace the cord connecting the handset to its base unit. The handset must remain within the range of its base unit. Cordless phones provide the same high quality voice as wired handsets without any added network complexity because the base unit looks exactly like a wired telephone to the telephone network. In summary, Cordless phones provide low

mobility, and low power, two way tetherless, wireline voice quality communications.

Wide area wireless data systems Wide area systems provide low data rate communication capabilities over a large coverage area. These systems use high transmitter power from a few base stations having high antennas that cover large regions.[1] For example, Cellular Digital Packet Data (CDPD) provides peak data rates of up to 19.2 kbps. In summary, wide area wireless data systems provide low data rates to highly mobile users over a large regional area.

Wireless local area networks (WLANs) WLANs are an effort to replace the traditional wired networks found, for example, in today's offices. They provide data communications only within a confined region, *e.g.*, a building or a campus. An advantage of the limited range is the ability to provide high data rates. The IEEE 802.11 specification [4] is an effort to standardize these systems. In summary, WLANs provide low mobility, high speed data communications within a confined region.

Paging/Messaging systems Paging systems provide one way communication capabilities to highly mobile users over a wide coverage area. The transmitters used for broadcasting a message to a user can be highly complex and can also use high powers. A pager is a low complexity, lightweight, and economical device that notifies its user if he has received a telephone call. Current generation pagers can also communicate a short message to the recipient. In summary, paging systems provide one way messaging capabilities over a wide coverage area to highly mobile users.

Satellite based mobile systems Satellite based mobile systems provide, by far, the largest coverage area of all wireless systems. By using satellites as base stations that provide two way communication capabilities with users on earth, satellite based mobile systems provide truly global coverage. However, the large distance between the satellites and the users necessitates high transmit powers making the mobile phones bulky. Moreover, the voice/data quality supported is poor and expensive. In brief, satellite

[1] Metricom's Ricochet [2] is an exception to this rule. It uses a large number of small, inexpensive base stations with low elevation antennas.

based mobile systems provide two way limited quality voice/data/messaging over a global coverage area.

Wireless home networks Wireless home networks [5] interconnect, within the confines of a home, entertainment appliances like televisions, consumer appliances like refrigerators and toasters, as well as co-located equipment like personal computers, printers, wireline telephones, and cellular phones. The Bluetooth radio system [6], which seeks to inexpensively replace cables between devices with self-organizing wireless links, is an effort in this direction. These networks provide coverage only within a confined region and support only low mobility.

Cellular mobile radio systems Cellular radio systems provide two way voice communication capabilities, and also data communications in the next generation systems, even at vehicular speeds over regional or national coverage area. In cellular systems, the coverage area is divided into cells, each cell being handled by a base station (BS). A Mobile Station (MS) located in a cell communicates with the cell's serving BS using channel(s) that are either assigned a priori or dynamically to the cell. We look at these systems in greater detail in Chapter 2. In summary, cellular mobile radio systems provide two way voice and data communications to highly mobile users over a regional or national coverage area.

The above list is not exhaustive. However, it does illustrate the wide disparity between various systems in terms of the levels of mobility supported by them.

Even in a given wireless system, the level of support for various services is also a function of mobility. For example, in next generation cellular system standards, the peak data rate drops from 384 Mbps for slow moving pedestrian users to 144 kbps for fast moving vehicular users [1, 7]. Consequently, the priorities assigned to supporting various applications in next generation standards depend on how fast the user is moving. Knisely *et al.*'s classification [8] in Table 1.1 illustrates this point.

One therefore sees that user mobility, while being an important reason for the popularity of wireless systems, is also a challenging issue to address when designing these systems. This dissertation focuses on analyzing the impact of user mobility on various aspects of

wireless system design. The results are directed towards cellular mobile radio systems, which are affected by mobility in a variety of ways, as discussed in detail in the next section. The ability of cellular radio systems to suitably handle the implications of user mobility will determine whether they can satisfactorily meet the growing demand for wireless services.

Application	Mobile	Semi-Mobile	Fixed
WWW/Internet access	Medium	High	High
Database access	High	High	High
File transfer	Medium	High	High
Electronic mail	Medium	High	High
Broadcast data	Medium	Medium	Medium
Video telephony	Medium	Medium	Medium
Program sound	Low	Low	Low
Asynchronous circuit data	Medium	Medium	Medium
Group 3 fax (faster than 14.4 kbps)	Medium	Medium	Medium

Table 1.1: Data service priority for cellular systems as a function of mobility

1.2 Cellular Resource Allocation and User Mobility

Cellular systems need to be optimized at various levels to be able to achieve the required high spectral efficiencies and handle multiple classes of traffic. The discussion below highlights the many relevant resource allocation issues that arise in designing these systems, and the impact of user mobility on them.

Resources in a channel A variety of resources like transmit power, modulation constellation shape and size, error correction coding scheme, multiple transmit and receive antennas, etc., are available to the system designer to achieve the required quality of service over a wireless channel. User motion, changes in interference, and changes in surroundings make the channel between the base station and the mobile user a time varying channel. Therefore, techniques for dynamically adapting the resources to the link condition become necessary for achieving high spectral efficiencies.

A key factor that determines the efficacy of adaptive schemes is the accuracy of the channel estimates used by them. The faster a user moves, the faster the wireless

channel de-correlates, and the lesser the accuracy of the estimates. In Chapter 3, we analyze the impact of user mobility on one such link adaptation technique.

Channels as resources Cellular systems based on TDMA/FDMA divide the spectrum assigned to them into a number of orthogonal frequency channels and non-overlapping time slots. Different approaches have been proposed to allocate and share these channels among the users. They can be broadly classified into the following two categories.

Packetized channel sharing schemes Packet based transmission schemes, in which the information is transmitted in the form of packets, offer a flexible mechanism to allocate and share channels among different classes of traffic. In addition to being flexible, packetized channel sharing schemes also can be spectrally efficient. This is because multiple bursty data users can be accommodated in the same channel. By making use of the many silence durations that occur during a talk spurt, even multiple voice users can be multiplexed on the same channel. To satisfy the constant rate requirements of voice users, packet based schemes use reservation mechanisms that reserve slots for voice packet transmissions during talk spurts.

When a voice user moves to another cell, it loses its reservation unless the resource allocation scheme actively takes steps to allocate the user a reservation in the new cell. In Chapter 4, we analyze the impact of user mobility on the performance of a packetized reservation scheme that does not make this new reservation.

Dedicated channel assignment schemes Dedicated channel assignment schemes allocate an entire channel[2] to a user for the entire duration of his call [9]. These schemes are suited for handling periodic traffic sources like voice users, who require a constant transmission rate.

When a user moves from one cell to another, his call has to be handed off to the new cell. The base station of the new cell therefore needs to allocate traffic

[2]In two way communications a pair of channels is allocated for uplink and downlink.

and control channels to the user. Failure to do so leads to a dropped call, or a significant degradation in transmission quality due to poor reception of signals transmitted/received at cell boundaries.

The hand-off traffic that cellular systems handle has increased as the cell sizes have become smaller to accommodate more users. Channel allocation techniques that can handle this increased hand-off traffic are therefore necessary. We study the impact of user mobility on such dedicated channel assignment schemes in Chapter 5.

User mobility also leads to other challenges at the network layer [10], which we do not look into. The network layer routes traffic meant for a mobile wireless end host to its current location. For data traffic, this is done by IP addresses, which uniquely identify a particular end host. However, IP addresses, by virtue of their hierarchical addressing scheme[3], are also used to find a route between two end hosts. Putting these two together results in a situation fraught with contradiction for mobile computing. The Mobile IP standard [11] is an effort to address this problem.

1.3 Dissertation Outline

Having presented arguments to show the impact of user mobility on wireless systems, and on resource allocation schemes in cellular radio systems in particular, we now give an outline of this dissertation.

In Chapter 2, we describe cellular systems that are the focus of our study. In particular, we describe the evolution of current and future cellular standards, their architecture, the characteristics of wireless channels on which they operate, and how hand-offs occur.

In Chapter 3, we first analyze the impact of user mobility in the form of a time varying channel between the mobile station (user) and its serving BS. We do so in the context of the link adaptation problem. In link adaptation techniques, users and BSs cooperate with each other to estimate the time varying link condition. The transmission rate on the uplink

[3]This helps reduce the size of look up tables used by routers for routing data traffic.

or downlink or both is then varied as a function of the fed back link estimate. The link adaptation thresholds determine which transmission rate is used for a given link estimate. In case the channel condition is estimated to be below an acceptable threshold, the user may choose not to transmit at all to avoid generating unnecessary interference. We study the impact of channel correlation, which determines the accuracy of the link estimates, on the optimal no-transmission threshold for a co-channel interference-limited TDMA system. We show that the optimal thresholds depend on the channel correlation. Also, the sensitivity of the system's performance to the threshold values depends on how heavily the system is loaded.

In Chapter 4, we analyze the impact of user mobility when it leads to hand-offs from one cell to another. We do so in the context of a packetized reservation based access protocol called Packet Reservation Multiple Access (PRMA) [12], which can simultaneously handle both periodic, delay intolerant and aperiodic, delay tolerant traffic. We analyze the impact of user hand-off rate on the voice packet dropping probability in PRMA. Hand-off of a voice user from one cell to another leads to a premature loss of his reservation, and consequently entails additional packet transmission delays. We shall see that packet errors can also lead to similar transmission delays. Therefore, the impact of packet errors due to channel fading and co-channel interference on PRMA is also analyzed in this chapter. We develop a signal flow graph based analytical technique for this purpose.

In both the above problems, the models we consider assume that a given number of users that have already been admitted into the system are contending with each other for the resources. The admission control mechanism itself plays a significant role in guaranteeing a required quality of service. In Chapter 5, we focus on this admission control problem in the context of dedicated channel assignment schemes. We investigate prediction based heuristic techniques for introducing new call access control for prioritizing hand-offs in unprioritized dedicated channel assignment schemes. We show that using mobility statistics at the new call access stage helps the system to handle non-uniform traffic. We present our conclusions in Chapter 6.

An important issue that arises during analysis is the many different phenomena that take place simultaneously in cellular systems, *e.g.*, fading, interference, traffic arrival and

departure, mobility, etc. This necessitates the use of approximate techniques for analyzing cellular systems in the chapters that follow, and is an interesting and ripe area for research.

Chapter 2 Cellular Mobile Radio Systems

Cellular mobile radio systems – the focus of this dissertation – are the solution to the problem of accommodating thousands of users in a very limited bandwidth. They are designed to provide services to slow moving pedestrians as well as fast moving vehicular traffic. We now describe these systems in more detail.

In Section 2.1, we briefly trace the evolution of cellular systems. This will provide the context for the standards we refer to later. The cellular system basics and architecture are explained in Section 2.2 and Section 2.3, respectively. We describe the path loss models and the time varying multipath fading and attenuation models, which are used in subsequent chapters, in Section 2.4. Hand-offs in cellular systems are described in Section 2.5.

2.1 Evolution of Cellular Systems

Cellular radio systems have evolved significantly in the two decades of their existence. While the first generation (1G) and second generation (2G) cellular systems were designed to primarily handle voice traffic, the next wave of third generation (3G) systems will also offer significant data traffic handling capabilities. The research community has also begun working on the fourth generation systems, which promise to be significantly more sophisticated.

Most cellular radio systems in operation today in the world are 2G systems, but the 1G systems are also in widespread use in the US. The 1G Advanced Mobile Phone Service (AMPS) network, first deployed in 1983 [13], is based on analog technology with FM modulation. The 2G systems instead employ digital modulation and have a provision for implementation of integrated speech and data services. There are multiple 2G standards in use in Europe, US, and Japan. These standards differ primarily in their air interfaces – Time Division Multiple Access/Frequency Division Multiple Access (TDMA/FDMA) or Code Division Multiple Access (CDMA). Global System for Mobile Communications

(GSM) and IS-54/IS-136 are examples of TDMA standards, whereas IS-95 is an example of a CDMA standard. The data handling capacity of 2G systems is limited. As a result, improvements have been proposed for increasing the data rates supported by them. For example, High Speed Circuit Switched Data (HSCSD) and General Packet Radio Service (GPRS) are improvements proposed for GSM. In-depth reviews of the evolution of cellular wireless communications can be found in [8, 14, 15, 16, 17].

Recently, wireline data traffic has grown by leaps and bounds, primarily due to the popularity of the Internet. This in turn has led to an increasing effort to support and handle both data and voice traffic in the 3G wireless systems. Unfortunately, the 3G systems themselves have several competing air interfaces. Universal Mobile Telecommunications Services (UMTS) is the 3G evolution preferred for GSM. It uses Wideband CDMA (WCDMA) as its radio interface technology [1]. Another proposed radio interface standard is cdma2000 [8]. The Enhanced Data Rates for GSM Evolution (EDGE) [18] standard provides an evolutionary path from current 2G TDMA standards to 3G services that lie in the same spectral bands. High Data Rate (HDR) [19] and 1XTREME are similar evolutionary standards for CDMA systems.

2.2 Cellular Radio System Basics

Cellular radio systems utilize the fall-off in power of a transmitted signal with distance to reuse the same frequency channel or time slot at another spatially separated location. The entire coverage area is therefore divided into non-overlapping cells, each cell being controlled by a base station. Using the same frequency in geographically separated cells gives rise to co-channel interference.[1] The cells using the same frequency are spaced far enough apart to prevent the co-channel interference from dominating the signal. In traditional cellular system design, the reuse distance is chosen based on a worst case design criterion that guarantees a minimum Signal to Interference plus Noise Ratio (SINR) in 90–95% of the cell area.

[1] Adjacent channel interference is another source of interference. It arises due to practical, non-ideal filters that cannot completely annul the signals from users using other channels.

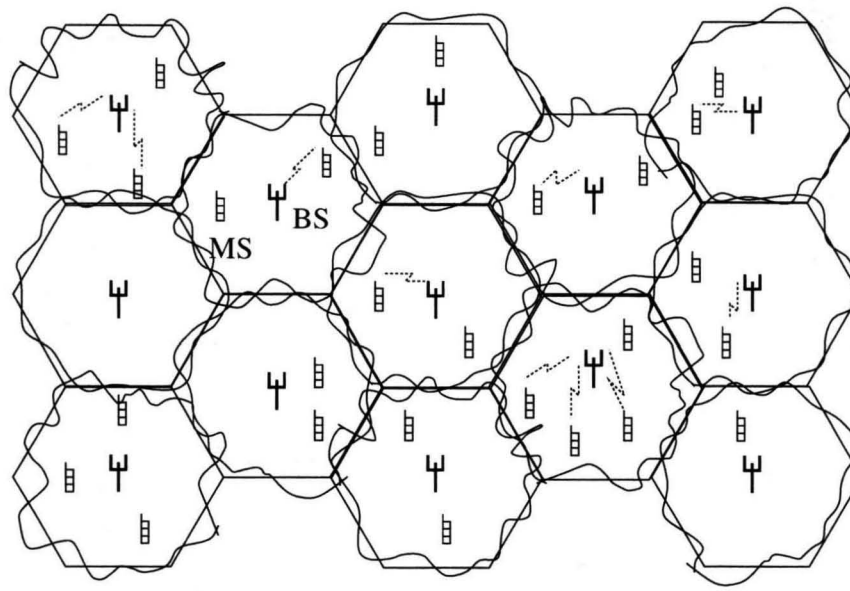

Figure 2.1: Cellular layout: actual coverage shapes and hexagonal footprint approximation

The actual radio coverage area of a cell, over which a minimum quality of service (in terms of SINR, for example) can be guaranteed, is known as its footprint [20, Chp. 2]. The footprint's shape depends on propagation conditions existing in the cell [21]. Hexagonal, circular, and square footprints are typical footprint shape approximations used in the literature analyzing cellular systems. This is because of the well known fact that regular hexagons and squares tessellate, *i.e.*, cover a two-dimensional plane without overlapping. The shapes in real cellular system deployments may vary from these approximations due to shadowing, fading, etc. For example, a study [21] based on power measurements conducted at 900 MHz in Manhattan indicates that the footprint is a concave diamond. A typical cellular layout with a hexagonal footprint is shown in Fig. 2.1. Also shown in the figure are the constant SINR contour lines that determine the actual cell shape.

A key feature of cellular systems is that they are interference-limited; the interference generated by other transmitters is significantly more than thermal noise. While thermal noise can be overcome by increasing the transmit power, doing so does not help in interference-limited systems since the interference power proportionally increases.

2.3 Generic Cellular System Architecture

Cellular radio systems consist of three generic subsystems. Borrowing terminology from the UMTS system architecture, the subsystems are User Equipment (UE), Radio Access Network (RAN), and Core Network (CN).[2] UE holds the subscriber's identity and subscription information. RAN handles the radio interface part of the cellular system. It consists of Base Stations[3] (BS) and Radio Network Controller (Base Station Controller (BSC) in GSM). The aspects related to the wireless channels, which exist between the MSs and BSs, are handled by the RAN. CN handles the switching of calls or traffic between external networks and BSCs. It also consists of databases which carry information about user location and service profile, and a list of visiting users from other cellular networks.

We describe the UMTS architecture in more detail in Appendix 2.A. A detailed description of 2G system architectures and 3G UMTS architecture can be found in [20, Chp. 10] and [1], respectively.

2.4 Channel Characteristics

When electromagnetic waves travel through the environment, they are reflected, scattered, and diffracted by walls, buildings, natural terrain, and other objects. Characterizing this propagation in detail calls for solving Maxwell's equations with boundary conditions determined by the physical characteristics of the obstructing objects [3]. These characterizations are clearly complex, and often the physical characteristics are themselves unknown. Various statistical and deterministic models have therefore been proposed to characterize the wireless channel, and thereby aid in system specification, analysis, and design. We shall use these models in the analysis presented in the following chapters.

The channel is characterized by its path loss due to signal propagation, attenuation due to natural and man-made objects, and fading due to multipath effects. The mathematical models for these are described below.

[2] In GSM, these would be called Mobile Station subsystem, Base Station subsystem, and Network Switching and Operation Support subsystem.

[3] The Base Station is also referred to as Node B in UMTS.

2.4.1 Path Loss Models

Path loss models illustrate the dependence of the received signal power on the distance between the transmitter and the receiver.

Simple analytical model The following power decay profile with distance d is widely used in the literature for analyzing cellular systems because of its simplicity [3].

$$P_r = P_t K \left(\frac{d_0}{d}\right)^{\gamma_0}, \tag{2.1}$$

where P_r is the received signal power, P_t is the transmitted power, and K is a dimensionless constant dependent on the antenna characteristics, wavelength, etc. d_0 is a reference distance and γ_0 is the path loss exponent. For example, in free space propagation, $P_r = P_t \left[\frac{G_l \lambda_c}{4\pi d}\right]^2$, where G_l is the product of the transmit and receive antenna gains in the line of sight (LOS) direction, and λ_c is the wavelength of the transmitted wave. While γ_0 is 2 for free space propagation, it is typically between 3 and 5 for microcellular environments due to multipath effects. γ_0 may itself change with the distance d [3].

Empirical models Several empirical models for path loss, based on actual data collected in various cellular environments, have also been proposed in the literature to accurately factor in variables like receive and transmit antenna heights, tree coverage density, etc. A detailed description of the various models can be found in [22, Chp. 2]. It must be noted that these models change considerably from one carrier frequency to another. They also depend on the terrain in which the measurements were made. To illustrate this point, we discuss in Appendix 2.B two common propagation models that explicitly characterize path loss as a function of carrier frequency, antenna heights, terrain, and distance.

We shall use the simpler analytical model of (2.1) in our analysis because it captures the essence of signal propagation in cellular systems without resorting to highly context specific and complicated empirical models.

2.4.2 Attenuation Models

Log-normal shadowing Log-normal shadowing models the variation in outdoor radio propagation environments caused by attenuation due to obstacles like buildings and natural objects. The probability distribution of the received signal power ψ is given by [3]

$$p(\psi) = \frac{\xi}{\sqrt{2\pi}\sigma_\psi \psi} \exp\left(-\frac{(10\log_{10}(\psi) - \mu_\psi)^2}{2\sigma_\psi^2}\right), \ \psi > 0, \tag{2.2}$$

where $\xi = 10/\ln(10)$, μ_ψ is the mean of $\psi_{\text{dB}} = 10\log_{10}(\psi)$, and σ_ψ^2 is the variance of ψ_{dB}.

2.4.3 Multipath Fading Models

Reflections, scattering, and diffraction from natural and man-made objects in the vicinity of the BS and MS, and possibly direct line of sight propagation, cause multiple distorted copies of the transmitted signal, with different propagation delays, to arrive at the receiving antenna. Small changes in the delays of the arriving signals can cause large changes in their phases. The constructive and destructive interference of these signals results in noticeable changes in received signal amplitude and phase. This phenomenon is called multipath fading [22, Chp. 2][20, Chp. 4][3].

We first list the different one-dimensional probability distributions for the channel fade amplitude sampled at a given instant in time. Having done so, we then give the corresponding two-dimensional probability distributions that model the time correlation in the channel as a function of user speed.

Rayleigh fading The Rayleigh fading model applies to a scenario where there is no LOS path between the transmitter and receiver antennae. The received complex baseband low pass signal is then modeled as a zero mean complex Gaussian random process. The received complex envelope $z(t)$ has a distribution given by [22, Chp. 2],[3]

$$p_z(\alpha) = \frac{2\alpha}{\Omega} \exp\left(-\frac{\alpha^2}{\Omega}\right), \ \alpha \geq 0, \tag{2.3}$$

where $\Omega = E[z^2]$. The phase of the received signal $\phi(t)$ is a uniformly distributed random variable over the interval $[-\pi, \pi)$, and its distribution takes the form

$$p_\phi(x) = \begin{cases} \frac{1}{2\pi}, & -\pi \leq x < \pi \\ 0, \text{ otherwise} \end{cases} . \tag{2.4}$$

We use the Rayleigh distribution in Chapter 4 when we compute expressions for the probability of packet error as a function of SIR.

Ricean fading When there does exist a LOS component, the in-phase and quadrature components of the received signal are modeled as non-zero mean Gaussian processes. The complex envelope has a Ricean distribution of the form [22, Chp. 2][3]

$$p_z(\alpha) = \frac{2\alpha(K+1)}{\Omega} \exp\left(-K - \frac{(K+1)\alpha^2}{\Omega}\right) I_0\left(2\alpha\sqrt{\frac{K(K+1)}{\Omega}}\right), \ \alpha \geq 0, \tag{2.5}$$

where $\Omega = E[z^2]$, $I_0(.)$ is the 0^{th} order Modified Bessel function [23, Chp. 8], and K is called the Rice factor. K is the ratio of the power of the LOS components to that of the scattered components. $K = 0$ corresponds to Rayleigh fading, and as $K \to \infty$ the distribution becomes an impulse (no fading). The probability distribution of the phase of the received signal is given by (2.4).

Nakagami fading The Nakagami fading model is an empirical model that provides a closer match to experimental data than either Rayleigh or Ricean distributions in a wide variety of environments. The received envelope follows a distribution given by [24][22, Chp. 2]

$$p_z(\alpha) = \frac{2m^m \alpha^{2m-1}}{\Gamma(m)\Omega^m} \exp\left(-\frac{m\alpha^2}{\Omega}\right), \ m \geq \frac{1}{2}, \tag{2.6}$$

where $\Omega = E[z^2]$. m is called the Nakagami fading parameter; $m = 1$ corresponds to Rayleigh fading, while $\frac{1}{2} \leq m < 1$ corresponds to fading conditions more severe than Rayleigh fading. The distribution becomes an impulse as $m \to \infty$. The Rice distribution can also be closely approximated by means of a mapping between K and

m [22, Chp. 2]. The phase distribution is assumed to be uniform, as in (2.4).

We use the above Nakagami model for fading channels in Chapter 3.

Envelope correlations

We now list the joint two-dimensional distributions for some of the statistical fading models given above.

Nakagami fading Let z_1 and z_2 denote the envelopes of the received signal at two time instants Δt apart. Let ρ denote the correlation coefficient [25, Chp. 7] between z_1 and z_2. The wide sense stationary joint distribution of z_1 and z_2 is given by [24]

$$\begin{aligned} p_{z_1,z_2}(\alpha_1,\alpha_2) &= \frac{4(\alpha_1\alpha_2)^m}{(1-\rho)\Gamma(m)\rho^{(m-1)/2}}\left(\frac{m}{\Omega}\right)^{m+1} \\ &\quad \times I_{m-1}\left(\frac{2m\sqrt{\rho}\alpha_1\alpha_2}{(1-\rho)\Omega}\right)\exp\left(-\frac{m(\alpha_1^2+\alpha_2^2)}{(1-\rho)\Omega}\right), \end{aligned} \tag{2.7}$$

where $I_{m-1}(.)$ is the $(m-1)^{\text{th}}$ order modified Bessel function of the first kind [23, Chp. 8], $\Gamma(m)$ is the Gamma function [23, Chp. 8], ρ is the channel correlation coefficient, m is the Nakagami fading parameter, and $E[z_1^2] = E[z_2^2] = \Omega$.

Channel correlation coefficient: In an isotropic scattering model[4], which occurs for $m = 1$ (Rayleigh fading), the correlation coefficient can be shown to be [27, Chp. 1]

$$\rho = \frac{\pi}{4-\pi}\left({}_2F_1\left[-\frac{1}{2},-\frac{1}{2};1;J_0^2(2\pi f_d\Delta t)\right]-1\right), \tag{2.8}$$

where f_d is the Doppler spread, Δt is the sampling time interval of the envelope of the signal, $J_0(.)$ is the 0^{th} order Bessel function of the first kind [23, Chp. 8], and ${}_2F_1[.,.;.;.]$ is the hyper-geometric function [23, Chp. 9]. This expression is well approximated by [27, Chp. 1]

$$\rho = J_0^2(2\pi f_d\Delta t). \tag{2.9}$$

[4]In an isotropic scattering model, the multipath plane waves incident on the receiving antenna arrive uniformly from all directions. This model is typically valid in microcells, for example, when the BS and the MS are at the same height. In macrocells, where the BS's antenna is at a higher elevation, the isotropic scattering assumption is typically valid only for the down link [26].

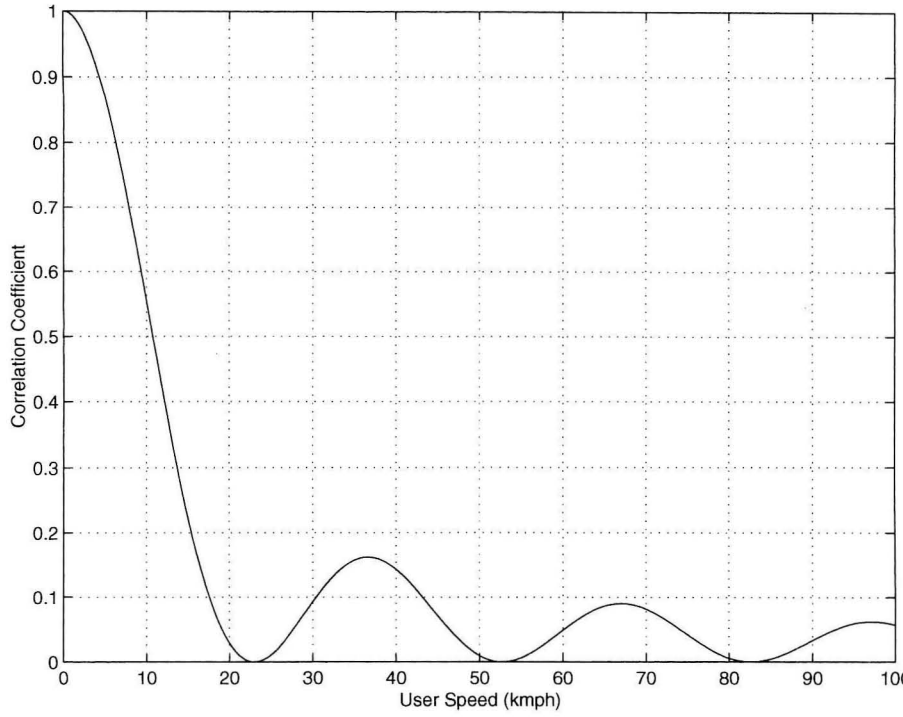

Figure 2.2: Channel correlation *vs.* user speed for $f_c = 1800$ MHz and $\Delta t = 10$ ms

The Doppler spread f_d, as a function of user speed v, is given by

$$f_d = \frac{v}{\lambda_c},$$

where λ_c is the wavelength of the transmitted signal. The correlation coefficient is plotted as a function of user speed v in Fig. 2.2 for $f_c = 1800$ MHz and $\Delta t = 10$ ms.

It must be noted that the formula for ρ in (2.8) is exact only for Rayleigh fading (Nakagami parameter $m = 1$). For other values of m, this is an approximation.

Log-normal shadowing Experimental measurements at 1700 MHz by Gudmundson [28] indicate for log-normal shadowing a wide sense stationary distribution with an exponential autocorrelation function $R_z(.)$ of the form

$$R_z(\Delta t) = \sigma_\psi^2 \exp\left(\frac{v\Delta t}{X_c}\right), \tag{2.10}$$

where v is the speed of the mobile and Δt is the difference in times when the envelope

is measured. X_c is called the de-correlation distance, and it is typically equal to the average size of the attenuating objects.

2.5 Hand-offs in Cellular Systems

Hand-off is the mechanism that transfers a MS's ongoing call from the current cell's BS to another cell's BS. In case the MS moves to another cellular system controlled by another operator, an inter-system hand-off is said to be carried out. The two systems need to be compatible and have agreements in place for this to occur.

Chapters 4 and 5 deal with the impact of hand-offs on the performance of a packetized statistical multiplexing scheme and dedicated channel assignment scheme, respectively. We therefore describe below the basis on which hand-offs occur in cellular systems.

2.5.1 Hand-off Initiation

Hand-offs are initiated when the quality of the link between the MS and the serving BS, to which the MS is connected, becomes unacceptable [22, Chp. 10]. Several parameters like SIR, Received Signal Strength (RSS), Bit Error Rate (BER), distance, and traffic load can be used to estimate the link quality. For example, in hand-off algorithms that use signal strength measurements as link quality estimates, a hand-off is made when the signal strength of the target BS exceeds that of the serving BS by at least H dB [22, Chp. 10].

In 1G systems, the link estimation is done solely by the BS. In 2G systems, the mobiles also assist in the hand-off decision process, leading to more reliable link estimates. The MS measures the signal power it receives from surrounding BSs, and reports them to the serving BS. The hand-off algorithm then calculates time averages of these signal strengths, and makes the hand-off decision. It must be noted that the window length over which the link quality measurements are averaged plays an important role in determining the performance of the hand-off decision algorithm. Adapting the window lengths to the speed of the MS can make the hand-off algorithms robust to changes in link quality due to changes in signal propagation [22, Chp. 10][29].

2.5.2 Cell Dwell Time

An important parameter that influences the performance of resource allocation schemes is the cell dwell time[5], defined as the time spent by a mobile in a cell. Note that cell dwell time is not the same as the duration of a call since a user may move through many cells during the course of his call. Cell dwell time statistics are determined by factors like user speed, cell shape and size, direction of roads inside the cell, etc. As a result, different cell dwell time probability distributions can be found in the literature. We list them below.

Guérin [30], based on simulations under some general assumptions of a cellular system with hexagonal footprints, showed that a negative exponential distribution accurately characterizes the cell dwell time distribution. His analytical model, derived under certain simplified vehicle motion assumptions, corroborates this result. Hong and Rappaport [31] evaluated the cell dwell time based on a model where a mobile's speed and direction in a cell are independent from one cell to another. They arrived at a different distribution for the dwell time. However, they also found that the negative exponential distribution was a good approximation.

On the other hand, analyses of data measured in cellular systems in operation in various parts of the globe have yielded data that buck the trend. Jedrzycki and Leung [32] analyzed data collected from a number of cell sites in British Columbia, Canada. They reported that the log-normal distribution, not the exponential distribution, fit their real traffic data with a high confidence level. Similarly, Barceló and Jordán [33] conducted a field study of channel occupancy intervals in a cellular system in Barcelona. They also reported that the log-normal distribution or a mixture of log-normal distributions was a much better fit than the exponential distribution. Orlik and Rappaport [34] have proposed a sum of hyper-exponential distributions to characterize the cell dwell time distribution.

In our study, we use the analytically simpler exponential distribution for the cell dwell time. The analytical simplicity afforded by this model makes it popular in the literature [35, 36, 37, 38].

[5]The cell dwell time is also referred to as the channel holding time. We avoid using the latter term since the schemes we consider may reassign channels allocated to calls in progress.

Appendix

2.A UMTS Cellular System Architecture

The 3G UMTS architecture is shown in Fig. 2.3 [1, Chp. 5][39]. The network elements are grouped into User Equipment (UE), UMTS Radio Access Network (UTRAN), and Core Network (CN). UTRAN handles all the radio-related functionality, while CN is responsible for switching and routing calls and data connections to external networks.

UE consists of the radio terminal, which is used by subscribers to communicate with UTRAN, and the UMTS Subscriber Identity Module (USIM) that holds the subscriber's identity and his subscription information, and performs authentication algorithms, etc., at the terminal.

The UTRAN consists of one or more Radio Network Subsystems (RNS). A RNS consists of a single Radio Network Controller (RNC) and one or more Node Bs. The RNC owns and controls the radio resources of UTRAN. It is responsible for outer loop power control and hand-over decisions. It may also perform functions like macro-diversity combining and splitting. Node B performs functions like channel coding, interleaving, rate adaptation, spreading, and inner loop power control.

CN consists of Home Location Register (HLR), Mobile Switching Center/Visitor Location Register (MSC/VLR), Gateway Mobile Switching Center (GMSC), Gateway General Packet Radio Service Support Node (GGSN), and Serving General Packet Radio Service Support Node (SGSN). HLR is a database located in the user's home system that stores the master copy of the user's service profile. The information stored about a user is created when he subscribes to the system, and is retained as long as the subscription is active. MSC is the switch that serves the circuit switched services of UE. Information about which MSC a roaming user is in is stored in the HLR. The VLR database holds a copy of the service profile of all visiting users and their exact location within the serving system. Visiting users are users that are subscribed to other MSCs. GMSC is the switch which connects the UMTS network to external circuit switched networks. The GGSN is GMSC equivalent for packet switched services. It is the one which connects the UMTS network to packet

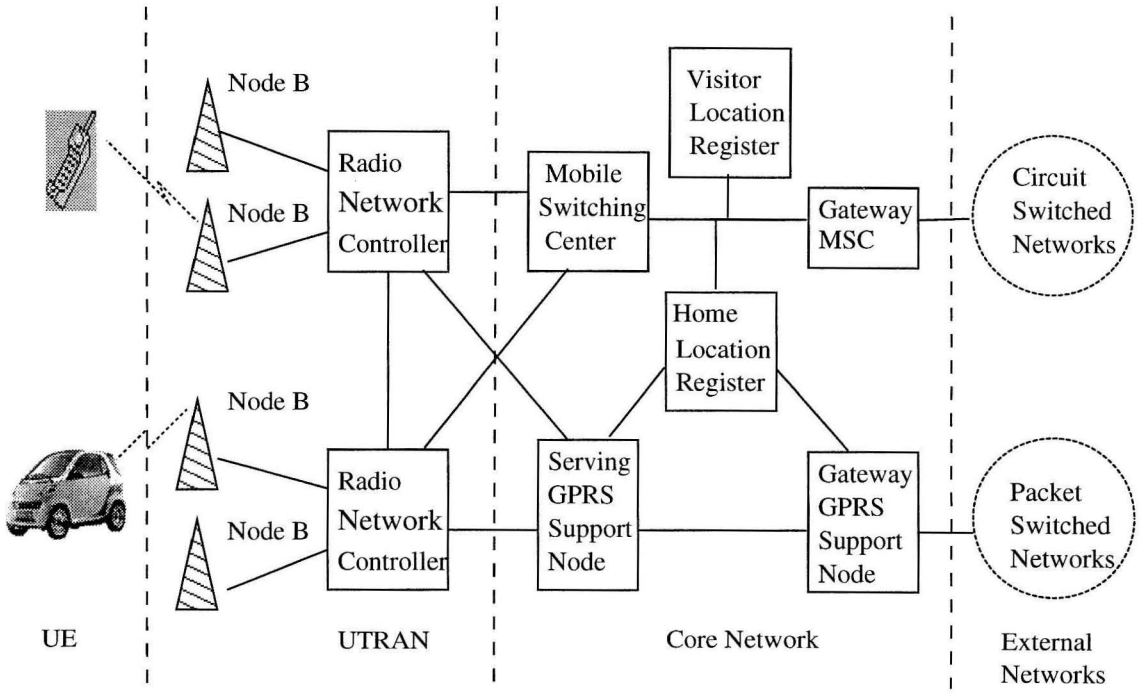

Figure 2.3: Cellular system architecture

switched networks like the Internet. Similarly, the SGSN is the VLR/MSC equivalent for packet switched services.

2.B Two Empirical Path Loss Models

COST231-Hata model As per the COST231-Hata model for use in the 1500 MHz–2000 MHz frequency range for outdoor microcells, the path loss L_p in dB, defined as $10\log_{10}(P_t/P_r)$, is given by

$$L_p = A + B\log_{10}(d) + C, \tag{2.11}$$

where

$$\begin{aligned} A &= 46.3 + 33.9\log_{10}(f_c) - 13.82\log_{10}(h_b) - \alpha(h_m), \\ B &= 44.9 - 6.55\log_{10}(h_b), \end{aligned}$$

$$C = \begin{cases} 0, & \text{medium city and suburban areas} \\ & \text{(with moderate tree density)} \\ 3, & \text{metropolitan centers} \end{cases}.$$

Here, h_b is the BS's antenna height, h_m is the MS's antenna height, f_c is the carrier frequency in MHz, and d is the distance between the BS and MS.

Okumura and Hata model On the other hand, the path loss in dB according to the Okumura and Hata model for 150 MHz $\leq f_c \leq$ 1000 MHz cellular systems is given by

$$L_p = \begin{cases} A + B\log_{10}(d), & \text{for urban area} \\ A + B\log_{10}(d) - C, & \text{for suburban area} \\ A + B\log_{10}(d) - D, & \text{for open area} \end{cases}, \tag{2.12}$$

where

$$\begin{aligned} A &= 69.55 + 26.16\log_{10}(f_c) - 134.82\log_{10}(h_b) - a(h_m), \\ B &= 44.9 - 6.55\log_{10}(h_b), \\ C &= 5.4 + 2\left(\log_{10}\left(\frac{f_c}{28}\right)\right)^2, \\ D &= 40.94 + 4.78\left(\log_{10}(f_c)\right)^2 - 19.33\log_{10}(f_c), \\ a(h_m) &= \begin{cases} (1.1\log_{10}(f_c) - 0.7)h_m - (1.56\log_{10}(f_c) - 0.8), & \text{medium/small city} \\ 8.28\left(\log_{10}(1.54h_m)\right)^2 - 1.1, \ f_c \geq 400 \text{ MHz}, & \text{large city} \\ 3.2\left(\log_{10}(11.75h_m)\right)^2 - 4.97, \ f_c < 400 \text{ MHz}, & \text{large city} \end{cases} \end{aligned}$$

As in the previous model in (2.11), h_b is the BS's antenna height, h_m is the MS's antenna height, f_c is the carrier frequency in MHz, and d is the distance between the BS and MS.

Chapter 3 Link Adaptation

Link adaptation techniques intelligently vary the rate at which data is transmitted on the wireless link as a function of the link's condition. In this chapter, we analyze a link adaptation technique that adapts the data transmission rate by changing the modulation and error correction coding scheme, and operates in a co-channel interference-limited cellular environment.

3.1 Chapter Organization

Before delving into link adaptation any further, we first give an outline of this chapter. Link adaptation is introduced in Section 3.2. The link adaptation threshold problem and our work are described in Section 3.3, and the system model is explained in Section 3.4. The analysis is given in Section 3.5 and Section 3.6. The results are presented in Section 3.7, and the conclusions in Section 3.8.

3.2 Introduction to Link Adaptation

The cellular systems in the 1G and 2G standards were designed to maximize coverage, *i.e.*, guarantee a minimum Signal to Interference plus Noise Ratio (SINR), for voice users [40]. This resulted in higher SINR for users closer to the base station (BS).[1] The better channel conditions closer to the BS can be exploited to provide higher transmit rates for data users, thus improving the spectral efficiency of the system. Such an idea can be found in Reuse Partitioning (RP) [41], a technique to increase the spectral efficiency of a system handling voice users.[2] However, there is no rate adaptation in RP, and it does not dynamically exploit

[1] This is under the assumption that no power control is employed to compensate for the link condition.

[2] In RP, every cell is divided into concentric zones. Each zone is associated with an overlaid cell plan. The reuse distance for frequencies used in the zones decreases as the zones get closer to the BS. RP therefore exploits the better channel conditions available closer to the BS by reducing the reuse distance.

the changes in SINR due to changes in fading and interference.

Dynamic rate adaptation is achieved in TDMA systems by link adaptation schemes that change the modulation constellation size, the error correction code redundancy, and the number of time slots allocated to each user. For example, in EDGE [18] two modulation schemes are used (8PSK [42, Chp. 3] and GMSK [43]), and the coding rate is varied[3] from 0.49 to 1 (uncoded). In CDMA systems, the spreading code length can be varied and multiple spreading codes can also be allocated to users. For example, in cdma2000 [40] two error correction codes of rate 1/2 and 1/3 are available, and the spreading factor can be changed from 2 to 64. In addition to increasing the spectral efficiency of the system, rate adaptation techniques also increase the peak data rates available to users, thereby enabling the cellular systems to handle applications with burstier traffic requirements.

Link adaptation techniques also exploit the high tolerance to retransmission delay of data users. The cellular system can operate at higher error rate levels and use Automatic Retransmission (ARQ) schemes [44, Chp. 15] to get the data through to the receiver. It is also possible for them to use Incremental Redundancy (IR) [45], which is a Type II Hybrid ARQ scheme [46]. In IR, the information is first sent with minimal coding. In case the decoding fails, additional coded bits are sent until it succeeds.

3.3 Link Adaptation Thresholds

Link adaptation techniques use estimates of SINR, frame error rate (FER), or bit error rate (BER) to choose the modulation and error coding scheme with which to transmit data. We shall refer to a combined modulation and error coding scheme as a **transmission mode** from now on. The link adaptation thresholds determine which transmission mode is used for a given link estimate. The performance of a link adaptation technique therefore crucially depends on its thresholds. A conservative choice for the thresholds that favors using low error rate, high redundancy, and thereby, low transmit rate schemes, instead of high error rate, high transmit rate schemes, leads to an inefficient utilization of resources. On the other hand, an aggressive choice for thresholds that favors high error rate, high transmit

[3]Different error correction codes are obtained by puncturing a parent rate 1/3 convolutional code.

rate schemes leads to unnecessary packet errors and retransmissions.

In this chapter, we consider an adaptation technique that dynamically chooses its transmission mode based on SINR estimates. Practical implementations typically use FER or BER instead of SINR as the link quality estimator [47]. This is because estimating SINR is computationally more expensive than estimating either FER or BER. However, FER, BER, and SINR are equivalent in our case since we assume error-free, albeit delayed, feedback in our model, which is described later. In addition, in interference-limited cellular systems thermal noise is insignificant when compared to interference. We therefore study an SIR (Signal to Interference Ratio) driven link adaptation technique in this chapter.

Various heuristic techniques have been proposed in the literature for determining the link adaptation thresholds. For a given SIR estimate $\widehat{\mathrm{SIR}}$, Furuskär, Mazur, Müller, and Olofsson [18] use the mode i that maximizes a 'net throughput' function[4]

$$i = \arg\max_j R^{(j)}(1 - \mathrm{BLER}^{(j)}(\widehat{\mathrm{SIR}})), \tag{3.1}$$

where $R^{(j)}$ and $\mathrm{BLER}^{(j)}(.)$ are the information transmission rate and the burst error rate function, respectively, for mode j. The intuition behind this technique is to maximize the number of successfully transmitted bits. It is for this reason that a mode's rate $R^{(j)}$ is multiplied by a factor which decreases as the error rate $\mathrm{BLER}^{(j)}(.)$ increases. As pointed out in [49], this formula does not correctly take into account the impact of packets which cannot be decoded correctly. In fact, these data packets remain in the system and need to be retransmitted. This leads to an increase in interference and, hence, an increase in the packet loss rate itself.

Qiu and Chuang [49] instead proposed that the transmission mode should be chosen using the modified rule:

$$i = \arg\max_j R^{(j)}(1 - \mathrm{BLER}^{(j)}(\mathrm{SIR}_e^{(j)})), \tag{3.2}$$

where $\mathrm{SIR}_e^{(j)}$ is given by the implicit equation $\mathrm{SIR}_e^{(j)} = \widehat{\mathrm{SIR}}(1 - \mathrm{BLER}^{(j)}(\mathrm{SIR}_e^{(j)}))$. They

[4]A similar idea has also been proposed for the European RACE R2084 Advanced Time Division Multiple Access (ATDMA) system by Dunlop, Irvine, and Cosimini [48].

arrived at this modification based on an approximation that assumed a linear relation between the expected number of retransmissions and the interference seen by a user. Their approximation also neglected the interactions between users transmitting in different modes.

One of the contributions of [49] is the introduction of an additional mode called Mode 0, in which the user does not transmit if he estimates his link quality to be poor. The usefulness of the no-transmission mode is well understood for a single user fading channel with Gaussian noise. In particular, Goldsmith and Varaiya [50] showed[5] that under the optimal rate and power adaptation policy for single user fading channels, a user does not transmit if the channel fade is below a cut-off value. However, the impact of Mode 0 on an interference-limited system is not clear, and it is this situation that we address in this chapter.

It is difficult to a priori define formulae, like those in (3.1) and (3.2), for choosing the optimum transmission mode given a link estimate, which account for all the interactions taking place in the cellular system between channel fading, packet errors, retransmissions, interference, and data traffic entering the system. The heuristic formulae in (3.1) and (3.2) also do not take into account the impact of channel correlation on the optimal link adaptation thresholds. It is channel correlation that ultimately determines how accurately an SIR estimate predicts its actual value. What determines channel correlation is user speed and delay in feeding back estimates – two significant parameters in cellular system design.

In this chapter, we analyze Mode 0 in two different traffic scenarios, and find the optimal link adaptation thresholds. We study the impact of channel correlation on the optimal thresholds. In the first scenario, packets arrive in the users queues according to a given stochastic process, and stay there until they are successfully transmitted. The system considered is stable, in that the users' queues are stable, *i.e.*, the queues always have only a finite number of packets to transmit. We derive expressions for the average packet waiting time, which depends on how much traffic arrives into the system, channel fading statistics, user locations, and link adaptation thresholds. We use this as the metric for comparing the performance of different systems.

[5]It must be pointed out that these information theoretic results were derived assuming perfect, *i.e.*, instantaneous and accurate, side information at the transmitter and receiver.

The second scenario that we consider is the saturated queues scenario, in which users always have packets to transmit, *i.e.*, their queues contain an infinite number of packets. It must be pointed out here that the average packet waiting time cannot be defined for this scenario because of the infinite number of packets in a user's queue. Consequently, the metric for evaluating the overall system performance is the average system throughput, which is defined as the number of packets successfully transmitted per unit time in the system. We derive expressions for the average system throughput as a function of the channel fading statistics, user locations, and link adaptation thresholds.

The impact of channel correlation turns out to be very different for these two scenarios. The analysis is done for a co-channel interference-limited TDMA cellular system with the users facing similar path loss and fading models. We discuss some limitations of our analysis in Section 3.8.

3.4 System Model

We now describe in detail the system model employed for studying the link adaptation problem. In particular, SIR's dependence on channel fading and co-channel interference is described. The relationship between a given modulation scheme and error correction code, and the burst error probability is then given. Finally, we describe the estimator model used for estimating the link condition.

A hexagonal cellular layout with a given frequency reuse pattern is considered. The n users in each cell are located on a circle of radius r from their serving BSs. Placing the users on a circle ensures that they face the same path loss attenuation.

It is possible to consider link adaptation on the uplink as well as the downlink. For the downlink, the BS can by itself choose which user's burst to transmit. For the uplink, however, only one user, among the many possible contenders, can transmit in a given time slot. Coordination by the BS is therefore required. Except for this difference, the problem set up, in terms of the number of possible interferers, is similar for both uplink and downlink. The analysis we present in this paper is for the uplink. We assume that the coordination mechanism is already in place, for example, by means of a separate broadcast control channel,

and that one user among the possible contenders is chosen randomly by the BS to transmit.

We call the amount of information transmitted in a time slot a **burst**. Based on the transmission mode, multiple packets can get transmitted in a burst.[6] No power control is assumed. Let $L^{(i)}$ denote the number of symbols transmitted in a Mode i transmission burst.

SIR The SIR at the receiving BS at time t is given by

$$\mathrm{SIR}_t = \frac{\alpha_t^2}{\mathcal{I}_t} \frac{P}{r^{\gamma_0}}, \tag{3.3}$$

where P is the transmitted power, r is the distance of the user from its BS, γ_0 is the path loss exponent, and α_t and $\mathcal{I}_t$ are the channel fade amplitude and interference, respectively, at time t.

Channel fading The Nakagami channel fading model, given in (2.7), is used. The fade is assumed to remain constant over the duration of transmission of the burst. We give the model below in terms of the notation we use in this chapter. Let $\alpha_{t-\tau}$ and α_t be the channel fade amplitudes at time instants[7] $t-\tau$ and t, respectively. The joint pdf $p(\alpha_{t-\tau}, \alpha_t)$ is given by [24]

$$\begin{aligned} p(\alpha_{t-\tau}, \alpha_t) &= \frac{4(\alpha_{t-\tau}\alpha_t)^m}{(1-\rho)\Gamma(m)\rho^{(m-1)/2}} \left(\frac{m}{\Omega}\right)^{m+1} \\ &\quad \times I_{m-1}\left(\frac{2m\sqrt{\rho}\alpha_{t-\tau}\alpha_t}{(1-\rho)\Omega}\right) \exp\left(-\frac{m(\alpha_{t-\tau}^2+\alpha_t^2)}{(1-\rho)\Omega}\right), \end{aligned} \tag{3.4}$$

where $I_{m-1}(.)$ is the $(m-1)^{th}$-order modified Bessel function of the first kind [23, Chp. 8], $\Gamma(m)$ is the Gamma function [23, Chp. 8], ρ is the channel correlation, m is the Nakagami fading parameter, and $E\left[\alpha_{t-\tau}^2\right] = E\left[\alpha_t^2\right] = \Omega$. Ω is normalized to 1. The marginal probability distribution for the fade ampli-

[6]Note that this terminology is different from that used in the standards where a packet is divided into multiple blocks, and each block consists of four bursts.

[7]For notational convenience, the time index t is an integer. It indexes frames.

tude then takes the form

$$p(\alpha_t) = \frac{2}{\Gamma(m)} \left(\frac{m}{\Omega}\right)^m \alpha_t^{2m-1} \exp\left(-\frac{m\alpha_t^2}{\Omega}\right). \tag{3.5}$$

The channel fading processes are assumed to be independent from one user to another.

The log-normal shadowing model with its empirically determined exponential autocorrelation function, described in Section 2.4.2, is another possible probability distribution that can be considered. However, the analytical intractability of this model prevents us from studying it.

Interference In addition to channel fading, the SIR is determined by the interference at the receiving BS. We consider only the first tier interferers since they are the dominant source of interference.[8] If k_t ($0 \leq k_t \leq 6$) cells interfere at time t, the interference $\mathcal{I}_t$ is given by

$$\mathcal{I}_t = \sum_{i=1}^{k_t} \frac{P}{d_i^{\gamma_0}}, \tag{3.6}$$

where d_i is the distance of the transmitting user in the i^{th} interfering cell from the receiving BS. A typical interference[9] scenario is shown in Fig. 3.1 for reuse cluster size 3. For analytical tractability, the fading of the interference signals is not considered in our model; the impact of this fading remains to be studied.

Eqn. (3.6) can be approximated by

$$\mathcal{I}_t = k_t \frac{P}{d_{\text{av}}^{\gamma_0}}, \tag{3.7}$$

[8]For $\gamma_0 = 2$, the decrease in interference power due to free space path loss is compensated by the increase in the number of interferers. The first tier approximation is then inaccurate. However, as mentioned in Section 2.4.1, γ_0 lies between 3 and 5 for cellular systems.

[9]In case antenna sectoring [20] with 3 sectors per cell is used, the number of similar first tier interferers is still 6 for reuse cluster sizes greater than 1. Therefore, the above formula carries through for this case. However, for reuse cluster size 1, antenna sectoring gives rise to two sets of first tier interferers at two different average distances from the receiving BS. Interference is then characterized by a two-dimensional random variable $(k_t^{(1)}, k_t^{(2)})$, where $k_t^{(1)} + k_t^{(2)} = k_t$.

where d_{av} is the average interfering user distance from the receiving BS. For the purpose of approximating interference, d_{av} is best evaluated using a harmonic type mean (instead of an arithmetic mean) from $\frac{1}{d_{\text{av}}^{\gamma_0}} = \frac{1}{6}\sum_{i=1}^{6}\frac{1}{d_i^{\gamma_0}}$.

Unlike channel fading statistics, interference statistics cannot be specified a priori. This is because the interference statistics are themselves dependent on other system parameters. We shall look at this in greater detail in Section 3.5.2, where we derive the interference statistics.

From (3.3) and (3.7), the expression for SIR_t is therefore

$$\text{SIR}_t = \frac{\alpha_t^2}{k_t}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}. \tag{3.8}$$

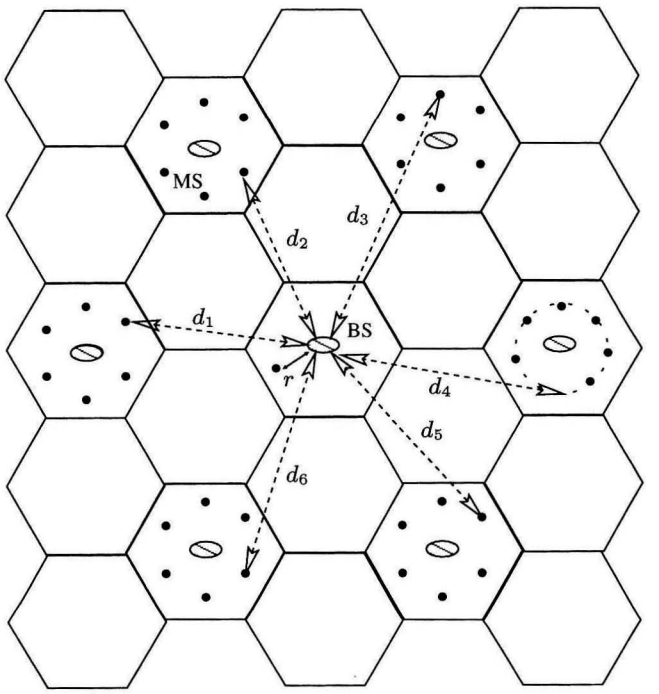

Figure 3.1: Co-channel interference from first ring of interferers

Symbol error probability The probability of symbol error as a function of SIR γ, for a transmission mode with MQAM or MPSK as the modulation scheme, is well approx-

imated by an exponential function[10] [51, 52] of the form $c_1^{(i)} \exp(-c_2^{(i)}\gamma)$, where $c_1^{(i)}$ and $c_2^{(i)}$ depend on the modulation scheme used in Mode i.

Burst error probability The burst error probability $P_E^{(i)}(\gamma)$ for a given SIR γ depends on the error correction code. We consider two standard approaches for modeling the performance of an error correction code.[11]

Coding gain based In this approach, the use of an error correction code is assumed to be equivalent to increasing the uncoded SIR by a factor $g^{(i)}$, which is the *coding gain* [42, Chp. 5] of the code used in Mode i. For analytical tractability, the coding gain $g^{(i)}$ for a given code is taken to be independent of SIR.[12] $P_E^{(i)}(\gamma)$ is then given by

$$P_E^{(i)}(\gamma) = 1 - \left[1 - c_1^{(i)} \exp(-c_2^{(i)} g^{(i)} \gamma)\right]^{L^{(i)}}. \tag{3.9}$$

Weight enumerator based An approach based on the union bound approximation [44, Chp. 12] that uses weight enumerators of codes also leads to expressions for burst error probability that are weighted sums of exponentials.

In both the above approaches, $P_E^{(i)}(\gamma)$ can be written as a weighted sum of exponentials, and the analysis is similar. We shall use the first approach in our analysis and results.

3.4.1 Channel Estimator

The values of SIR measured in the previous slots, corresponding to time instants $t - \tau$ and earlier, are assumed to be fed back error free to the user. Note that this is an idealization

[10] The exponential function approximation is not very accurate at low SIR. However, for analytical tractability, we use this function for all SIR values.

[11] We do not model Incremental Redundancy (IR) in this chapter. In IR, the information is first transmitted with minimal coding. Additional parity bits are transmitted only if the decoding at the receiver fails. This process is continued until the decoding succeeds.

[12] In practice, the coding gain for an error correction code varies with SIR or, equivalently, with burst error probability.

since the estimates, in addition to being delayed by τ time instants, will also have measurement errors that are not considered here. The feedback delay is typically a multiple of some number of frames. The estimator generates an estimate $\widehat{\mathrm{SIR}}_t$ for the SIR that a user shall see at time t. We consider a first order linear predictor model in which [18]

$$\widehat{\mathrm{SIR}}_t = \mathrm{SIR}_{t-\tau}. \tag{3.10}$$

The analysis can be easily modified to handle estimators of type $\widehat{\mathrm{SIR}}_t = \Psi(\mathrm{SIR}_{t-\tau})$, where $\Psi(.)$ is any invertible function. Analysis of higher order estimators requires, at the very least, higher order statistics for fading.

3.4.2 Link Adaptation Rule

Let $\{0, 1, \ldots, H\}$ denote the set of modes a user can be in. A given user will choose to be in Mode 0 at time t if his SIR estimate $\widehat{\mathrm{SIR}}_t$ satisfies

$$\widehat{\mathrm{SIR}}_t < l_0. \tag{3.11}$$

A user chooses Mode i, in which i packets get transmitted in a burst, if

$$l_{i-1} \leq \widehat{\mathrm{SIR}}_t < l_i. \tag{3.12}$$

Note that for the highest mode H, $l_H = \infty$.

3.5 Stable Queues Scenario

Every user maintains a queue (with an infinite buffer) in which the packets to be transmitted to the BS arrive with mean rate λ packets/frame and second moment V. A packet stays in the user's queue until it is successfully transmitted, *i.e.*, received and decoded correctly by the BS. To transmit a burst, a user must

1. Have a packet to transmit,

2. Not be in Mode 0, and

3. Be chosen by the BS to transmit from among the other users in the cell who also satisfy the first two criteria. We assume that the BS chooses randomly one user from the users that want to transmit, and allows him to transmit his burst.

Terminology: If a user satisfies the first two criteria, we shall say that he **wants to** transmit. We shall say that a user is **allowed to** transmit if he meets all the above three criteria.

Figure 3.2 gives a pictorial illustration of the operation of the link adaptation technique. Four data users are shown in a cell. User 1 does not have packets to transmit, while user 2 has packets in his queue but is in Mode 0. Among users 3 and 4 that both want to transmit, user 4 is allowed by the BS to transmit.

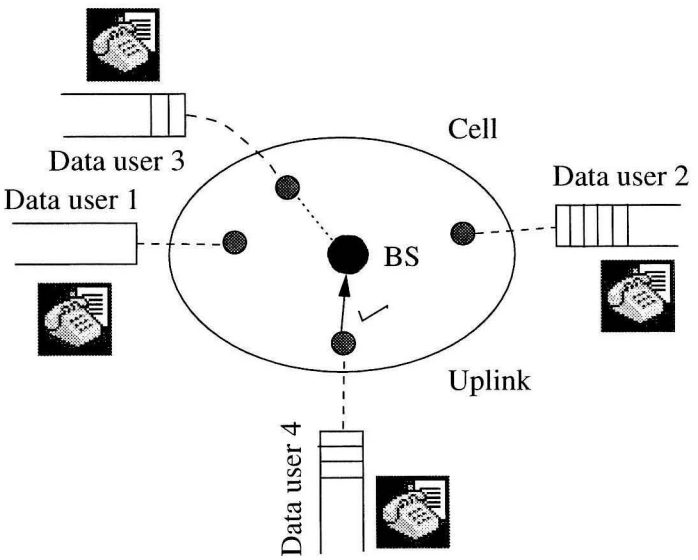

Figure 3.2: Illustration of the link adaptation technique

3.5.1 Performance Metric for Comparison

The average packet waiting time is the performance metric we use for comparing the performance of link adaptation techniques with different system parameters. The lower the average packet waiting time, the better the system performance.

The analysis presented below is for the two transmission modes case, where a user is either in Mode 0 or wants to transmit one packet in a burst using Mode 1. A burst is equivalent to a packet in this case, and the two terms are used interchangeably in this section. Therefore, as per the link adaptation rule specified in Section 3.4.2, a user is in Mode 0 if $\widehat{\mathrm{SIR}}_t < l_0$. He wants to transmit in Mode 1 otherwise. As we discuss later in Appendix 3.A.1, the analysis of the stable queues scenario for the three or more transmission modes case is an open problem.

3.5.2 Analysis

We first derive a closed form expression for the average packet waiting time given the average burst success probability, given that a user is allowed to transmit, $s^{(1)}$, the probability p_0 a user is in Mode 0, the packet arrival statistics λ and V, and the number of users per cell n.

Average Packet Waiting Time

The average packet waiting time for a user with packet arrival statistics defined by average rate λ and second moment V can be shown to be[13]

$$W = \frac{V - 2\lambda^2 + \lambda}{2\lambda(s^{(1)}c(1-p_0) - \lambda)}, \tag{3.13}$$

where c, the multiple access contention factor, is the probability that a user who wants to transmit is allowed to transmit. If C (a random variable) is the number of other users that also want to transmit, we have $c = E\left[\frac{1}{C+1}\right]$. This is because the BS chooses randomly, *i.e.*, with uniform probability, the user that gets to transmit from among the users that want to transmit.

The contention factor for a stable system, derived in Appendix 3.A.1, is approximately given by

$$c = \frac{\lambda}{s^{(1)}} + \frac{V - 2\lambda^2 + \lambda}{V - (n+1)\lambda^2 + \lambda}\left(1 - \frac{n\lambda}{s^{(1)}}\right). \tag{3.14}$$

[13]The proof is similar to that for Lemma 1 in Appendix 3.A.1.

Note that c depends on λ, V, $s^{(1)}$, and the number of users n per cell. Also notice that this is an approximation since it does not explicitly take into account the effect of p_0 on c. As p_0 increases, c increases, albeit insignificantly.

Interference Statistics

For a given traffic arrival process, the link adaptation thresholds and the channel fading statistics determine the interference statistics. Together, these determine the burst success probability $s^{(1)}$ (given that a user is allowed to transmit) and the probability p_0 of a user being in Mode 0. We first derive the first and second order statistics for the number of interferers k_t at time t, if s and p_0 are known. We subsequently use these to arrive at expressions for $s^{(1)}$ and p_0.

The fading processes and traffic arrivals for users in different cells are independent of each other, and the set of interferers for each cell is different. Therefore, to a good approximation, the cells at a given instant of time behave independently of each other. The probability $\pi(k_t)$ that a given cell encounters interference from k_t cells (out of the 6 first tier interfering cells) is then given by

$$\pi(k_t) = \begin{pmatrix} 6 \\ k_t \end{pmatrix} (1-q)^{k_t} q^{6-k_t}, \tag{3.15}$$

where q denotes the probability that no transmission occurs in a cell. This is equivalent to the probability that no user in the cell wants to transmit.

Equating the mean rate of arrival and departure of packets in a cell yields

$$q = 1 - \frac{n\lambda}{s^{(1)}}. \tag{3.16}$$

The conditional probability $\pi(k_t|k_{t-\tau})$ that k_t cells interfere at time t, given that $k_{t-\tau}$ cells interfered at time $t-\tau$, is given by

$$\pi(k_t|k_{t-\tau}) = \begin{pmatrix} 6 \\ k_t \end{pmatrix} q_c(k_{t-\tau})^{6-k_t} (1 - q_c(k_{t-\tau}))^{k_t}, \tag{3.17}$$

where

$$q_c(k_{t-\tau}) = p_{\text{fn}}^{(\tau)} + \frac{k_{t-\tau}}{6}\left(1 - p_{\text{fn}}^{(\tau)} - p_{\text{nf}}^{(\tau)}\right).$$

$p_{\text{fn}}^{(\tau)}$ is the probability that a transmission occurs in a cell at time t given that no transmission occurred in it at time $t-\tau$. Similarly, $p_{\text{nf}}^{(\tau)}$ is the probability that no transmission occurs in a cell at time t given that a transmission occurred in it at time $t-\tau$.

The above expressions are derived in Appendix 3.A.2. Evaluation of $p_{\text{fn}}^{(\tau)}$ is given in Appendix 3.A.3. For example, for $\tau = 1$ and a Bernoulli packet arrival process, in which one packet arrives with probability λ and no packets arrive with probability $1-\lambda$, $p_{\text{fn}}^{(1)} \approx 1-(1-\lambda)^n$.

Evaluating $s^{(1)}$

We will see in Section 3.5.3 that, due to channel correlation, the mean fade power Ω_m seen by a user when he has packets to transmit is lower than the channel's actual mean fade power Ω.[14] The following derivations therefore use Ω_m as the mean fade power, instead of Ω, for the joint Nakagami pdf in (3.4).

From (3.6), $\text{SIR}_t = \frac{\alpha_t^2}{k_t}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}$ and $\widehat{\text{SIR}}_t = \text{SIR}_{t-\tau} = \frac{\alpha_{t-\tau}^2}{k_{t-\tau}}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}$. Therefore, the probability $s^{(1)}(k_{t-\tau}, k_t)$ that a packet is successfully transmitted (given that $\widehat{\text{SIR}}_t \geq l_0$), as a function of k_t and $k_{t-\tau}$, is

$$s^{(1)}(k_{t-\tau}, k_t) = \frac{\int_{\sqrt{l_0 k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0}}}^{\infty} d\alpha_{t-\tau} \int_0^{\infty} \left(1 - c_1^{(1)} \exp\left(-c_2^{(1)} g^{(1)} \frac{\alpha_t^2}{k_t}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}\right)\right)^{L^{(1)}} p(\alpha_{t-\tau}, \alpha_t) d\alpha_t}{\mathcal{P}\left(\alpha_{t-\tau} \geq \sqrt{l_0 k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0}}\right)}, \tag{3.18}$$

where $\mathcal{P}(E)$ denotes the probability of occurrence of event E. Simplification results in [23,

[14] The channel correlation not only leads to $\Omega_m < \Omega$, but also changes the statistics of the channel seen by the transmitter. Using the Nakagami joint distribution, given in (3.4), for calculating $s^{(1)}$ is therefore an approximation by itself.

53]

$$s^{(1)}(k_{t-\tau}, k_t) = \sum_{h=0}^{L^{(1)}} (-1)^h \binom{L^{(1)}}{h} \frac{{c_1^{(1)}}^h m^m}{\left(m + \frac{c_2^{(1)} h g^{(1)} \Omega_m}{k_t} \left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}\right)^m} \left[\frac{\Gamma(m) - \Gamma\left(m, l_0 k_{t-\tau} \left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0} b^{(1)}(k_t, h)\right)}{\Gamma(m) - \Gamma\left(m, l_0 k_{t-\tau} \left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0} \frac{m}{\Omega_m}\right)}\right], \tag{3.19}$$

where

$$\begin{aligned} b^{(1)}(k_t, h) &= \frac{m}{\Omega_m} \frac{\left(m + \frac{c_2^{(1)} h g^{(1)} \Omega_m}{k_t} \left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}\right)}{\left(m + \frac{c_2^{(1)} h g^{(1)} \Omega_m}{k_t} \left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0} (1-\rho)\right)}, \\ \Gamma(m, a) &= \int_0^a x^{m-1} e^{-x} dx. \end{aligned}$$

Taking expectations over the random variables $k_{t-\tau}$ and k_t, the successful burst transmission probability $s^{(1)}$ is computed from

$$s^{(1)} = \sum_{k_{t-\tau}=0}^{6} \sum_{k_t=0}^{6} s^{(1)}(k_{t-\tau}, k_t) \pi(k_{t-\tau}) \pi(k_t | k_{t-\tau}). \tag{3.20}$$

The probabilities $\pi(k_{t-\tau})$ and $\pi(k_t|k_{t-\tau})$ have been derived earlier in Section 3.5.2. Notice how these in turn depend on $s^{(1)}$ through (3.15)–(3.17).

Evaluating p_0

As described in Section 3.4, a user with packets to transmit decides not to transmit, *i.e.*, he is in Mode 0, if $\widehat{\text{SIR}}_t = \text{SIR}_{t-\tau} < l_0$. Hence, the probability that a user is in Mode 0 is given by

$$\begin{aligned} p_0 &= \mathcal{P}\left(\frac{\alpha_{t-\tau}^2}{k_{t-\tau}} \left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0} < l_0\right), \\ &= \sum_{k_{t-\tau}=0}^{6} \mathcal{P}\left(\alpha_{t-\tau}^2 < l_0 k_{t-\tau} \left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0} | k_{t-\tau}\right) \pi(k_{t-\tau}), \end{aligned} \tag{3.21}$$

$$= \frac{1}{\Gamma(m)} \sum_{k_{t-\tau}=0}^{6} \Gamma\left(m, l_0 k_{t-\tau} \left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0} \frac{m}{\Omega_m}\right) \pi(k_{t-\tau}). \tag{3.22}$$

Eqn. (3.22) follows from (3.21) using the Nakagami distribution in (3.5).

3.5.3 Skewed Sampling of Fading Channel

A user observes the channel only if he has packets to transmit. The probability that a packet gets successfully transmitted, and that a user is not in Mode 0 in the first place, is lower for a deep fade. Also, a packet remains in the user's queue until it is successfully transmitted. As a result, the deeper the fade (and consequently the lower the SIR, for the same level of interference), the longer the duration of time a packet stays in the user's queue. Therefore, the average fade power Ω_m observed by a user is less than $\Omega = 1$.

An exact evaluation of Ω_m involves calculating the average of the fade power observed by the user over all the various possible fade and interference evolutions, making the problem intractable. We describe below a simplified block fading based approximation, which assumes that the system parameters remain constant over a duration of time d_c and then decorrelate thereafter.

Block Fading Approximation

Let α denote the channel fade. Let $b(\alpha)$ denote the probability that a user is in Mode 0, if the channel fade is α. Let $s(\alpha, k)$ denote the successful burst transmission probability when the fade is α and the number of interferers is k.

A user observes a fade α exactly once if he is not in Mode 0 and if his burst is successfully transmitted. This happens with probability $b(\alpha)s(\alpha, k)$. Similarly, the user samples the fade α twice with probability $(1 - s(\alpha, k)b(\alpha))b(\alpha)s(\alpha, k)$, and so on up to d_c time instants. Notice that the fade α is taken to be the same for all these time instants.

After d_c time instants, the fade is assumed to become uncorrelated to the initial fade α. Thereafter, the user observes average values of the fade and the burst success probability. The decorrelated fade power measured after d_c time instants is then $\overline{\alpha^2} = \Omega = 1$, and the burst success probability is $\overline{s(\alpha, k)}$. The average number of times the user samples

the average fade power Ω is then $\frac{1}{s(\alpha,k)}$. This block fading model is shown in Fig. 3.3. In general, the longer the decorrelation length d_c, the lower the average fade power seen by a user that has packets to transmit. The channel correlation coefficient ρ is one factor that clearly influences how soon the link decorrelates.

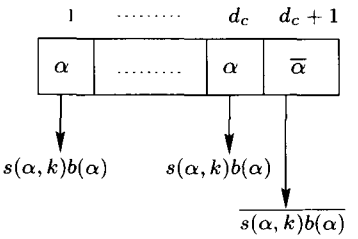

Figure 3.3: Block fading and interference model

It is important to note the extent of simplifications this model makes for evaluating Ω_m:

- The interference dynamics are not accounted for since the number of interferers k is taken to be the same up to d_c time instants.
- The fade is assumed to remain constant up to d_c time instants, and then to decorrelate completely immediately thereafter.
- $b(\alpha)$ does not explicitly take into account the effect of the multiple access contention factor c. For low c, a user does not transmit not only because he is in Mode 0, but also because he is not allowed to transmit by the BS.

A comparison of analytical results with simulation results, in Section 3.7.1, shows that these approximations do not significantly affect the accuracy of the analytical results.

Under the above model, the average power of a channel fade seen by a user that has packets to transmit is, by definition, given by

$$\Omega_m = \frac{E_{\alpha,k}\left[\begin{array}{l}\alpha^2(s(\alpha,k)b(\alpha)+\ldots+d_c(1-s(\alpha,k)b(\alpha))^{d_c-1}s(\alpha,k)b(\alpha)) \\ +\left(d_c\alpha^2+\frac{\overline{\alpha^2}}{s(\alpha,k)}\right)(1-s(\alpha,k)b(\alpha))^{d_c}\end{array}\right]}{E_{\alpha,k}\left[\begin{array}{l}s(\alpha,k)b(\alpha)+\ldots+d_c(1-s(\alpha,k)b(\alpha))^{d_c-1}s(\alpha,k)b(\alpha) \\ +(d_c+\frac{\overline{\alpha^2}}{s(\alpha,k)})(1-s(\alpha,k)b(\alpha))^{d_c}\end{array}\right]}, \quad (3.23)$$

where $E_{\alpha,k}[.]$ denotes the expectation over the random variables α and k. Their probability distributions are as given in (3.5) and (3.15), respectively. Section 3.A.4 in the appendix evaluates a closed form solution for Ω_m from the above equation.

Since the fade is assumed to be constant in a block of duration d_c, the probability that a user is in Mode 1 depends on the number of interferers, and is given by $b(\alpha) = \sum_{k=0}^{\eta(\alpha)-1} \pi(k)$, where $\eta(\alpha)$ is the minimum number of interferers required for a user to not transmit[15] if his fade is α. Therefore, it is defined by the relations $\frac{\alpha^2}{\eta(\alpha)-1}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0} \leq l_0 < \frac{\alpha^2}{\eta(\alpha)}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}$. $s(\alpha, k)$, for SIR $= \frac{\alpha^2}{k}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}$ is as given in (3.9).[16]

In brief, $s^{(1)}$, p_0, and c are evaluated by solving a system of coupled non-linear equations (3.14)–(3.17), (3.19), (3.20), (3.22), and (3.23). The average packet waiting time W is then calculated from (3.13). The interference-limited nature of the cellular system, which leads to inter-dependencies between its constituent cells, is brought out by the occurrence of the coupled system of equations. The results for the stable queue scenario are presented in Section 3.7, along with the results for the saturated queues scenario that we analyze in the following section.

3.6 Saturated Queues Scenario

We now consider the scenario where users always have packets to transmit. The packet arrival process then plays no role in this model, simplifying the problem of analyzing the system. Therefore, the multiple transmission modes case can be analyzed for this simpler scenario.

At *every instant* of time, the users decide which mode to be in based on their SIR estimates. In this scenario, to transmit a burst, a user must

1. Not be in Mode 0.

2. Be chosen by the BS to transmit from among the other users who also want to transmit. As in Section 3.5, we assume that the BS chooses randomly one user from

[15]Since $\eta(\alpha) \leq 6$, for high enough α the user will transmit irrespective of the number of interferers. For this case no feasible $\eta(\alpha)$ exists and $b(\alpha) = \sum_{k=0}^{6} \pi(k) = 1$.

[16]Note that (3.9) gives a pessimistic (lower) value for $s(\alpha, k)$ since it is not conditioned on the fact that the user is not in Mode 0.

among the users that want to transmit.

Terminology: We shall say that a user **wants to** transmit if he meets the first criterion. We shall say that a user is **allowed to** transmit if he satisfies both the above criteria. Note the difference in terminology from the stable queues scenario considered earlier in Section 3.5, where users had to satisfy an additional requirement to be able to transmit.

3.6.1 Performance Metric for Comparison

Since users always have packets in their queues, *i.e.*, they have an infinite backlog of packets to transmit, average packet waiting time – the performance metric in the stable queues scenario – cannot be meaningfully defined for the saturated queues scenario. The metric for evaluating system performance is now the average system throughput, defined as the average number of packets received error-free by the BS per unit time. The higher the throughput, the better the system performance.

3.6.2 Analysis

Throughput

The system throughput μ, derived in Appendix 3.B.1, can be easily shown to be

$$\mu = (1-q)\left(\sum_{i=1}^{H} i s^{(i)}\right), \tag{3.24}$$

where q is the probability that all the users in a cell are in Mode 0. $s^{(i)}$ is the probability that a burst is transmitted in Mode i ($i \geq 1$) and is successfully received by the BS.

We now derive expressions for q and $s^{(i)}$ in terms of the other system variables.

Expression for $s^{(i)}$

The derivation for $s^{(i)}$, $i \geq 1$, is similar to that in Section 3.5.2 for the stable queues scenario. Since a user always has packets to transmit, and therefore observes the channel at

every time instant, the average fade power measured by him is $\Omega = 1$ itself. This is unlike what we saw in Section 3.5.3 for the stable queues scenario.

The probability of successful burst transmission $s^{(i)}(k_{t-\tau}, k_t)$, when the number of interferers at transmission time t is k_t and at measurement time $t-\tau$ is $k_{t-\tau}$, is given by

$$
\begin{aligned}
s^{(i)}(k_{t-\tau}, k_t) = & \\
& \frac{\int_{\sqrt{l_{i-1}k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0}}}^{\sqrt{l_i k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0}}} d\alpha_{t-\tau} \int_0^\infty \left(1 - c_1^{(i)} \exp\left(-c_2^{(i)} g^{(i)} \frac{\alpha_t^2}{k_t} \left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}\right)\right)^{L^{(i)}} p(\alpha_{t-\tau}, \alpha_t) d\alpha_t}{\mathcal{P}\left(\alpha_{t-\tau} \geq \sqrt{l_0 k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0}}\right)}, \\
= & \sum_{h=0}^{L^{(i)}} (-1)^h \binom{L^{(i)}}{h} \frac{{c_1^{(i)}}^h m^m}{\left(m + \frac{c_2^{(i)} h g^{(i)}}{k_t}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}\right)^m} \\
& \times \left[\frac{\Gamma\left(m, l_i k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0} b^{(i)}(k_t, h)\right) - \Gamma\left(m, l_{i-1} k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0} b^{(i)}(k_t, h)\right)}{\Gamma(m) - \Gamma\left(m, l_0 k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0} m\right)}\right], \quad (3.25)
\end{aligned}
$$

where

$$
\begin{aligned}
b^{(i)}(k_t, h) &= \frac{m\left(m + \frac{c_2^{(i)} h g^{(i)}}{k_t}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}\right)}{\left(m + \frac{c_2^{(i)} h g^{(i)}}{k_t}\left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}(1-\rho)\right)}, \\
\Gamma(m, a) &= \int_0^a x^{m-1} e^{-x} dx.
\end{aligned}
$$

$s^{(i)}$, the probability of successful burst transmission in Mode i, is therefore

$$
s^{(i)} = \sum_{k_t=0}^{6} \sum_{k_{t-\tau}=0}^{6} s^{(i)}(k_{t-\tau}, k_t) \pi(k_t | k_{t-\tau}) \pi(k_{t-\tau}). \quad (3.26)
$$

The interference statistics $\pi(k_{t-\tau})$ and $\pi(k_t|k_{t-\tau})$ are derived below.

Expression for q

q is the probability that no transmission occurs in a cell, *i.e.*, all the users in the cell are in Mode 0. As shown in Appendix 3.B.2, q is given by

$$q = \sum_{k_{t-\tau}=0}^{6} \left[\frac{\Gamma(m, l_0 m k_{t-\tau} \left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0})}{\Gamma(m)} \right]^n \pi(k_{t-\tau}). \tag{3.27}$$

Recall that m is the Nakagami fading parameter and l_0 is the Mode 0 SIR threshold.

Interference Statistics

The channel fading is independent from one cell to another. Moreover, the set of interferers for each cell is different. Therefore, the binomial approximation for evaluating $\pi(k_t)$, which we had used earlier in Section 3.5.2 for the stable queues scenario, is valid for the saturated queues scenario also. $\pi(k_t)$ is therefore given by

$$\pi(k_t) = \begin{pmatrix} 6 \\ k_t \end{pmatrix} (1-q)^{k_t} q^{6-k_t}, \tag{3.28}$$

where q, given in (3.27), is the probability that no transmission occurs in a cell.

The derivation of the second order interference statistics is the same as that in the stable queues scenario. The second order interference statistics for the stable queues scenario were given in Section 3.5.2 and derived in Appendix 3.A.2. In brief, the conditional probability $\pi(k_t|k_{t-\tau})$ that k_t cells interfere at time t given that $k_{t-\tau}$ cells interfered at time $t-\tau$ takes the form

$$\pi(k_t|k_{t-\tau}) = \begin{pmatrix} 6 \\ k_t \end{pmatrix} q_c(k_{t-\tau})^{6-k_t} (1 - q_c(k_{t-\tau}))^{k_t}, \tag{3.29}$$

where

$$q_c(k_{t-\tau}) = p_{\text{fn}}^{(\tau)} + \frac{k_{t-\tau}}{6}(1 - p_{\text{fn}}^{(\tau)} - p_{\text{nf}}^{(\tau)}).$$

Recall from Section 3.5.2 that $p_{\text{fn}}^{(\tau)}$ is the probability that a transmission occurs in a cell at time t given that no transmission occurred in it at time $t-\tau$, and $p_{\text{nf}}^{(\tau)}$ is the probability that no transmission occurs in a cell at time t given that a transmission occurred in it at time $t-\tau$.

The main difference between the earlier stable queues scenario and the saturated queues scenario, which we are now analyzing, is the value of $p_{\text{fn}}^{(1)}$, from which $p_{\text{fn}}^{(\tau)}$ is evaluated. As explained in detail in Appendix 3.B.3, in the saturated queues scenario $p_{\text{fn}}^{(1)} \approx 1$.

In brief, the average system throughput, given in (3.24), is a function of q and $s^{(i)}$ ($1 \leq i \leq H$). q and $s^{(i)}$ are obtained by solving a system of coupled non-linear equations (3.25)–(3.29). The results for the saturated queues scenario are presented in Section 3.7.2.

3.7 Results

We first present results for the stable queues scenario in Section 3.7.1. In Section 3.7.2, we present results for the saturated queues scenario.

3.7.1 Stable Queues Scenario

The common parameters used in the results are: users per cell $n = 5$, path loss exponent $\gamma_0 = 3$, modulation coefficients $c_1^{(1)} = 0.20$ and $c_2^{(1)} = 0.50$, coding gain $g^{(1)} = 2$, burst/packet length $L^{(1)} = 16$, Nakagami fading parameter $m = 1$ (Rayleigh fading), and feedback delay $\tau = 1$. The users are symmetrically located at a distance $r = 0.75$ from their serving BSs. The reuse cluster size is 1. The results are shown for a Bernoulli packet arrival process with mean $\lambda = 0.06$.

Figure 3.4 plots the average packet waiting time, evaluated analytically and by means of simulations, as a function of the Mode 0 threshold l_0 for a high channel correlation coefficient $\rho = 0.82$. Recall that in the two modes case, a user wants to transmit a packet at time t if $\widehat{\text{SIR}}_t \geq l_0$. The simulation set-up is described briefly in Appendix 3.A.5. The analytical results are shown for decorrelation lengths $d_c = 2$ and $d_c = 3$. For low l_0, the larger decorrelation block length ($d_c = 3$) provides a better match with the simulation results, while a smaller decorrelation block length ($d_c = 2$) is more accurate for larger values of l_0. We can see that Mode 0 reduces the average packet waiting time by up to 15% as compared to the case where no Mode 0 is used ($l_0 = 0$).

Figure 3.5 plots the average packet waiting time for a low channel correlation coefficient

$\rho = 0.41$, all other parameters remaining the same. The analytical results are shown for $d_c = 1$ and $d_c = 2$. For smaller l_0, $d_c = 2$ is more accurate than $d_c = 1$, while the reverse is true for larger l_0. The average packet waiting time is the lowest when $l_0 = 0$, *i.e.*, when Mode 0 is not used.

Comparing the results in Fig. 3.4 and Fig. 3.5, we therefore see that Mode 0 does not yield any gains in system performance if the correlation between the estimated SIR and the actual SIR is low enough, *i.e.*, the link estimates are not very reliable. In other words, the optimal link adaptation threshold depends on the reliability of the estimates. This is not taken into account in the previous papers on link adaptation thresholds [18, 54, 49].

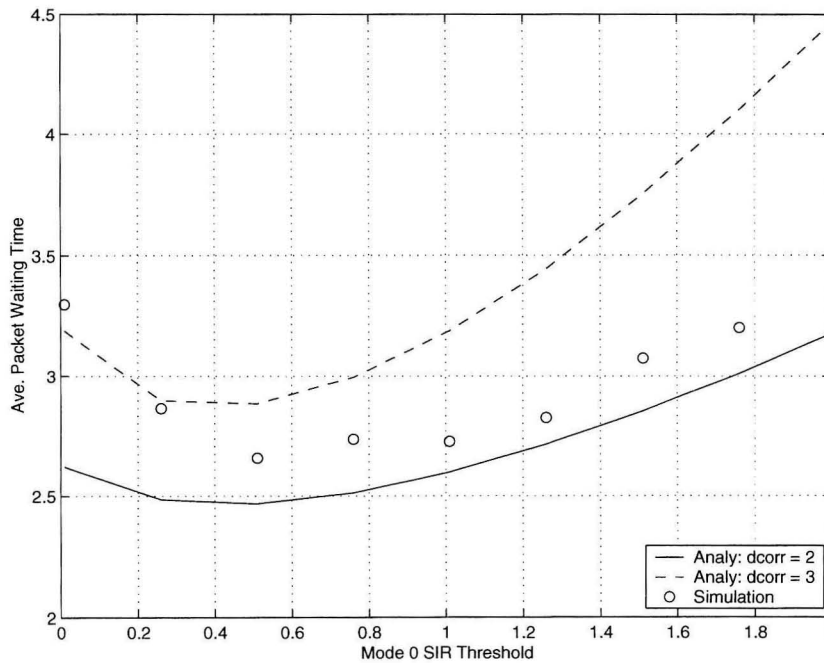

Figure 3.4: Ave. packet waiting time *vs.* l_0: high channel correlation ($\rho = 0.82$)

3.7.2 Saturated Queues Scenario

The results are first shown for the two transmission modes case. Recall that, in this case, a user can either be in Mode 0 or Mode 1 (a single fixed modulation and coding scheme).

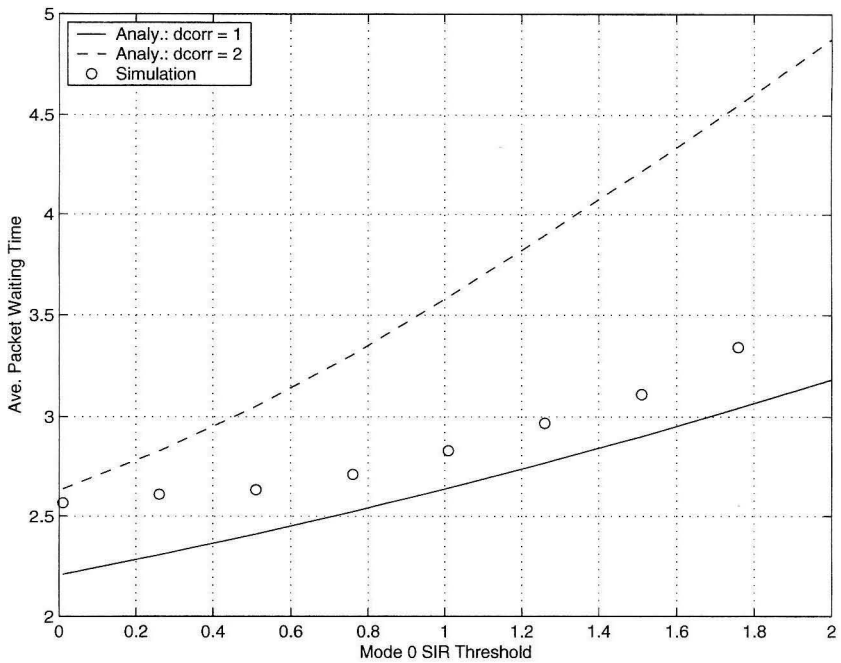

Figure 3.5: Ave. packet waiting time *vs.* l_0: low channel correlation ($\rho = 0.41$)

The results for the three modes case are then presented. Here, a user not in Mode 0 can choose to transmit using one of two given modes – Mode 1 and Mode 2.

Two Modes Case

Figure 3.6 plots the system throughput as a function of the Mode 0 SIR threshold for two different channel correlations: $\rho = 0.82$ and $\rho = 0.41$. The values of the system parameters are as follows: users per cell $n = 5$, Nakagami fading parameter $m = 1$, symbols per burst $L^{(1)} = 16$, feedback delay $\tau = 1$, path loss exponent $\gamma_0 = 3.0$, reuse cluster size 1, coding gain $g^{(1)} = 2$, and modulation coefficients $c_1^{(1)} = 0.20$ and $c_2^{(1)} = 0.50$. Also shown in the figure are the results obtained from simulations. While the analytical results track the results from simulations well, the mismatch between them increases for higher Mode 0 SIR thresholds. This is primarily due to the approximation for $p_{\text{fn}}^{(1)}$ described in Appendix 3.B.3.

Note that the SIR threshold ($l_0 = 2.0$) at which throughput is maximized does not change with channel correlation. This is unlike the behavior observed in Section 3.7.1 where all the users had stable queues. As expected, the throughput of the system does degrade as the channel correlation decreases.

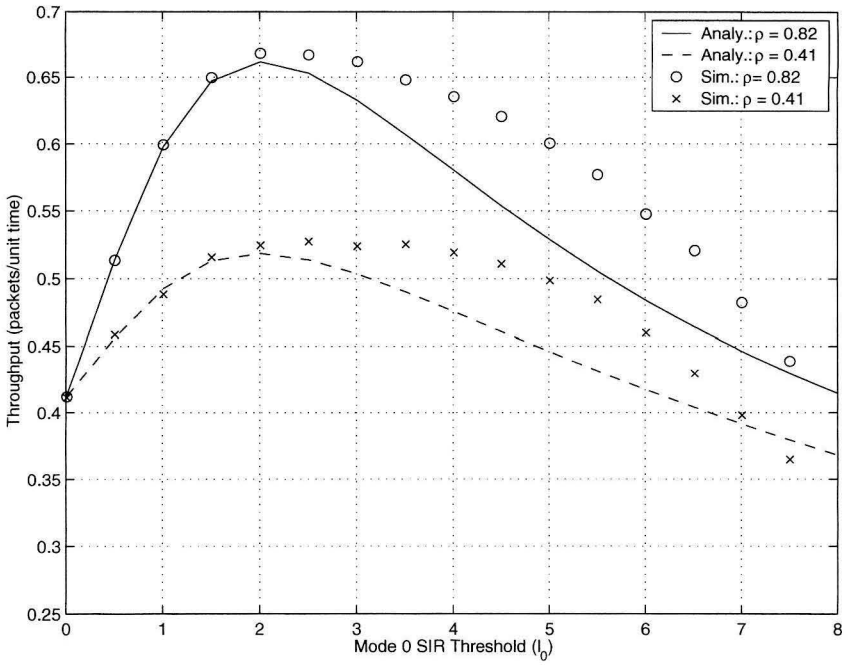

Figure 3.6: Two modes case: throughput *vs.* l_0 for $\rho = 0.82$ and $\rho = 0.41$

We now examine the effect of the Nakagami parameter m on the optimal link adaptation threshold. $m = 1$ corresponds to the Rayleigh fading channel; the channel becomes more and more like an additive white Gaussian noise channel (without fading) as m increases. Figure 3.7 plots throughput *vs.* l_0, obtained using analysis, for $m = 0.5, 1.0, 2.0$, and 10.0 when $\rho = 0.82$. Figure 3.8 does the same for $\rho = 0.41$. The points of maximum throughput have been circled in the two figures.

It is interesting to observe that when $\rho = 0.82$ the system throughput actually decreases for larger m. This is because the benefits of link adaptation are lost when all the users face identical non-fading channels. Note that the system behavior is very different for $\rho = 0.41$. In this case, the maximum throughput increases as m increases. However, the senstivity of the system throughput to the Mode 0 threshold l_0 also increases. We again observe that the maximum throughput is always achieved at the same l_0 irrespective of the value of m.

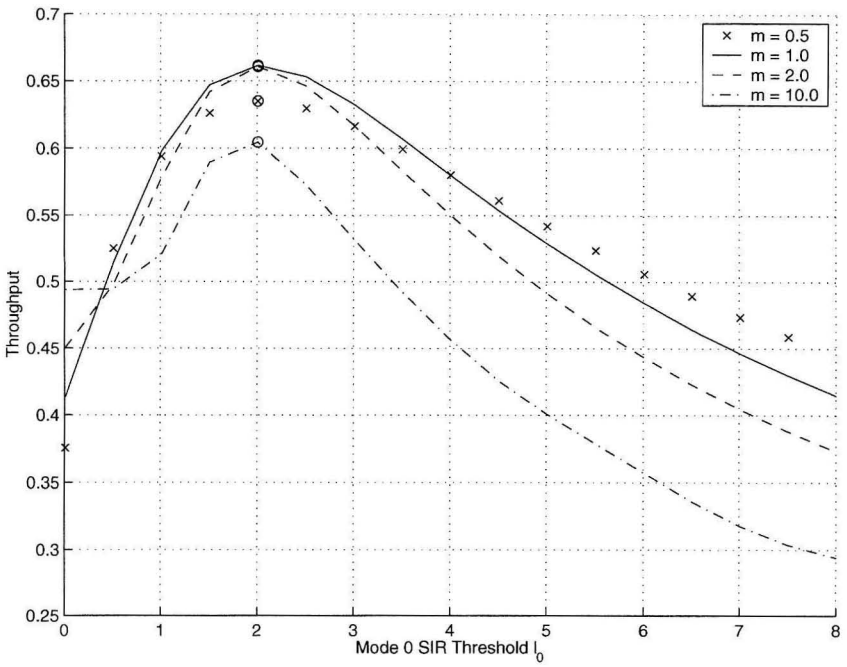

Figure 3.7: Effect of m (two modes case): throughput *vs.* l_0 for high channel correlation ($\rho = 0.82$)

Three Modes Case

In the three modes case, two thresholds need to be chosen: l_0 and l_1. Figure 3.9 plots the throughput *vs.* l_1 for different Mode 0 thresholds l_0, for channel correlation $\rho = 0.82$. The results obtained using analysis and simulations are both plotted for comparison. The values of the system parameters are as follows: users per cell $n = 5$, $m = 1$, $\gamma_0 = 3$, reuse cluster size 1, $L^{(1)} = L^{(2)} = 16$, $c_1^{(1)} = 0.20$, $c_2^{(1)} = 0.50$, $c_1^{(2)} = 0.20$, $c_2^{(2)} = 0.21$, $g^{(1)} = 2$, and $g^{(2)} = 2$. Figure 3.10 does the same for $\rho = 0.41$.

It can be seen that for low Mode 0 thresholds l_0, there exists a region of SIR in which using Mode 1 improves system performance. For example, when $\rho = 0.82$ and $l_0 = 1.0$, Mode 1 should be used when $1.0 = l_0 \leq \widehat{\text{SIR}} < l_1 = 3.0$. However, the optimal thresholds at which system throughput is maximized (at $l_0 = l_1 = 4.0$ when $\rho = 0.82$) correspond to the case where Mode 1 is not used – the user is either in Mode 0 or Mode 2. This type of behavior where one of the modes is turned off was also observed for other typical values of $c_1^{(\cdot)}$, $c_2^{(\cdot)}$, and $g^{(\cdot)}$. As in the two modes case, we again observe that the optimal link adaptation thresholds are insensitive to the channel correlation. The difference between

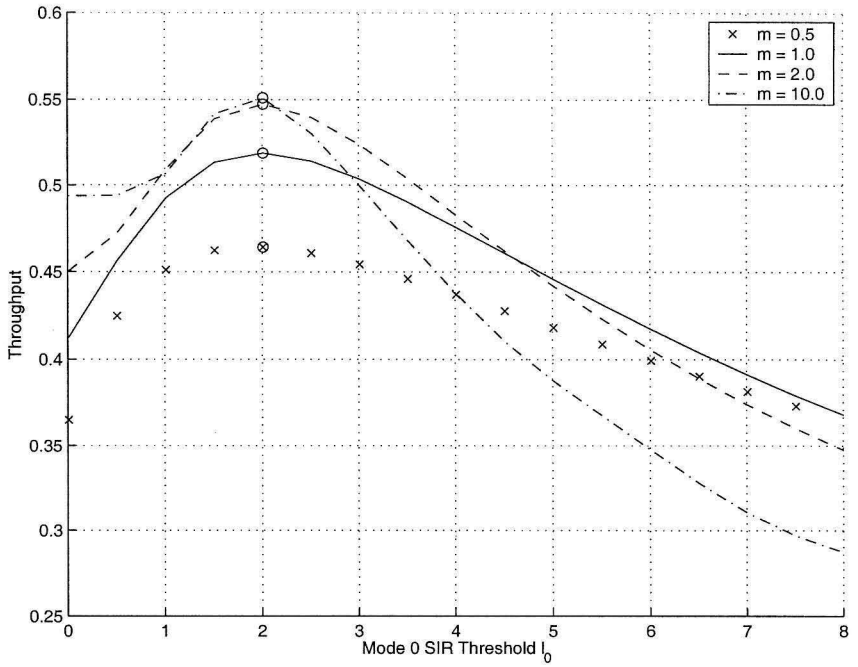

Figure 3.8: Effect of m (two modes case): throughput *vs.* l_0 for low channel correlation ($\rho = 0.41$)

analytical and simulation results for higher Mode 0 thresholds, similar to that observed in Section 3.7.2, should be noted.

3.8 Conclusions

We have studied the optimization and performance of cellular systems that adapt to the link condition using SIR measurements. We derived analytical expressions for the average packet waiting time for data users in an interference-limited cellular system under a stable queues assumption. The scheme we studied used two modes – the no-transmission mode and the transmission mode. We saw that the optimal no-transmission mode threshold depends on the channel correlation, which is a function of the delay in feeding back estimates and user speed. The no-transmission mode did reduce the average packet delay for a fading channel with high correlation. On the other hand, it bore no gains for a channel with low correlation. This result is not surprising: it indicates that adaptation is effective only for slowly varying channels where accurate link estimates can be obtained.

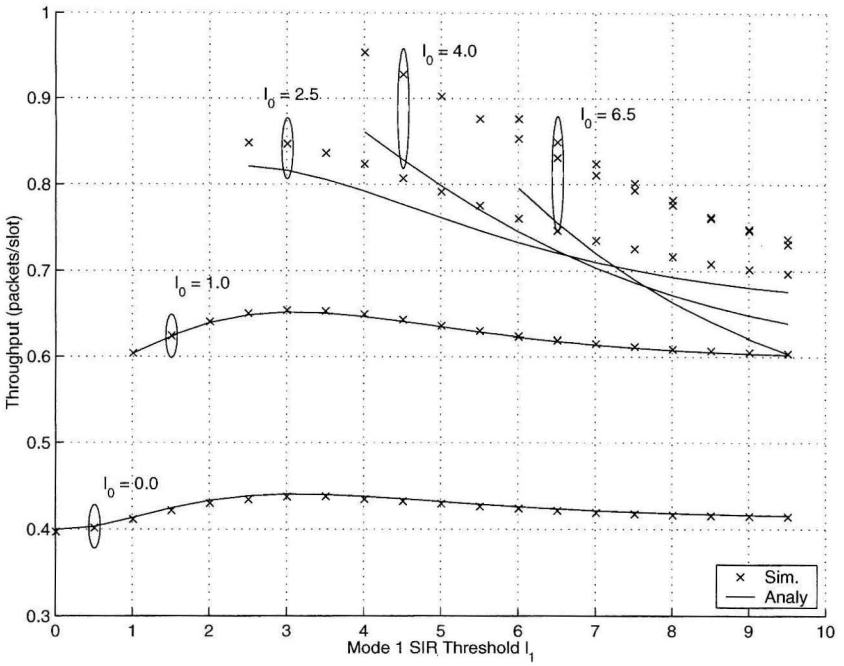

Figure 3.9: Three modes case: throughput *vs.* l_1 and l_0 for high channel correlation ($\rho = 0.82$)

We also studied the system throughput of a link adaptation technique in a saturated queues scenario, where users always have packets to transmit. The multiple transmission modes scenario was analyzed. Unlike the stable queues scenario, the optimal no-transmission mode threshold was found to be insensitive to channel correlation. This explains, to an extent, the intuition behind the heuristic formulae proposed in the literature. However, the system throughput did decrease as the channel correlation decreased.

In summary, one can say that the no-transmission mode is effective in a correlated fading environment where good estimates of the link condition can be obtained. In the saturated queues scenario, the optimal thresholds are insensitive to the channel correlation, and optimum performance is obtained when the intermediate transmission modes generally are discarded.

It is important to note that the analytical results presented in this chapter are for a system model in which users face similar path loss and statistical fading models. Link adaptation will yield its maximum gains in an unequal user situation where, on the average, users closer to the base station use high rate, low redundancy transmission modes, while the users

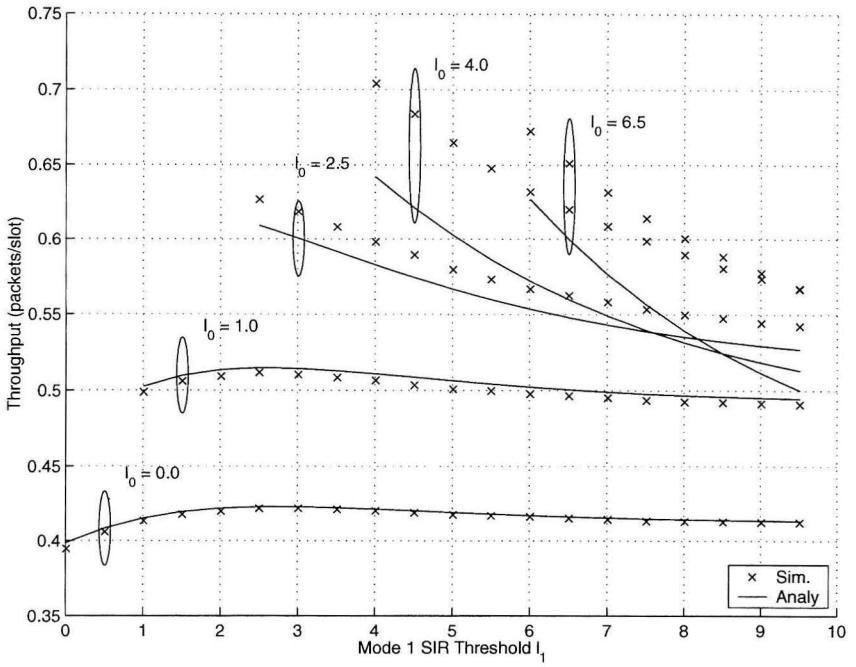

Figure 3.10: Three modes case: throughput *vs.* l_1 and l_0 for low channel correlation ($\rho = 0.41$)

near the cell boundaries use low rate, high redundancy transmission modes or do not transmit at all. Analysis of the link adaptation technique for the unequal user situation remains an unsolved problem. Moreover, the issue of fairness needs to be resolved before the no-transmission mode is employed in such unequal user scenarios. While the no-transmission mode may improve the overall system performance, using it may also result in users close to the cell boundary not transmitting at all. Analyses of the effect of incremental redundancy and of power control combined with link adaptation are other interesting extensions to the work presented in this chapter.

Appendix

3.A Stable Queues Scenario

3.A.1 Contention Factor c

We first consider the no Mode 0 case ($p_0 = 0$). In this case, a user wants to transmit a packet if and only if his queue is non-empty.

Consider a packet which arrives in the queue of user i. Define Q_t to be the total number of packets in all the users' queues, when this packet arrives. Define $A_t^{(i)}$ to be the number of packets which arrive in user i's queue at time t. As per the notation defined in Section 3.4, $E\left[A_t^{(i)}\right] = \lambda$ and $E\left[(A_t^{(i)})^2\right] = V$, for all users i.

Lemma 1 *The average packet waiting time W for the no Mode 0 case is given by*

$$W = \frac{V - (n+1)\lambda^2 + \lambda}{2\lambda(s^{(1)} - n\lambda)}.$$

Proof: A packet will get transmitted by any one of the users in the cell so long as $Q_t > 0$, if Mode 0 is not used. The system evolution equation is given by

$$Q_{t+1} = Q_t + \sum_{i=1}^{n} A_t^{(i)} - 1\{Q_t, \text{Packet success}\}, \tag{3.30}$$

where

$$1\{Q_t, \text{Packet success}\} = \begin{cases} 1, & \text{if } Q_t > 0 \text{ and packet is successfully transmitted} \\ 0, & \text{otherwise} \end{cases}.$$

From (3.30), it follows that

$$E\left[Q_{t+1}\right] = E\left[Q_t\right] + \sum_{i=1}^{n} E\left[A_t^{(i)}\right] - E\left[1\{Q_t, \text{Packet success}\}\right].$$

Since $E[Q_{t+1}] = E[Q_t]$ in steady state, we get

$$E[1\{Q_t, \text{Packet success}\}] = \sum_{i=1}^{n} E\left[A_t^{(i)}\right] = n\lambda. \tag{3.31}$$

Similarly,

$$E\left[Q_{t+1}^2\right] = E\left[\left(Q_t + \sum_{i=1}^{n} A_t^{(i)} - 1\{Q_t, \text{Packet success}\}\right)^2\right]. \tag{3.32}$$

Using (3.31) and the fact that $E\left[Q_{t+1}^2\right] = E\left[Q_t^2\right]$ in steady state, we get upon simplifying (3.32) that

$$E[Q_t] = \frac{n\left(E\left[(A_t^{(i)})^2\right] + E\left[A_t^{(i)}\right] - (n+1)E\left[A_t^{(i)}\right]^2\right)}{2\left(s^{(1)} - nE\left[A_t^{(i)}\right]\right)}.$$

From Little's theorem [55]

$$W = \frac{E[Q_t]}{n\lambda} = \frac{V - (n+1)\lambda^2 + \lambda}{2\lambda(s^{(1)} - n\lambda)}.$$

■

Theorem 1 *The contention factor c, when $l_0 = 0$, is given by*

$$c = \frac{\lambda}{s^{(1)}} + \frac{V - 2\lambda^2 + \lambda}{V - (n+1)\lambda^2 + \lambda}\left(1 - \frac{n\lambda}{s^{(1)}}\right).$$

Proof: When $l_0 = 0$, a user with packets in his queue will want to transmit. The probability that he is allowed to transmit is c. Therefore, the service rate for packets in user i's queue is $s^{(1)}c$. Using a derivation similar to that in Lemma 1 for $n = 1$ yields

$$W = \frac{V - 2\lambda^2 + \lambda}{2\lambda(s^{(1)}c - \lambda)}. \tag{3.33}$$

Comparing the above equation with that in Lemma 1 yields the formula for c. ■

c could be evaluated, in the above derivations for the two modes case with $p_0 = 0$, because the queues for the various users could be aggregated. This is not possible when $p_0 \neq 0$ since the service rate of the aggregate queue will now also depend on how many of the user's queues are non-empty. For the multiple mode case, the queue aggregation technique again does not work for a similar reason. Consequently, an exact expression for c when $p_0 > 0$ is, as yet, unsolved. The multiple access factor for $p_0 > 0$ is approximated with the value for c derived above in Theorem 1. As the Mode 0 threshold increases, c in fact increases, albeit insignificantly. We find this approximation to be accurate to within 10% of the actual waiting time, for p_0 as high as 0.6.

3.A.2 Second Order Interference Statistics

Recall that the co-channel interference arises from the 6 first tier cells. In general, the number of interferers k_t at time t is not independent of the number of interferers $k_{t-\tau}$ at time $t - \tau$, especially when the channel fading process is correlated. We derive an approximate expression for the conditional probability $\pi(k_t|k_{t-\tau})$ based on the following simplified on/off Markov model for cell transmissions, shown in Fig. 3.11. The simulation results show this to be an accurate approximation, especially for lower p_0.

A transmission in a cell occurs when at least one of the users in it wants to transmit. At any instant of time, a cell can be in one of two states: transmitting a packet (on), or not transmitting (off), as shown in Fig. 3.11. Define $p_{\text{nf}}^{(\tau)}$ to be the probability that a cell is off at time t, given that it was on at time $t - \tau$. Likewise, $p_{\text{fn}}^{(\tau)}$ is the probability that a cell is on at time t given that it was off at time $t - \tau$.

Lemma 2

$$p_{\text{nf}}^{(1)} = \frac{q}{1-q} p_{\text{fn}}^{(1)}.$$

Proof: Recall from Section 3.5.2 that q is the probability that no transmission occurs in a cell. This is the same as the steady state probability of the cell being in the off state.

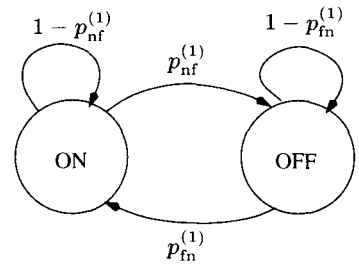

Figure 3.11: On/Off transmission model for a cell

From the Markov model in Fig. 3.11, the probability that the cell is in the off state is $p_{\text{nf}}^{(1)}/(p_{\text{nf}}^{(1)} + p_{\text{fn}}^{(1)})$. Hence, the result.

■

Define $E[k_t|k_{t-\tau}]$ to be the mean interferer count at time t given that $k_{t-\tau}$ interferers transmitted at time $t-\tau$.

Lemma 3

$$E[k_t|k_{t-\tau}] = 6p_{\text{fn}}^{(\tau)} + k_{t-\tau}(1 - p_{\text{nf}}^{(\tau)} - p_{\text{fn}}^{(\tau)}).$$

Proof: Let the cells that interfered at time $t-\tau$ be identified by the labels $1, 2, \ldots, k_{t-\tau}$, and let the other $(6 - k_{t-\tau})$ cells be identified by the labels $(k_{t-\tau}+1), \ldots, 6$. For cell i, define the indicator function

$$1_i(t) = \begin{cases} 1, \text{ if cell i is on at time } t \\ 0, \text{ otherwise} \end{cases}.$$

From the on/off Markov model for a cell defined above, we have

$$E[1_i(t)] = \begin{cases} 1 - p_{\text{nf}}^{(\tau)}, & 1 \le i \le k_{t-\tau} \\ p_{\text{fn}}^{(\tau)}, & k_{t-\tau} < i \le 6 \end{cases}. \tag{3.34}$$

The number of interferers at time t is given by

$$k_t = 1_1(t) + \cdots + 1_{k_{t-\tau}}(t) + 1_{k_{t-\tau}+1}(t) + \cdots + 1_6(t).$$

Taking expectations on both sides, we have

$$E\left[k_t | k_{t-\tau}\right] = k_{t-\tau}(1 - p_{\text{nf}}^{(\tau)}) + (6 - k_{t-\tau})(1 - p_{\text{nf}}^{(\tau)}).$$

Rearranging terms yields the required expression. ■

3.A.3 Evaluating $p_{\text{fn}}^{(\tau)}$ and $p_{\text{nf}}^{(\tau)}$

Expressions for $p_{\text{fn}}^{(1)}$ and $p_{\text{nf}}^{(1)}$

When $p_0 = 0$, a cell is in the off state if none of the users have packets to transmit. Therefore, the cell goes from the off state to the on state if a packet arrival occurs for at least one of the n users in the cell. Let p_e denote the probability that no packet arrives in a given user's queue. Since the packet arrival process at each of the users is independent, we have

$$p_{\text{fn}}^{(1)} = 1 - p_e^n. \tag{3.35}$$

For example, for Bernoulli traffic arrivals, $p_e = 1 - \lambda$. Given $p_{\text{fn}}^{(1)}$, $p_{\text{nf}}^{(1)}$ is evaluated from Lemma 2. When $p_0 > 0$, an exact expression for p_{fn} is unknown since it in turn requires an a priori knowledge of the conditional interference statistics. As an approximation, we use the value for $p_{\text{fn}}^{(1)}$ given in (3.35).

Expressions for $p_{\text{fn}}^{(\tau)}$ and $p_{\text{nf}}^{(\tau)}$

Given $p_{\text{fn}}^{(1)}$ and $p_{\text{nf}}^{(1)}$, the transition probability matrix $T^{(1)}$ for the Markov chain in Fig. 3.11 is

$$T^{(1)} = \begin{bmatrix} 1 - p_{\text{fn}}^{(1)} & p_{\text{fn}}^{(1)} \\ p_{\text{nf}}^{(1)} & 1 - p_{\text{nf}}^{(1)} \end{bmatrix}. \tag{3.36}$$

The transition probability matrix $T^{(\tau)}$ for a transition time τ is then given by

$$T^{(\tau)} = \begin{bmatrix} 1-p_{\text{fn}}^{(\tau)} & p_{\text{fn}}^{(\tau)} \\ p_{\text{nf}}^{(\tau)} & 1-p_{\text{nf}}^{(\tau)} \end{bmatrix} = \begin{bmatrix} 1-p_{\text{fn}}^{(1)} & p_{\text{fn}}^{(1)} \\ p_{\text{nf}}^{(1)} & 1-p_{\text{nf}}^{(1)} \end{bmatrix}^{\tau} = \left(T^{(1)}\right)^{\tau}. \tag{3.37}$$

Hence,

$$\begin{aligned} p_{\text{fn}}^{(\tau)} &= T^{(\tau)}(1,2), \\ p_{\text{nf}}^{(\tau)} &= T^{(\tau)}(2,1). \end{aligned} \tag{3.38}$$

Define $q_c(k_{t-\tau}) = 1 - \frac{E[k_t|k_{t-\tau}]}{6}$. As discussed in Section 3.5.2, each cell behaves, to a good approximation, independently of the others. Therefore, the required conditional probability $\pi(k_t|k_{t-\tau})$ is given by

$$\pi(k_t|k_{t-\tau}) = \begin{pmatrix} 6 \\ k_t \end{pmatrix} q_c(k_{t-\tau})^{(6-k_t)}(1-q_c(k_{t-\tau}))^{k_t}.$$

3.A.4 Evaluating Ω_m

Recall from Section 3.5.3 that the mean fade power Ω_m seen by a user, given that he has a packet to transmit, is given by

$$\Omega_m = \frac{E_{\alpha,k}[\alpha^2 A(\alpha,k) + B(\alpha,k)]}{E_{\alpha,k}[A(\alpha,k) + B(\alpha,k)]}, \tag{3.39}$$

where

$$\begin{aligned} A(\alpha,k) &= s(\alpha,k)b(\alpha) + \ldots + d_c(1-s(\alpha,k)b(\alpha))^{d_c-1}s(\alpha,k)b(\alpha) \\ &\quad + d_c(1-s(\alpha,k)b(\alpha))^{d_c}, \\ B(\alpha,k) &= \frac{\overline{\alpha^2}}{\overline{s(\alpha,k)}}(1-s(\alpha,k)b(\alpha))^{d_c}. \end{aligned}$$

Upon simplifying the above expressions, and using the binomial expansion for $s(\alpha, k)$, we get

$$
\begin{aligned}
A(\alpha,k) &= \sum_{i=0}^{d_c-1} e_i p(\eta(\alpha))^i \sum_{h=0}^{iL^{(1)}} (-1)^h c_1^{(1)h} \binom{iL^{(1)}}{h} \exp\left(-hc_2^{(1)} g^{(1)} \left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0} \frac{\alpha^2}{k}\right), \\
B(\alpha,k) &= \sum_{i=0}^{d_c} f_i p(\eta(\alpha))^i \sum_{h=0}^{iL^{(1)}} (-1)^h \binom{iL^{(1)}}{h} c_1^{(1)h} \exp\left(-c_2^{(1)} hg^{(1)} \left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0} \frac{\alpha^2}{k}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
e_i &= \begin{cases} (-1)^{i-1} \sum_{j=0}^{d_c-i} (i+j) \binom{i+j-1}{i-1} - d_c \binom{d_c}{i}, & i \geq 1 \\ 1, & i = 0 \end{cases}, \\
f_i &= (-1)^i \binom{d_c}{i}.
\end{aligned} \tag{3.40}
$$

The value for Ω_m can then be written down from (3.39) using the following relations:

$$
\begin{aligned}
E_{\alpha,k}\left[\alpha^2 \exp\left(-hc_2^{(1)} g^{(1)} \left(\tfrac{d_{\text{av}}}{r}\right)^{\gamma_0} \tfrac{\alpha^2}{k}\right)\right] &= \textstyle\sum_{k=0}^{6} \left(\frac{m}{m+z(k,h)}\right)^{m+1} \pi(k), \\
E_{\alpha,k}\left[\exp\left(-hc_2^{(1)} g^{(1)} \left(\tfrac{d_{\text{av}}}{r}\right)^{\gamma_0} \tfrac{\alpha^2}{k}\right)\right] &= \textstyle\sum_{k=0}^{6} \left(\frac{m}{m+z(k,h)}\right)^{m} \pi(k),
\end{aligned} \tag{3.41}
$$

where

$$
z(k,h) = \frac{c_2^{(1)} hg^{(1)}}{k} \left(\frac{d_{\text{av}}}{r}\right)^{\gamma_0}.
$$

3.A.5 Simulations

The simulations are carried out on a 19 hexagonal cell cluster for a duration of 1.31×10^5 time steps. The users lie on a circle at a distance of $r = 0.75$ from their serving BSs. The packet arrival process generated in the simulations for each user is Bernoulli. At each time step, the SIR for each user at the receiving BS is evaluated from the channel fade and the interference. The channel fading process for the users is generated using a correlated Rayleigh fading ($m = 1$ Nakagami parameter) simulator.

The interference power at the BS receiving the burst is calculated from the knowledge of the transmitting users in the interfering (first tier) BSs. As mentioned in Section 3.4, cross

fading, *i.e.*, fading for the interfering signals, is not considered. Based on the previous value of SIR, a user decides whether to transmit a burst or not. The BS chooses randomly one user from among the users that want to transmit. The burst is taken to be a failure if a random variable, generated so as to be uniformly distributed between 0 and 1, is less than the burst error probability value for that SIR. The fixed coding gain based burst error probability function in (3.9) is used for evaluating the burst error probability.

3.B Saturated Queues Scenario

3.B.1 Throughput μ

Lemma 4

$$\mu = (1-q)\left(\sum_{i=1}^{H} i s^{(i)}\right).$$

Proof: Given that a user is transmitting, the probability that he chooses to transmit in Mode i (i packets in one burst), $1 \leq i \leq H$, and the probability that his transmission is received error-free is $s^{(i)}$. The average number of packets successfully transmitted by a user, given that he is transmitting, is therefore $\sum_{i=1}^{H} i s^{(i)}$.

The probability that a transmission occurs in a cell (by any one of its users) is $1-q$. The BS chooses randomly a user from among the ones that want to transmit. Since the decision by the BS to let a user transmit does not depend on the mode it intends to transmit in, the required result follows.

■

3.B.2 Evaluating q

Lemma 5

$$q = \sum_{k_{t-\tau}=0}^{6} \left[\frac{\Gamma(m, l_0 m k_{t-\tau}\left(\frac{d_{\text{av}}}{r}\right)^{-\gamma_0})}{\Gamma(m)}\right]^n \pi(k_{t-\tau}).$$

Proof: q is the probability that no user in the cell transmits to the serving BS. This

occurs when $\widehat{\mathrm{SIR}}_t^{(i)} < l_0$ for all users in the cell, where $\widehat{\mathrm{SIR}}_t^{(i)}$ denotes the SIR estimate at time t for user i ($1 \leq i \leq n$). Hence,

$$q = \mathcal{P}\left(\widehat{\mathrm{SIR}}_t^{(1)} < l_0, \ldots, \widehat{\mathrm{SIR}}_t^{(n)} < l_0\right), \tag{3.42}$$

$$= E_{k_{t-\tau}}\left[\mathcal{P}\left((\alpha_{t-\tau}^{(1)})^2 < l_0 k_{t-\tau}\left(\frac{d_{\mathrm{av}}}{r}\right)^{-\gamma_0}, \ldots, (\alpha_{t-\tau}^{(n)})^2 < l_0 k_{t-\tau}\left(\frac{d_{\mathrm{av}}}{r}\right)^{-\gamma_0}\right)\right], \tag{3.43}$$

$$= E_{k_{t-\tau}}\left[\Pi_{i=1}^{n}\mathcal{P}\left((\alpha_{t-\tau}^{(1)})^2 < l_0 k_{t-\tau}\left(\frac{d_{\mathrm{av}}}{r}\right)^{-\gamma_0}\right)\right], \tag{3.44}$$

$$= \sum_{k_{t-\tau}=0}^{6}\left(\frac{\Gamma(m, l_0 m k_{t-\tau}\left(\frac{d_{\mathrm{av}}}{r}\right)^{-\gamma_0})}{\Gamma(m)}\right)^n \pi(k_{t-\tau}). \tag{3.45}$$

Eqn. (3.44) follows from (3.43) since the fading processes are independent for different users. Eqn. (3.45) follows from (3.44) from the cdf of the Nakagami distribution in (3.5). ■

3.B.3 Second Order Interference Statistics

The second order interference statistics are evaluated using the on/off cell transmission Markov model of Fig. 3.11. We had used this model for arriving at the second order interference statistics for the stable queue scenario also.

Evaluating $p_{\mathrm{fn}}^{(1)}$ and $p_{\mathrm{nf}}^{(1)}$

$p_{\mathrm{nf}}^{(1)}$ is obtained from $p_{\mathrm{fn}}^{(1)}$ using Lemma 2. To evaluate $p_{\mathrm{fn}}^{(1)}$ we employ an approximation that is motivated from the scenario where the probability p_0 that a user is in Mode 0 is low. When $p_0 = 0$, all the users in the cell will want to transmit. Therefore, a transmission occurs in the cell at every time instant; the cell is never off. For $p_0 > 0$, we assume that a cell is off, *i.e.*, no transmission occurs in the cell, for at most 1 slot. The cell then goes back to the on state. This implies that $p_{\mathrm{fn}}^{(1)} = 1$, a reasonable approximation to make when users in the cell always have packets to transmit and the probability of being in the no-transmission mode is small. The accuracy of this approximation can be verified from a

comparison of the analytical and simulation results presented in Section 3.7.2. As the Mode 0 SIR threshold l_0 increases, the probability that a user wants to transmit will decrease. As a result, the cell will spend a greater fraction of time in the off state, making $p_{\text{fn}}^{(1)} < 1$.

Evaluating $p_{\text{fn}}^{(\tau)}$ and $p_{\text{nf}}^{(\tau)}$

Since the on/off Markov model for transmissions in a cell is the same as that for the stable queues scenario, the relationship between $p_{\text{fn}}^{(\tau)}$ and $p_{\text{nf}}^{(\tau)}$, and $p_{\text{fn}}^{(1)}$ and $p_{\text{nf}}^{(1)}$ is as given in Appendix 3.A.3. In brief, the Markov transition probability matrix $T^{(1)}$ for $\tau = 1$ is

$$T^{(1)} = \begin{bmatrix} 1 - p_{\text{fn}}^{(1)} & p_{\text{fn}}^{(1)} \\ p_{\text{nf}}^{(1)} & 1 - p_{\text{nf}}^{(1)} \end{bmatrix}.$$

The transition probability matrix $T^{(\tau)}$ for a transition time interval of τ instants is therefore

$$T^{(\tau)} = \begin{bmatrix} 1 - p_{\text{fn}}^{(\tau)} & p_{\text{fn}}^{(\tau)} \\ p_{\text{nf}}^{(\tau)} & 1 - p_{\text{nf}}^{(\tau)} \end{bmatrix} = \left(T^{(1)}\right)^{\tau}.$$

Hence, $p_{\text{fn}}^{(\tau)} = T^{(\tau)}(1,2)$ and $p_{\text{nf}}^{(\tau)} = T^{(\tau)}(2,1)$.

Chapter 4 Packet Reservation Multiple Access

The focus of future cellular standards is on increasing capacity, providing higher data rates, supporting multiple services like messaging, e-mail, audio conferencing, and file transfer, and providing Internet access. These standards will eventually support high quality video conferencing and concurrent audio, video, and data services for multimedia. Each of these services may have a different quality of service requirement and a different traffic profile. As a consequence packet based transmission schemes, which offer a flexible and spectrally efficient air interface, have received a lot of attention. Since each packet has an address field which contains information about the packet origin and destination, packet routing through the nodes of the infrastructure network is also easily accomplished [56]. Packet based transmission and switching therefore facilitates network control and eases the burden on the core switching network; this is an issue that is gaining in importance as cellular systems start operating with smaller micro-cells. These schemes also offer statistical multiplexing gains obtained by combining bursty, low average traffic streams. In addition, they offer compatibility with the wireline infrastructure [56].

4.1 Packetized Multiple Access Schemes

In the link adaptation problem considered in Chapter 3, the allocation of a channel for transmission to a user was actively controlled by the base station (BS). Packetized multiple access schemes offer a more distributed solution to this channel resource allocation problem. In such schemes, the decision to transmit a packet is partly made by the users admitted into the system, themselves. A variety of multiple access schemes have been proposed for cellular systems, and studied extensively in the literature [57]. Idle Signal Multiple Access (ISMA) [58, 59], Dynamic TDMA (D-TDMA) [60], Resource Auction Multiple Access (RAMA) [61], Packet Reservation Multiple Access (PRMA) [12], and Centralized Packet Reservation Multiple Access (C-PRMA) [62] are examples of such schemes. We focus

on PRMA in this chapter and analytically evaluate the impact of user mobility and packet errors on its performance.

4.2 Chapter Outline

This chapter is organized as follows. We first describe PRMA and the relevant literature in Section 4.3. The motivation for our work is then described in Section 4.4. The basic model for a voice terminal[1] in PRMA is explained in Section 4.5. The impact of user mobility on PRMA is modeled and analyzed in Section 4.6. The impact of packet errors is then analyzed in Section 4.7. In both of these sections, the modified terminal model is given first, followed by the analysis and the results. We present our conclusions in Section 4.8.

4.3 PRMA Description

Packet Reservation Multiple Access (PRMA) [12] is a multiple access technique for transmitting packets over the short range wireless channels of micro-cells, where propagation delays are negligible. It enables dispersed terminals (users) to transmit fixed length packets to a receiving BS over a shared channel. PRMA is designed to simultaneously handle a mixture of packets from both periodic, delay intolerant sources like voice and aperiodic (bursty), delay tolerant sources like data.

In PRMA, the time axis is organized into slots, whose duration is equal to the packet transmission time. The slots are in turn grouped into frames. The BS classifies the slots as either reserved or available (unreserved). An active terminal that has packets tries, with a probability p, to transmit a packet in an unreserved time slot. The permission probability p is generated autonomously by each user. It is assumed to be a constant, and is an important design parameter.[2] If two or more contending terminals simultaneously transmit a packet in an available slot, a collision is said to occur, and all the packets involved in it are lost. An alternate capture model has also been considered in the literature [12, 65, 66]. In this

[1] We use the terms *terminal* and *user* interchangeably in this chapter.

[2] Techniques for varying it dynamically in order to reduce packet collisions have been investigated in [63, 64], but we do not consider such techniques here.

model, the received packet with the highest energy has a good chance of getting detected accurately by the BS, and is therefore not lost. We shall use a collision model in our analysis.

In case the BS cannot decode the header of an arriving packet, it assumes that the voice terminal has finished its talk spurt and has relinquished its reservation. The BS declares the slot to be unreserved. The main function of the BS is to broadcast binary feedback at the end of each slot indicating whether the next slot is reserved or not. One can therefore see that PRMA requires little central control. The slot reservation information is assumed to be broadcast instantaneously by the BS, since the propagation delay is negligible in micro-cells.

In case a periodic (voice) user's packet is successfully received in an unreserved slot, the slot is reserved in subsequent frames for that user. A voice user voluntarily gives up its reservation whenever its talk spurt ends. This enables PRMA to exploit the silence durations in voice to accommodate more users per slot. No such reservation facility is accorded to delay tolerant, aperiodic (data) users; they need to contend every time they have a packet to transmit. Voice packets delayed beyond a threshold of D s are dropped. Therefore, in PRMA, voice packet dropping probability is the performance metric for voice terminals. For data users, the average packet delay is the performance metric.

Figure 4.1 illustrates the operation of PRMA. It shows a frame with 4 slots. Slots 1 and 3 are reserved. Two terminals A and B contend for slot 2, and as a result neither of their transmissions is successful. Voice terminal A successfully contends for slot 4, which is reserved for it in subsequent frames.

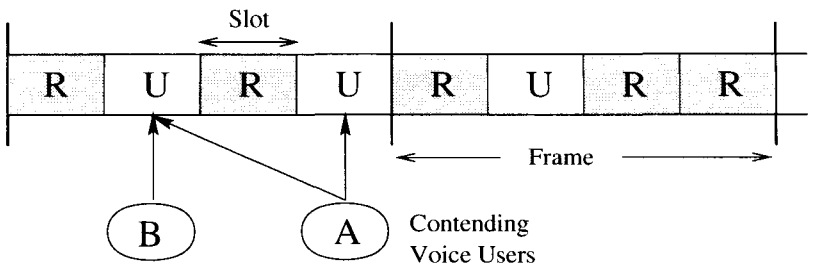

Figure 4.1: Illustration of PRMA

4.3.1 Previous Work

Extensive simulations have been carried out in the literature to study PRMA. Goodman and Wei [67] carried out simulations to examine the effect of several variables like number of channels, permission probability, frame duration, and speech activity detector model on PRMA's performance. Wong [68] simulated PRMA in a metropolitan micro-cellular radio environment to study the impact of call characteristics, mobility, and channel conditions on PRMA. As opposed to a single cell scenario considered in earlier papers, Frullone, Falciasecca, Grazioso, Riva, and Serra [56] simulated PRMA's operation in a cellular environment with fixed channel allocation. Frullone, Riva, Grazioso, and Carciofi [69] simulated PRMA in a cellular system with adaptive channel allocation strategies.

PRMA's simplicity also makes it amenable to analytical study. The explosion in state space of system state based Markov models makes them analytically intractable. Therefore, various approximation techniques have been considered in the literature. Nanda, Goodman, and Timor [70] were the first to analyze PRMA. They evaluated the voice packet dropping probability as a function of the number of simultaneous conversations in progress in a cell. Qi and Wyrwas [71] tried to extend this analysis to analyze the effect of packet errors on PRMA. An approach different from that in [71] was adopted by Wang, Wang, and Sukhbaatar [72] to analyze the capacity of PRMA with packet errors. Qiu and Li [66] improved upon this analysis to study the improvement in PRMA's performance due to packet capture.

4.4 User Mobility and Packet Errors in PRMA

When a voice user moves from one cell to another, he loses his slot reservation even though he still has packets to transmit. Packet errors, which are quite common when transmitting information over a harsh wireless channel, occur when the receiver (the BS in our case) fails to decode the transmitted set of bits. Channel fading and interference are primarily responsible for this. The critical information about a packet's source, destination, etc., is contained in its header, while the remaining part of a packet contains the data that is to be

sent. The implications of errors in the header part and data part are markedly different. Specifically, a premature loss of reservation, similar to that due to user mobility, occurs when the header part of a voice user's packet is corrupted. When the BS does not receive a packet due to an error in the packet's header, it assumes that the voice user has finished its talk spurt, and it declares the slot to be unreserved. No such loss of reservation occurs when only the data part of the packet is corrupted.

We develop an approximate analytical technique to analyze the impact of user mobility and packet errors on PRMA. Our analysis handles the fixed packet error rate model of [67, 68, 71] and also the interesting interference-limited cellular system model where the packet error rates are dependent on the interference, and vice versa. As we explain later, our analysis avoids many of the approximations made in [71], and is therefore more accurate.

4.5 Voice Terminal's Basic Model

We now present the basic model for a voice terminal operating in PRMA. This model does not include the effects of packet errors and mobility. We will modify the model in subsequent sections to include these effects.

The state of a voice terminal in PRMA has been modeled as a Markov chain in [12, 67]. A speech source creates a series of talk spurts and gaps during a conversation. It is classified by the speech activity detector [73, 12, 67] to either be in the *silent* mode or the *talk spurt* mode. Voice packets[3] are generated only in the talk spurt mode. This model is called the slow speech activity detector model. An alternate model considered in literature is the fast speech activity detector [74, 12, 67] where a talk spurt further consists of mini-silence durations. Since the slow speech activity detector has been shown to be more robust[4] than the fast one for PRMA [67], we shall use the former in our analysis.

A time slot is the basic unit of time in a PRMA system. Terminal mode transitions occur at the end of each time slot. The mode diagram for a terminal is shown in Fig. 4.2.

[3]Each voice packet consists of voice sampled over the duration of a frame.

[4]Voice terminals that use fast speech activity detectors can give up their reservations even during the mini-silence durations in a talk spurt, thus lowering the voice activity factor. However, the additional contention overhead annuls the statistical multiplexing gains that are possible due to the lower voice activity factor.

In the figure, the branch labels are the respective mode transition probabilities. When the terminal is inactive, it is in the silent mode $\mathcal{S}$. At the beginning of a talk spurt, it enters the contention mode $\mathcal{C}$. The terminal keeps contending for the next unreserved slot until it manages to successfully transmit a packet in one of them. In order to successfully transmit a packet in one of the unreserved slots, the following must be true:

- The terminal must have permission to transmit.
- None of the other contending terminals should have permission to transmit.
- The slot must be unreserved.
- The terminal's talk spurt must not end.

The slot in which the terminal makes its first successful transmission is reserved for it in subsequent frames. The terminal then enters the reservation mode $\mathcal{R}_{n-1}$. The subscript i in $\mathcal{R}_i$ is the number of slots that remain before the terminal gets to transmit again. At the end of a time slot, a terminal in mode $\mathcal{R}_{n-1}$ will, with probability 1, go to mode $\mathcal{R}_{n-2}$ since the slot reserved for it is now $n-2$ slots away. Similarly, from $\mathcal{R}_i$ ($i \geq 1$), the terminal will go to $\mathcal{R}_{i-1}$. From mode $\mathcal{R}_0$, the terminal returns to $\mathcal{R}_{n-1}$ if it has more packets to transmit. Otherwise, it returns to the silent mode. A contending terminal, *i.e.*, one that has not managed to secure a slot reservation, will re-enter the silent mode $\mathcal{S}$, dropping all the packets that are yet to be transmitted, when its talk spurt ends.

4.6 Impact of User Mobility

4.6.1 Terminal Model with Mobility

Mobility can cause the terminal to change cells during any time slot. We modify the terminal model of Fig. 4.2 to incorporate the effect of this hand-off from one cell to another. The modified model is shown in Fig. 4.3.

During any time slot, an active terminal can hand-off from one cell to another with probability m. The channel quality of a terminal at the cell boundary is typically poor and

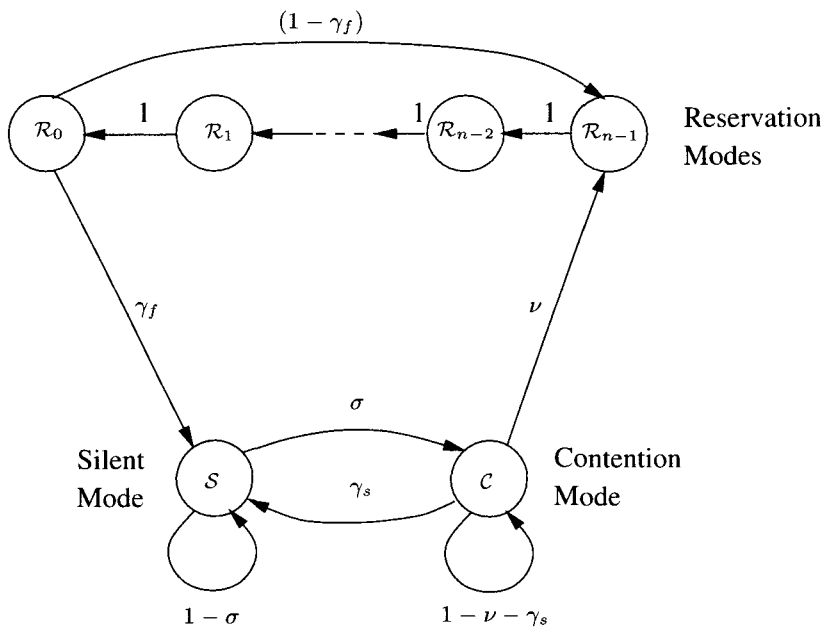

Figure 4.2: Basic terminal model

can result in his not being able to transmit. Hand-off is therefore modeled as a delay of y slots: $\mathcal{B}_1, \ldots, \mathcal{B}_y$. Any time a hand-off occurs, the terminal loses its reservation and transits to mode $\mathcal{B}_y$. From $\mathcal{B}_i$ $(2 \leq i \leq y)$ the terminal goes to $\mathcal{B}_{i-1}$, with probability 1, after one time slot. From $\mathcal{B}_1$ the terminal reenters the contention mode $\mathcal{C}$ if it has any more packets to transmit. The terminal in $\mathcal{B}_1$ goes back to $\mathcal{B}_y$ if it again changes cells during the y slot delay.[5] Note that the transition from $\mathcal{S}$ to $\mathcal{B}_y$ is not considered since a terminal in silent mode does not have packets to transmit. This assumption also ensures that the effect of hand-off is not felt over multiple talk spurts.

When there is no terminal mobility, the number of packets transmitted during a talk spurt is solely a function of the delay encountered in transmitting a packet successfully from the contention mode.[6] Subsequent packets are guaranteed reserved slots for transmission, and are therefore not dropped. This is the case analyzed by Nanda, Goodman, and

[5]A more accurate model would allow for transitions from any $\mathcal{B}_i$ to $\mathcal{B}_y$. However, this greatly increases the complexity of the analysis in terms of the number of closed loops that need to be considered in the corresponding signal flow graph in Fig. 4.4, which we describe later. It is therefore avoided by making the transition rate from $\mathcal{B}_1$ to $\mathcal{B}_y$ to account for all mobility induced mode transitions that can occur in the y slot delay period.

[6]This transmission corresponds to the transition from mode $\mathcal{C}$ to mode $\mathcal{R}_{n-1}$.

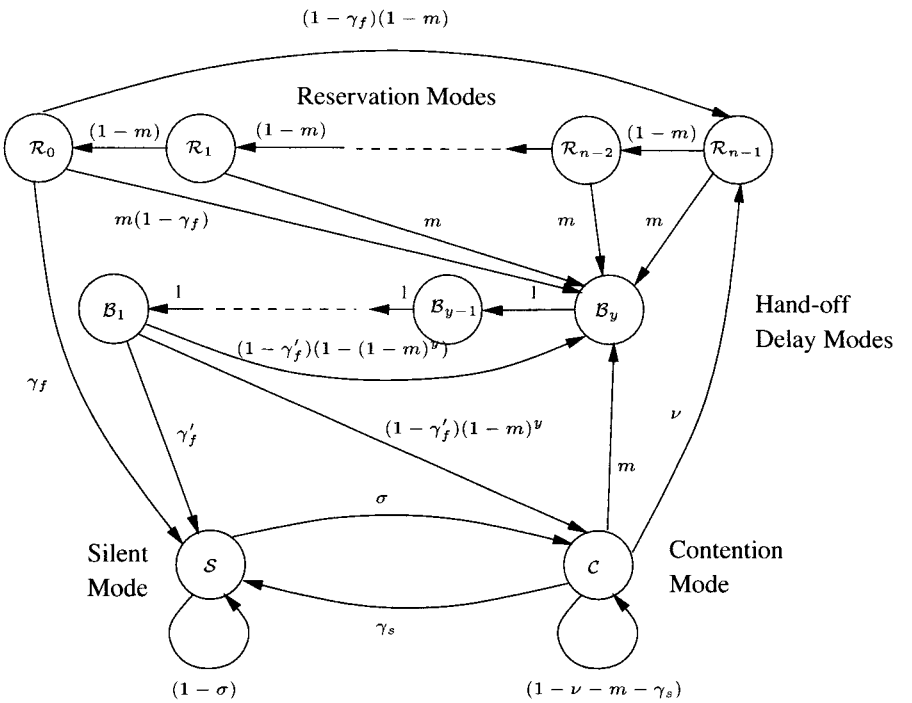

Figure 4.3: Terminal model with mobility

Timor [70]. The main difference that one encounters when terminal mobility is taken into account is that subsequent packets can get dropped due to a premature loss of reservation when the terminal moves into another cell. We therefore develop a path enumeration technique based on signal flow graphs to analyze the resulting packet dropping probability. We concentrate on evaluating the performance of PRMA for voice users only. Techniques for also handling data users are available in the literature [70, 71].

4.6.2 Packet Dropping Probability Evaluation

We first evaluate the voice packet dropping probability $\pi(\nu)$, given the probability ν that a contending terminal successfully transmits a packet.

Without loss of generality, let the first packet of a talk spurt be generated at time 0. The i^{th} packet of a talk spurt will be generated at time $n(i-1)$, and must be transmitted by time $d + n(i-1)$ to avoid violating the delay constraint. Similarly, for a talk spurt of length q, the last packet must be transmitted by time $n(q-1) + d$.

Let $a_{\mathcal{A}}(i,j)$ denote the probability that a terminal in mode $\mathcal{C}$ at time 0 ends up in mode $\mathcal{A}$ at time j and has transmitted exactly i packets successfully. After time nq, if the terminal is in any of the unreserved modes, viz. $\mathcal{B}_k$ $(1 \leq k \leq y)$ or $\mathcal{C}$, it will return to mode $\mathcal{S}$ without transmitting any more packets. Therefore, for modes $\mathcal{C}$ and $\mathcal{B}_k$, the average number of packets dropped in a given talk spurt of length q is $\sum_{i=0}^{q-1} \sum_{\mathcal{A}\in\{\mathcal{B}_1,\ldots,\mathcal{B}_y,\mathcal{C}\}} (q-i) a_{\mathcal{A}}(i,nq)$. If the terminal is instead in any of the reserved modes $\mathcal{R}_w$ $(0 \leq w \leq n-1)$, it can continue to transmit its remaining packets so long as it does not lose its slot reservation. Let $b_w^{(q-i)}(h)$ denote the probability that the terminal manages to transmit h more packets, given that it has transmitted i $(i < q)$ packets and is in mode $\mathcal{R}_w$ by time nq.

The average number of packets in a talk spurt is $1/\gamma_f$. The talk spurt length q of a terminal is a geometrically distributed random variable. The probability that the talk spurt is q packets long is $\gamma_f(1-\gamma_f)^{q-1}$. Upon averaging over the talkspurt lengths, $\pi(\nu)$ takes the form

$$\begin{aligned}\pi(\nu) &= \gamma_f \sum_{q=1}^{\infty} \sum_{i=0}^{q-1} \sum_{\mathcal{A}\in\{\mathcal{B}_1,\ldots,\mathcal{B}_y,\mathcal{C}\}} (q-i) a_{\mathcal{A}}(i,nq) \gamma_f (1-\gamma_f)^{q-1} \\ &\quad + \gamma_f \sum_{q=1}^{\infty} \sum_{i=0}^{q-1} \sum_{w=0}^{n-1} a_{\mathcal{R}_w}(i,nq) \left(\sum_{h=0}^{\min(\tau-1,q-i)} (q-i-h) b_w^{(q-i)}(h) \right) \gamma_f (1-\gamma_f)^{q-1}, \end{aligned} \tag{4.1}$$

where $d = \tau n$. $b_w^{(q-i)}(h)$, derived in Appendix 4.A.1, is given by

$$b_w^{(q-i)}(h) = \begin{cases} 1-(1-m)^{w+1}, & h = 0 \\ (1-m)^{w+1+n(h-1)}(1-(1-m)^n), & 1 \leq h < \min(q-i, \tau-1) \\ (1-m)^{(h-1)n+w+1}, & h = \min(q-i, \tau-1) \end{cases} . \tag{4.2}$$

For the sake of simplicity, we will assume a delay constraint $d = 2n$ for the remaining derivations. A similar reasoning applies for other values of $d = \tau n$ $(\tau \in Z^+)$. For $d = 2n$ $(\tau = 2)$, we have $b_w^{(q-i)}(1) = (1-m)^{w+1}$ and $b_w^{(q-i)}(0) = 1-(1-m)^{w+1}$.

In Fig. 4.3, a terminal can reach the reservation mode $\mathcal{R}_i$ only by first passing through mode $\mathcal{R}_{n-1}$. This implies that $a_{\mathcal{R}_w}(i,j) = (1-m)^{n-w-1} a_{\mathcal{R}_{n-1}}(i, j-n+w+1)$, for

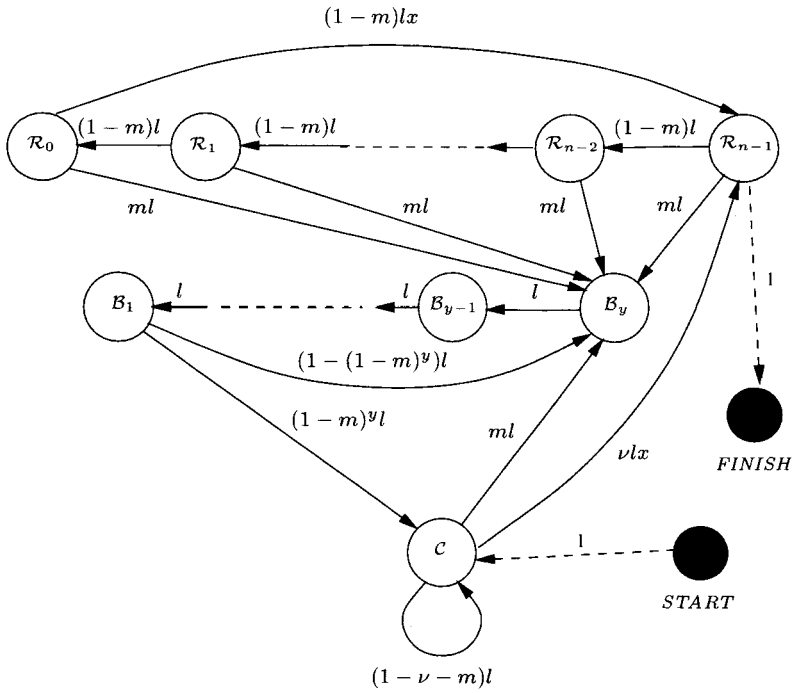

Figure 4.4: Terminal mode diagram, with mobility, during a talk spurt

$0 \leq w \leq n-1$. Simplifying (4.1) yields

$$\begin{aligned}\pi(\nu) &= \gamma_f^2 \sum_{q=1}^{\infty} \sum_{i=0}^{q-1} (q-i) \left[\sum_{\mathcal{A} \in \Omega} a_{\mathcal{A}}(i, nq) \right] (1-\gamma_f)^{q-1} \\ &\quad - \gamma_f^2 (1-m)^n \sum_{q=1}^{\infty} \sum_{i=0}^{q-1} \left[\sum_{w=1}^{n} a_{\mathcal{R}_{n-1}}(i, n(q-1)+w) \right] (1-\gamma_f)^{q-1}, \quad (4.3)\end{aligned}$$

where $\Omega = \{\mathcal{C}, \mathcal{R}_0, \ldots, \mathcal{R}_{n-1}, \mathcal{B}_1, \ldots, \mathcal{B}_y\}$ is the set of modes a terminal can be in during a talk spurt.

4.6.3 Generating Functions

To evaluate $\pi(\nu)$ we thus need to evaluate $a_\Omega(i,j) = \sum_{\mathcal{A} \in \Omega} a_{\mathcal{A}}(i,j)$ and $a_{\mathcal{R}_{n-1}}(i,j)$. We do so by first evaluating their generating functions $g_\Omega(x,l) = \sum_i \sum_j a_\Omega(i,j) x^i l^j$ and $g_{\mathcal{R}_{n-1}}(x,l) = \sum_i \sum_j a_{\mathcal{R}_{n-1}}(i,j) x^i l^j$.

Since a terminal does not enter the silent mode $\mathcal{S}$ during a talk spurt, it suffices to

consider the truncated mode diagram in Fig. 4.4 for calculating the average number of packets dropped in a talk spurt.[7] The branch labels of the mode transition diagram of a terminal, now interpreted as a signal flow graph, are modified as follows.

- Every branch label is multiplied by an indeterminate l to measure the length of a path traversing through the mode space.
- In addition, the gains for branches that correspond to mode transitions in which a successful packet transmission occurs are multiplied by indeterminate x. Therefore, the two branches ending in $\mathcal{R}_{n-1}$, viz. $\mathcal{R}_0 \rightarrow \mathcal{R}_{n-1}$ and $\mathcal{C} \rightarrow \mathcal{R}_{n-1}$, bear x also.

Consequently, a path of length j slots in which i packets get transmitted will have a factor $x^i l^j$ in its path gain. This technique is similar to the one used for evaluating the weight enumerators of convolutional error correction codes [44, Chp. 11].

Let $g_{\mathcal{A}}(x,l)$ denote the generating function of path gains for all paths starting at $\mathcal{C}$ and ending in a given mode $\mathcal{A}$. $g_{\mathcal{A}}(x,l)$ can be calculated by applying Mason's gain formula [75, Chp. 4] as given below.

$$g_{\mathcal{A}}(x,l) = \frac{1}{\Delta(x,l)} \sum_{q \in \mathcal{F}} F_q(x,l)\Delta_q(x,l), \tag{4.4}$$

where $F_q(x,l)$ is the *forward path gain* of the forward path q from $START$ (appended to $\mathcal{C}$) to $FINISH$ (appended to $\mathcal{A}$), $\Delta_q(x,l)$ is the *co-factor* of the path q, $\mathcal{F}$ is the set of forward paths, and $\Delta(x,l)$ is the *graph determinant*. Appendix 4.A.2 derives the expression for $g_{\mathcal{A}}(x,l)$ from Fig. 4.4.

Mason's gain formula sums the gains of only those paths that start at $\mathcal{C}$ and end at $\mathcal{A}$ and for which none of the intermediate modes are $\mathcal{C}$ or $\mathcal{A}$. In our case, however, we must also include the gains of paths in which the terminal can go through modes $\mathcal{C}$ and $\mathcal{A}$ any number of times. For example, a terminal can stay (loop around) in $\mathcal{C}$ for a multiple number of time slots before it jumps to $\mathcal{R}_{n-1}$. It is for this reason that we append two additional dummy modes $START$ and $FINISH$, and evaluate the gain from $START$ to $FINISH$. $START$ is connected to $\mathcal{C}$ by a branch with gain 1. Similarly, $FINISH$ is connected to $\mathcal{A}$ by a branch

[7] The role of the additional modes $START$ and $FINISH$ is explained in the following paragraphs.

of unity gain. All the required path gains are now included in the gain formula. Fig. 4.4 illustrates the technique for $\mathcal{A} = \mathcal{R}_{n-1}$.

In the signal flow graph in Fig. 4.4, the loops that have a factor x in their loop gains have the mode $\mathcal{R}_{n-1}$ in common, and therefore cannot be non-touching loops [75, Chp. 4]. It follows that $g_\Omega(x,l)$ and $g_{\mathcal{R}_{n-1}}(x,l)$ can be written in the form

$$g_\Omega(x,l) = \frac{\alpha_\Omega(l) + \beta_\Omega(l)x}{\gamma(l) + \delta(l)x}, \tag{4.5}$$

$$g_{\mathcal{R}_{n-1}}(x,l) = \frac{\alpha_{\mathcal{R}_{n-1}}(l) + \beta_{\mathcal{R}_{n-1}}(l)x}{\gamma(l) + \delta(l)x}, \tag{4.6}$$

where $\alpha_\Omega(l)$, $\beta_\Omega(l)$, $\alpha_{\mathcal{R}_{n-1}}(l)$, $\beta_{\mathcal{R}_{n-1}}(l)$, $\gamma(l)$, and $\delta(l)$ are all polynomials in l. This form enables us to use a simple recursion based technique for evaluating the bivariate power series expansions in x and l of $g_\Omega(x,l)$ and $g_{\mathcal{R}_{n-1}}(x,l)$.

4.6.4 Recursion

Let $a_\Omega^{(i)}(l)$ denote the coefficient of x^i in the series expansion of $g_\Omega(x,l)$. Then $g_\Omega(x,l) = \sum_i a_\Omega^{(i)}(l)x^i$. Note that $a_\Omega^{(i)}(l)$ is itself a polynomial in l. From (4.5), it is easy to see that

$$a_\Omega^{(i)}(l) = \begin{cases} \frac{\alpha_\Omega(l)}{\gamma(l)}, & i = 0 \\ \frac{\beta_\Omega(l)\gamma(l) - \alpha_\Omega(l)\delta(l)}{\gamma(l)^2}\left(-\frac{\delta(l)}{\gamma(l)}\right)^{(i-1)}, & i \geq 1 \end{cases}. \tag{4.7}$$

Therefore, for $i \geq 2$ we have the recursion relation

$$a_\Omega^{(i)}(l) = a_\Omega^{(i-1)}(l)\left(-\frac{\delta(l)}{\gamma(l)}\right). \tag{4.8}$$

Similar formulae apply to $a_{\mathcal{R}_{n-1}}^{(i)}(l)$ also.

4.6.5 Simplification of $\pi(\nu)$ Formula

$\delta(l)$ is the coefficient of x in the graph determinant $\Delta(x,l)$; so, it appears in the denominators of $g_\Omega(x,l)$ and $g_{\mathcal{R}_{n-1}}(x,l)$. The graph determinant is a function of the loop gains of the closed loops in the signal flow graph. In the mode diagram in Fig. 4.4, the minimum length

of a closed loop in which at least one packet gets transmitted is[8] n. Therefore, $\delta(l)$ can be written as $l^n\delta_s(l)$, where $\delta_s(l)$ is itself a polynomial in l. Instead of $a_\Omega^{(i)}(l)$, we therefore recursively evaluate the function $f_\Omega^{(i)}(l)$ defined by

$$\begin{aligned} f_\Omega^{(0)}(l) &= \tfrac{\alpha_\Omega(l)}{\gamma(l)} &&= a_\Omega^{(0)}(l), \\ f_\Omega^{(1)}(l) &= \tfrac{\beta_\Omega(l)\gamma(l)-\alpha_\Omega(l)\delta(l)}{\gamma(l)^2} &&= a_\Omega^{(1)}(l), \\ f_\Omega^{(i)}(l) &= -\tfrac{\delta_s(l)}{\gamma(l)} f_\Omega^{(i-1)}(l) &&= l^{-n(i-1)} a_\Omega^{(i)}(l),\ i \geq 2. \end{aligned} \tag{4.9}$$

Let $f_\Omega(i,j)$ and $f_{\mathcal{R}_{n-1}}(i,j)$ denote the coefficients of l^j in $f_\Omega^{(i)}(l)$ and $f_{\mathcal{R}_{n-1}}^{(i)}(l)$, respectively. Then

$$\begin{aligned} a_\Omega(i,j) &= \begin{cases} f_\Omega(0,j), & i=0 \\ f_\Omega(i,j-n(i-1)), & i \geq 1 \end{cases}, \\ a_{\mathcal{R}_{n-1}}(i,j) &= \begin{cases} f_{\mathcal{R}_{n-1}}(0,j), & i=0 \\ f_{\mathcal{R}_{n-1}}(i,j-n(i-1)), & i \geq 1 \end{cases}. \end{aligned} \tag{4.10}$$

The formula for $\pi(\nu)$ in (4.1) can now be simplified as follows.

$$\begin{aligned} \pi(\nu) &= \gamma_f^2 \sum_{q=1}^{\infty}\sum_{i=0}^{q-1}(q-i)a_\Omega(i,nq)(1-\gamma_f)^{q-1} \\ &\quad - \gamma_f^2(1-m)^n \sum_{q=1}^{\infty}\sum_{i=0}^{q-1}\left[\sum_{w=1}^{n} a_{\mathcal{R}_{n-1}}(i,n(q-1)+w)\right](1-\gamma_f)^{q-1}, \\ &= \gamma_f^2 \sum_{q=1}^{\infty} q f_\Omega(0,nq)(1-\gamma_f)^{q-1} \\ &\quad + \gamma_f^2 \sum_{i=1}^{\infty}\sum_{q=i+1}^{\infty}(q-i)f_\Omega(i,n(q-i+1))(1-\gamma_f)^{q-1} \\ &\quad - \gamma_f^2(1-m)^n \sum_{i=1}^{\infty}\sum_{q=i+1}^{\infty}\left[\sum_{w=1}^{n} f_{\mathcal{R}_{n-1}}(i,n(q-i)+w)\right](1-\gamma_f)^{q-1}, \\ &= \gamma_f^2(1-m)^n \sum_{i=1}^{\infty}\sum_{t'=1}^{\infty}\left[\sum_{w=1}^{n} f_{\mathcal{R}_{n-1}}(i,nt'+w)\right](1-\gamma_f)^{i+t'-1} \end{aligned}$$

[8] A terminal can transmit only once in n slots. This also imposes the condition that $y \geq n$.

$$+\gamma_f^2 \sum_{q=1}^{\infty} q f_\Omega(0, nq)(1-\gamma_f)^{q-1}$$
$$+\gamma_f^2 \sum_{i=1}^{\infty}\sum_{t'=1}^{\infty} t' f_\Omega(i, n(t'+1))(1-\gamma_f)^{i+t'-1}, \tag{4.11}$$

where $q-i$ is substituted by t' in the last equation. In Appendix 4.A.3, we show that (4.11) reduces to the one derived in [70] for the special no mobility case. In general, $\pi(\nu)$ for any $d = \tau n$ is given by

$$\begin{aligned}\pi(\nu) &= \gamma_f^2 \sum_{q=1}^{\infty} q f_\Omega(0, nq)(1-\gamma_f)^{q-1} \\ &\quad + \gamma_f^2 \sum_{i=1}^{\infty}\sum_{t'=1}^{\infty} t' f_\Omega(i, n(t'+1))(1-\gamma_f)^{i+t'-1} \\ &\quad - \gamma_f^2 \sum_{h=1}^{\tau-1}(1-m)^{nh} \sum_{t'=h}^{\infty}\sum_{i=1}^{\infty}\left[\sum_{w=1}^{n} f_{\mathcal{R}_{n-1}}(i, nt'+w)\right](1-\gamma_f)^{i+t'-1}\end{aligned} \tag{4.12}$$

It follows from Section 4.6.4 that the power series expansion of the rational function $f_\Omega^{(i)}(l)$ $(i \geq 2)$ is the output of an Infinite Impulse Response (IIR) filter [76, Chp. 2] with $f_\Omega^{(i-1)}(l)$ as the input and impulse response[9] $-\frac{\delta_s(l)}{\gamma(l)}$, as shown in Fig. 4.5. It is important to note that we are dealing with rational functions that may have some of their poles close to the unit circle. Therefore, numerically stable filtering techniques [76, Chp. 6] need to be used to evaluate the filter responses.

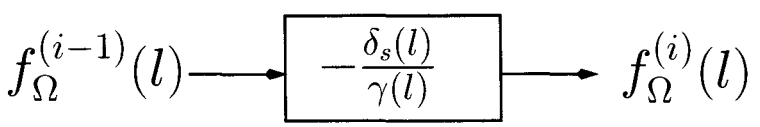

Figure 4.5: IIR filter implementation

[9]The variable l is equivalent to the variable z^{-1} that is typically used in the notation dealing with filters.

4.6.6 System Probability Distribution Estimate

In the previous section we derived the packet dropping probability $\pi(\nu)$ for a given successful packet transmission probability ν. The packet dropping probability evaulated only at a single state does not give a good estimate of its actual value [70]. We therefore derive a system probability distribution for the system state variables. $\pi(\nu(C, R))$ is evaluated for each of the system states, and is subsequently unconditioned according to the probability distribution estimate.

For a given cell, the state vector is given by $(S, C, R_0, \ldots, R_{n-1}, B_1, \ldots, B_y)$. Here, S is the number of terminals in a cell in silent mode $\mathcal{S}$, C is the number of terminals in mode $\mathcal{C}$, R_i is the number of terminals in reservation mode $\mathcal{R}_i$ $(0 \leq i \leq n-1)$, and B_i is the number of terminals in hand-off delay mode $\mathcal{B}_i$ $(1 \leq i \leq y)$. Let $R = \sum_{i=0}^{n-1} R_i$, denote the total number of reserved slots in a frame. A complete multi-dimensional Markov analysis of the system is intractable because it has to handle, approximately, $2^n t^{y+2}$ states. We therefore use an approximate analytical technique based on Equilibrium Point Analysis (EPA) [77, Chp. 2] to arrive at the system state probability distribution. EPA was first used for analyzing PRMA in [70]. Improvements to the EPA based approximation have also been suggested in the literature. Qi and Wyrwas [71] constructed a transition matrix for a Markov chain with state vector (R, C). Qiu and Li [66] and Wang, Wang, and Sukhbaatar [72] instead made a stationarity assumption for a marginal Markov process (distribution of R given S) to derive an approximate distribution. We first describe EPA and then come up with the required probability distribution along the lines suggested by [70].

4.6.7 Equilibrium Point Analysis (EPA)

Equilibrium Point Analysis was proposed by Fukuda [78] and used extensively by Tasaka [77] to analyze multiple access protocols. An equilibrium state is defined to be a state in which the expected change in the number of terminals in each mode is zero. The advantage of EPA is that only $n + y + 2$ equations need to be solved.

Let s denote the equilibrium value of S, the number of terminals in mode $\mathcal{S}$. Similarly, let c, r_i, and b_i denote the equilibrium values of C, R_i $(0 \leq i \leq n-1)$, and B_i $(1 \leq i \leq y)$,

respectively. For a terminal to leave $\mathcal{C}$ by means of a successful packet transmission, all of the following conditions must be true.

- The terminal can only transmit in a time slot that is unreserved. The probability that the slot is not reserved is $(1 - r_0)$.
- The terminal must have permission to transmit. This permission is generated with probability p at each of the terminals independently.
- None of the other terminals that are also contending for a slot have permission to transmit in that slot. If there are k other contending terminals, no collision occurs with probability $(1 - p)^k$.
- The terminal must not move into another cell. The probability that it does not move into another cell is $(1 - m)$.
- The terminal must not return to the silent mode. The probability that the terminal does not return to $\mathcal{S}$ is $(1 - \gamma_s)$.

Therefore, the transition probability from $\mathcal{C}$ to $\mathcal{R}_{n-1}$ is

$$\begin{aligned} \nu &= u(c)p(1 - r_0)(1 - m)(1 - \gamma_s), \text{ where} \\ u(c) &= \begin{cases} (1 - p)^{(c-1)}, & c \geq 1 \\ 1, & c < 1 \end{cases}. \end{aligned} \tag{4.13}$$

From Fig. 4.3, the equilibrium equations, obtained by equating the expected increase and decrease in the number of terminals in a given mode, are then given as follows.

At mode $\mathcal{S}$

$$s\sigma = r_0\gamma_f + b_1\gamma_f' + c\gamma_s, \tag{4.14}$$

At mode $\mathcal{C}$

$$c(\nu + m + \gamma_s) = \sigma s + b_1(1 - \gamma_f')(1 - m)^y, \tag{4.15}$$

At mode $\mathcal{R}_{n-1}$

$$r_{n-1} = c\nu + r_0(1 - \gamma_f)(1 - m), \tag{4.16}$$

At mode $\mathcal{R}_i$ $(0 \leq i \leq n-2)$

$$r_i = r_{i+1}(1-m), \tag{4.17}$$

At mode $\mathcal{B}_y$

$$b_y = b_1(1-(1-m)^y)(1-\gamma_f') + m(1-\gamma_f)r_0 + m\sum_{i=1}^{n-1} r_i + cm, \tag{4.18}$$

At mode $\mathcal{B}_i$ $(1 \leq i \leq y-1)$

$$b_i = b_{i+1}. \tag{4.19}$$

Let t denote the average number of terminals in a cell. We therefore have an additional relation

$$t = \sum_{i=0}^{n-1} r_i + s + c + \sum_{i=1}^{y} b_i. \tag{4.20}$$

Expressing all variables in terms of c, we get a non-linear equation of the form $F(c) = t$ that is solved to obtain c. Equilibrium values of the other system variables are then obtained using (4.13)–(4.19).

Notation: Let $\theta_{A|B}(a)$ denote the probability that the random variable A takes a value a conditioned on B. Let $\mathcal{P}(E)$ denote the probability of occurrence of event E. Let $\Theta_A(x)$ denote the moment function [25, Chp. 5] of random variable A.

Having obtained $\pi(.)$ from (4.11), the average packet dropping probability is given by

$$\overline{\pi} = \sum_{R=0}^{n}\sum_{C=1}^{t-R} \pi(\nu(C,R))\theta_{C|R}(C)\theta_R(R). \tag{4.21}$$

The successful transmission probability as a function of R and C is given by

$$\nu(C,R) = (1-\theta_{R_0|R}(1))p(1-p)^C(1-m), \tag{4.22}$$

where $\theta_{R_0|R}(1)$ is the probability that the current slot, in which a terminal wants to transmit, is reserved, *i.e.*, $\mathcal{R}_0 = 1$, given that a total of R slots are reserved. We now derive expressions for $\theta_{R_0|R}(1)$, $\theta_R(R)$, and $\theta_{C|R}(C)$.

Distribution of R_0 Given R, $\theta_{R_0|R}(.)$

The total number of reserved slots R in a cell is a sum of n random variables $R_0, \ldots, R_{n-1}$. R_i is a binary valued random variable with $r_i = \mathcal{P}(R_i = 1)$. Since each user generates the permission probability independently of other users, the n random variables are, to a good approximation, independent [70].[10] However, they are not identically distributed. We now derive the probability distribution $\theta_{R_0|R}(1)$.

In a cell, given the total number of reserved slots R in a frame, the probability $\theta_{R_0|R}(1)$ is given by Baye's theorem [25, Chp. 2] as

$$\theta_{R_0|R}(1) = \frac{\theta_{R|R_0=1}(R)\theta_{R_0}(1)}{\theta_{R_0}(1)\theta_{R|R_0=1}(R) + \theta_{R_0}(0)\theta_{R|R_0=0}(R)}. \tag{4.23}$$

Since $R_0, \ldots, R_{n-1}$ are independent, we have

$$\begin{aligned} \theta_{R|R_0=1}(R) &= \mathcal{P}\left(\sum_{i=1}^{n-1} R_i = R-1\right), \text{ and} \\ \theta_{R|R_0=0}(R) &= \mathcal{P}\left(\sum_{i=1}^{n-1} R_i = R\right). \end{aligned}$$

The independence assumption implies that the moment function $\Theta_{n-1}(x)$ of $\sum_{i=1}^{n-1} R_i$ is given by $\Theta_{n-1}(x) = \prod_{i=1}^{n-1}(r_i x + (1-r_i))$, where $\Theta_{R_i}(x) = (r_i x + (1-r_i))$ is the moment function of the binary valued random variable R_i.[11] Hence,

$$\theta_{R_0|R}(1) = \frac{1}{\left(1 + \frac{(1-r_0)}{r_0}\frac{\kappa(R-1)}{\kappa(R)}\right)}, \tag{4.24}$$

where $\kappa(j)$ is the coefficient of x^j in $\Theta_{n-1}(x)$.[12]

[10] $R_0, \ldots, R_{n-1}$ are independent since different time slots are occupied by different terminals that access channels independently.

[11] From (4.17), $r_i = r_0(1-m)^i$.

[12] When the n random variables are also identically distributed, (4.24) reduces to $\theta_{R_0|R}(1) = R/n$.

Distribution of R, $\theta_R(.)$

The probability distribution of R can be obtained from its moment function $\Theta_R(x) = \prod_{i=0}^{n-1}(r_i x + (1 - r_i))$. This follows from the fact that $\theta_R(k)$ is the coefficient of x^k in $\Theta_R(x)$.

Distribution of C Given R, $\theta_{C|R}(.)$

When the number of contending terminals is small, the number of contending users C, waiting in a queue to get a reservation, can be assumed to be geometrically distributed [70]. We then get the following distribution for C given R.

$$\theta_{C|R}(C) = \begin{cases} p_0(1-p_0)^C, & C < t - R - \lceil yb_y \rceil \\ (1-p_0)^C, & C = t - R - \lceil yb_y \rceil \\ 0, & \text{otherwise} \end{cases}, \tag{4.25}$$

where $p_0 = \frac{1}{c+1}$. The above approximation assumes that the rate at which the customers get serviced in the queue is independent of queue length. In PRMA, the successful packet transmission probability in fact decreases with an increase in the number of contending terminals since the number of collisions increases. Therefore, the approximation fails when the number of terminals is large. The total number of terminals t has been reduced by the term $\lceil yb_y \rceil$ to account for the terminals in the hand-off delay slots. Note that the tail of the geometric distribution has been modified to ensure that C is finite [70].

4.6.8 Results

We assume the following values for the system parameters that are shown in Fig. 4.3: average talk spurt duration $t_a = 0.50$ s, average silence duration $t_q = 1.35$ s, frame duration $t_f = 0.016$ s, number of slots in a frame $n = 10$, number of delay train slots $y = 10$, and the maximum voice packet transmission delay $d = 0.032$ s (20 time slots). Therefore, the time slot duration t_s is $t_f/n = 0.0016$ s. Since the terminal model is Markov, the probability σ that a silent terminal enters talk spurt mode is given by $1 - \exp(-t_s/t_q) = 0.00118$. The

probability γ_s that a terminal in mode $\mathcal{C}$ returns to the silent mode is $1 - \exp(-t_s/t_a) = 0.0032$. The probability γ_f that a terminal that has obtained a slot reservation returns to the silent mode is then $1 - (1 - \gamma_s)^n = 0.032$. Similarly, the probability γ'_f that a terminal's talk spurt does not end during its y slot hand-off delay is $1 - (1 - \gamma_s)^y = 0.032$.

Let H denote the average number of cells (excluding the one in which the talk spurt originated) a terminal visits during a talk spurt. The probability m that a mobile moves to another cell during a talk spurt is therefore Ht_s/t_a. We shall use H to compare the results for different user mobility rates.

Figure 4.6 shows the packet dropping probability, derived in (4.44), as a function of the number of terminals t for different mobility rates H when permission probability p is 0.1. It can be seen that the packet dropping probability increases by at most 10% when H increases from 0 to 0.05, and by at most another 10% as H increases from 0.05 to 0.1. Figure 4.7 shows the same for $p = 0.4$. For large t, notice that the curves become the same irrespective of the value of H.

We thus see that even though the BS is not actively involved in a hand-off, the degradation in the voice packet dropping probability of the PRMA protocol due to user mobility is marginal. Hence, the primary reason for voice packets getting dropped in PRMA is contention.

Simulations based on the Markov model of the terminal were also carried out. At the start of the simulations, all the terminals were in the silent mode. The duration of the simulation run was chosen to be 10^6 slots (10^5 frames). Figure 4.8 compares the results obtained using analysis and simulation for $H = 0.05$ and $p = 0.1$. Analytical results turn out to be pessimistic when compared with the simulation results. For $p = 0.1$, the analytical results were within 10% of simulation results. The analytical results deviate from the simulation results when there are a large number of terminals ($t > 30$) due to the approximation made in Section 4.6.7. In Fig. 4.8, the dashed line (- -) corresponds to $\overline{\pi}$ *vs.* t when the term corresponding to $R = n$ in (4.44) (which is $\sum_{C=1}^{t-n} \pi(\nu(C,n))\theta_{C|R=n}(C)\theta_R(n)$) is not taken into account.[13] We see that it accounts for a major portion of $\overline{\pi}$ when the number of terminals is large. The discrepancy between the simulation results and analytical results for a large

[13]For $R = n$, $\pi(\nu(C,n)) = 1$ since $\nu(C,n) = 0$.

number of terminals was also observed for other values of p.

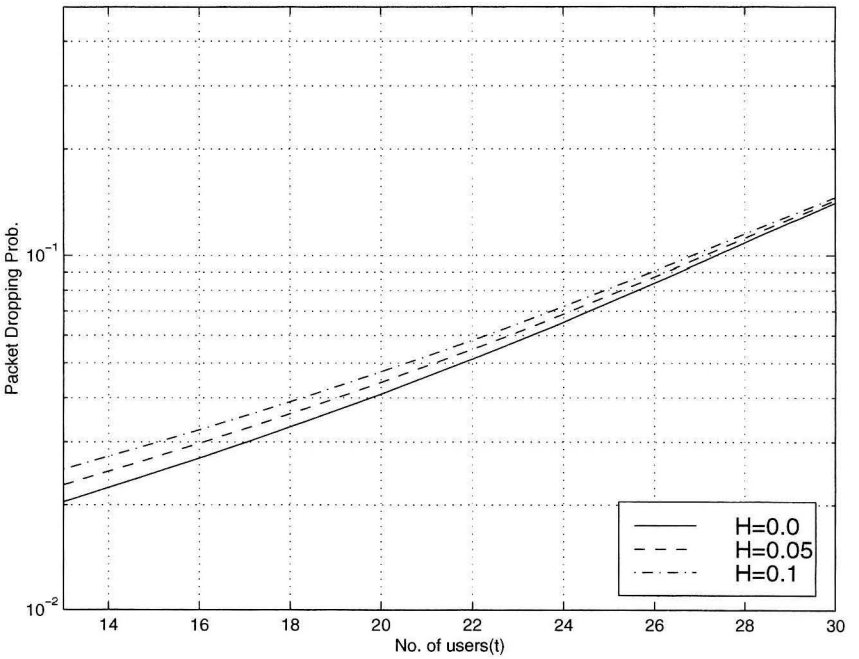

Figure 4.6: Voice packet dropping probability (analysis) for $p = 0.1$

4.7 Impact of Packet Errors

The previous section dealt with the impact of mobility on PRMA. We now analyze the impact of packet errors on PRMA. Different approaches for incorporating packet errors in a system model dealing with PRMA can be found in the literature. Wong [68] and Qi and Wyrwas [71] studied the effect of packet header errors on PRMA, but ignored packet data errors. Frullone *et al.* [56] and Grillo *et al.* [79] instead assumed ideal packet header error correction in their simulations and treated voice packets with packet data errors in them as *interfered* packets. They observed that around 5% of the transmitted voice packets were interfered packets for typical system parameter values. What effect such packets have on voice quality has not been elaborated upon in the literature.

In our analysis, we assume that voice packets with packet data errors in them are retransmitted, so long as their delay constraint is not violated. This scenario is feasible in PRMA since the feedback from the BS is assumed to be available right at the beginning of

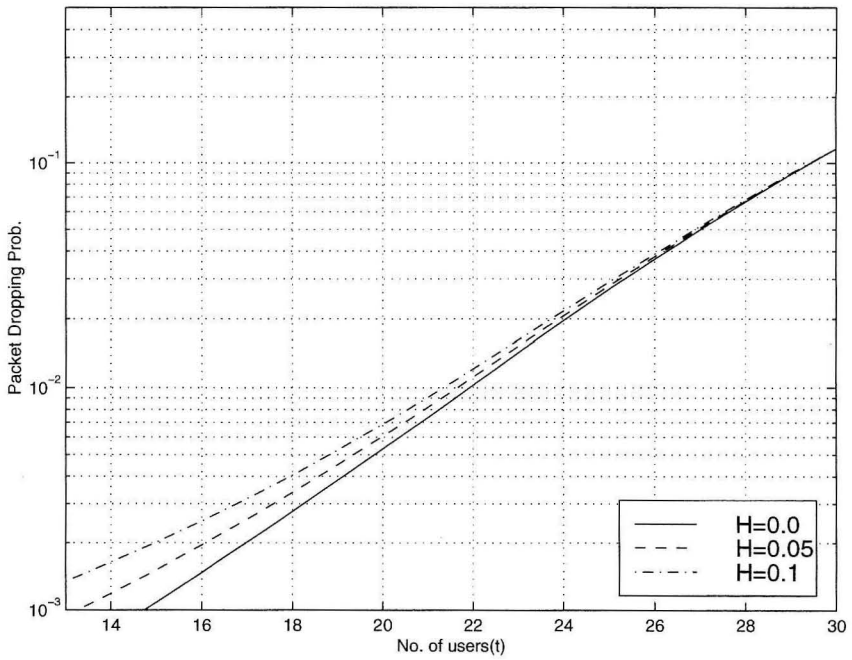

Figure 4.7: Voice packet dropping probability (analysis) for $p = 0.4$

the next slot. This assumption therefore ensures that a packet gets dropped only if it cannot meet the voice packet delay constraint. In addition, it avoids having the performance of PRMA for voice users characterized by two performance measures: the packet dropping probability as well as the packet interference probability.

Another issue is the model used for deciding when packet errors have occurred. For modeling interference dependent packet error rates, Frullone *et al.* [56] used an 8 dB threshold for received SIR (Γ) in their simulations.[14] They assumed that if $\Gamma \geq 8$ dB the transmitted packet is perfectly received. Otherwise, they labeled the data in the corrupted packet as *interfered.* It is difficult to correlate the actual performance of the protocol for a given error correction code and the performance obtained using their simplistic model. Using an EPA based analysis, we show that it is possible to model the packet error rates as a function of the channel fading statistics and interference, and as a function of the modulation and error correction coding scheme used for packet transmission.

[14] A similar threshold model has also been employed by Wang and Poon [80] in their simulations. Complete physical layer simulations were carried out by Hanzo, Cheung, Steele, and Webb [81]. They assumed a 32 kbps ADPCM speech codec, Reed-Solomon forward error correction codec, and a diversity-assisted 16-level star quadrature amplitude modulation.

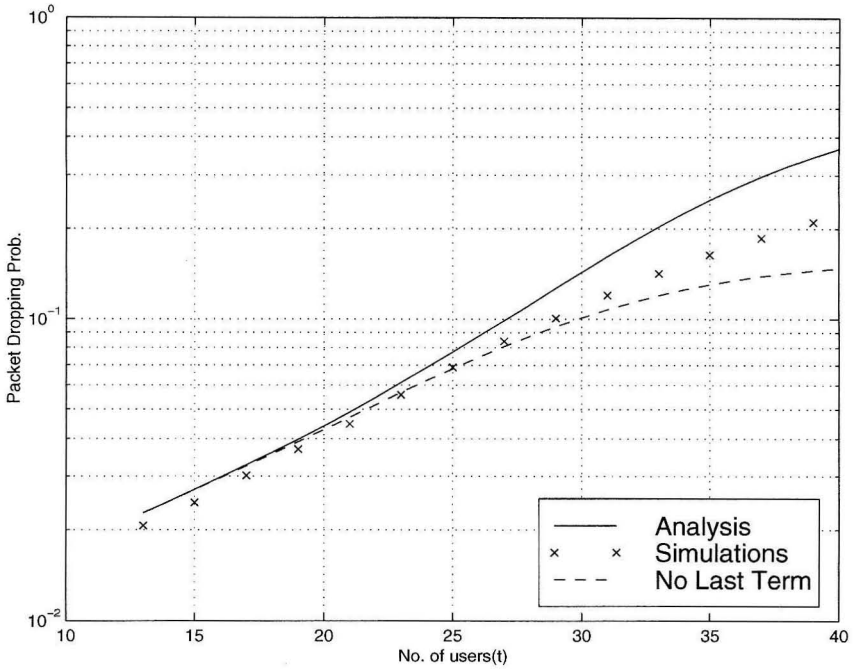

Figure 4.8: Voice packet dropping probability: analysis *vs.* simulation for $p = 0.1$, $H = 0.05$

4.7.1 Terminal Model with Packet Errors

The Markov model for a terminal with packet errors is shown in Fig. 4.9. When the terminal transmits a packet in mode $\mathcal{R}_0$, the voice packet header part can get corrupted with probability e_h, and the voice packet data part can get corrupted with probability e_d.[15] A header error causes the terminal to lose its reservation. It therefore rejoins the contention mode. In case of an error in the data part, the terminal just goes back to mode $\mathcal{R}_{n-1}$, with the transmitted packet being rejected by the receiving BS – there is no loss of reservation. The two branches shown in Fig. 4.9 from $\mathcal{R}_0$ to $\mathcal{R}_{n-1}$ correspond to packet transmissions with and without packet data error. The feedback from the BS is taken to be free of errors.

4.7.2 Packet Error Rate Model

We consider two models for evaluating the packet error rates faced by PRMA in a cellular system.

[15] In [71] and [68], e_h was assumed to be fixed, and e_d was taken to be 0.

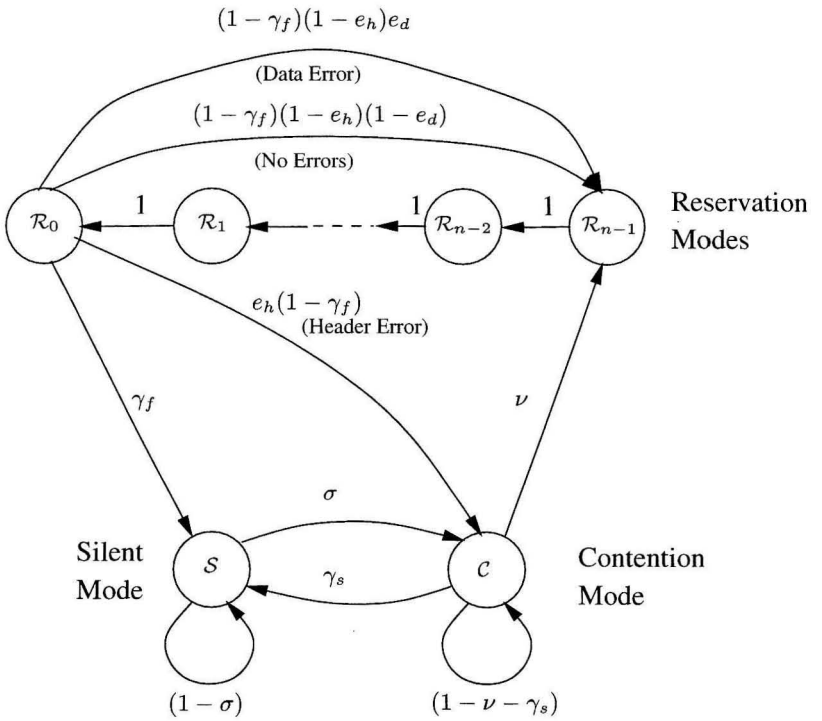

Figure 4.9: Terminal model with packet errors

Fixed error rates The packet header and data errors are fixed a priori. This model is appropriate in a noise-limited cellular system.

Interference dependent error rates The fixed error rate model is not appropriate in an interference-limited cellular environment where the average packet error rate is a function of the SIR Γ seen by the receiving BS. It also depends on the error correction capability of the coding used for the packet header and data. What makes the problem interesting is that the interference in turn depends on the error rates seen by the users. For example, the higher the data error rate, the more the retransmissions of packets, and hence the more the interference. If packet header rates are high, more users fall into the contention mode. Hence, the probability of multiple users transmitting in the same slot, and therefore generating interference, increases.

The analysis for the case with packet errors follows along the same lines as the one with mobility in Section 4.6.2–Section 4.6.5. The main differences are in the individual

expressions.

4.7.3 Packet Dropping Probability Evaluation

As in Section 4.6.2, we first evaluate $\pi(\nu)$ given the probability ν that a contending terminal will successfully transmit a packet. Let the first packet of a talk spurt be generated at time 0. The i^{th} packet in a talk spurt is generated at time $n(i-1)$ and must be transmitted by time $d+n(i-1)$ to avoid violating the delay constraint. Similarly, the last packet in a talk spurt of length q must be transmitted by time $n(q-1)+d$. Let $a_{\mathcal{A}}(i,j)$ be the probability that a terminal in mode $\mathcal{C}$ at time 0 ends up in mode $\mathcal{A}$ by time j and has transmitted exactly i packets successfully. After time nq, if the terminal does not have a slot reservation, *i.e.*, it is in mode $\mathcal{C}$, it returns to mode $\mathcal{S}$ and discards the remaining packets of the talk spurt. Therefore, for a terminal in mode $\mathcal{C}$ the average number of packets dropped in a given talk spurt of length q is $\sum_{i=0}^{q-1}(q-i)a_{\mathcal{C}}(i,nq)$. If the terminal is instead in any of the reserved modes $\mathcal{R}_w$ $(0 \leq w \leq n-1)$ at time nq, it will continue to transmit its remaining packets provided that it does not lose its slot reservation, and as long as the delay constraint is not violated.

As in Section 4.6.2, let $b_w^{(q-i)}(h)$ denote the probability that the terminal manages to transmit h more packets, given that it has transmitted i $(i<q)$ packets and is in mode $\mathcal{R}_w$ by time nq. As mentioned above, the last packet has to be transmitted before time $n(q-1)+d$. Let $d=\tau n$. The terminal can only transmit in the time interval $nq+1,\ldots,n(q-1)+d$ of length $d-n=(\tau-1)n$. In a time interval of length $(\tau-1)n$ at most $(\tau-1)$ voice packets can be transmitted. Therefore, a terminal can transmit at most $\min(q-i,\tau-1)$ packets after time nq. Hence, the average number of packets dropped by a terminal in mode $\mathcal{R}_w$ at time nq is $\sum_{i=0}^{q-1}\sum_{h=0}^{\min(q-i,\tau-1)}(q-i-h)a_{\mathcal{R}_w}(i,nq)b_w^{(q-i)}(h)$.

We first explain how to evaluate $b_w^{(q-i)}(h)$ for $d=2n$. For this case, at most one more packet can be transmitted after time nq. From Fig. 4.9, it is clear that a terminal in mode $\mathcal{R}_w$ can successfully transmit one packet with probability $(1-e_h)(1-e_d)$. Therefore,

$$b_w^{(q-i)}(0) \;=\; 1-(1-e_h)(1-e_d),$$

$$b_w^{(q-i)}(1) = (1-e_h)(1-e_d). \tag{4.26}$$

In general, we have

$$b_w^{(q-i)}(h) = \sum_{q=h}^{\min(q-i,\tau-1)-1} \binom{q}{h} (1-e_d)^h e_d^{q-h} e_h (1-e_h)^h + \binom{\min(q-i,\tau-1)}{h} (1-e_d)^h e_d^{q-h} (1-e_h)^{\min(q-i,\tau-1)}. \tag{4.27}$$

We use $d = 2n$ for the following derivations. The average number of packets in a talk spurt is $1/\gamma_f$. The probability that the talk spurt is q packets long is $\gamma_f(1-\gamma_f)^{q-1}$. From Fig. 4.9, it is also clear that $a_{\mathcal{R}_w}(i,j) = a_{\mathcal{R}_{n-1}}(i, j-n+w+1)$, for $0 \le w \le n-1$. We therefore get

$$\begin{aligned}
\pi(\nu) = & \\
& \gamma_f \sum_{q=1}^{\infty}\sum_{i=0}^{q-1}(q-i)a_{\mathcal{C}}(i,nq)\gamma_f(1-\gamma_f)^{q-1} \\
& + \gamma_f \sum_{q=1}^{\infty}\sum_{i=0}^{q-1}\sum_{w=0}^{n-1} a_{\mathcal{R}_w}(i,nq)\gamma_f(1-\gamma_f)^{q-1} \left(\sum_{h=0}^{\min(\tau-1,q-i)} (q-i-h) b_w^{(q-i)}(h) \right), \\
& = \gamma_f^2 \sum_{q=1}^{\infty}\sum_{i=0}^{q-1}(q-i) \left[\sum_{\mathcal{A}\in\Omega} a_{\mathcal{A}}(i,nq) \right] (1-\gamma_f)^{q-1} \\
& - \gamma_f^2(1-e_h)(1-e_d) \sum_{q=1}^{\infty}\sum_{i=0}^{q-1} \left[\sum_{w=1}^{n} a_{\mathcal{R}_{n-1}}(i, n(q-1)+w) \right] (1-\gamma_f)^{q-1},
\end{aligned} \tag{4.28}$$

where $\Omega = \{\mathcal{C}, \mathcal{R}_0, \ldots, \mathcal{R}_{n-1}\}$ is the set of modes a terminal can be in during a talk spurt.

4.7.4 Generating Functions and Simplification

To evaluate $\pi(\nu)$ we compute the two generating functions $g_\Omega(x,l) = \sum_i \sum_j a_\Omega(i,j)x^i l^j$ and $g_{\mathcal{R}_{n-1}}(x,l) = \sum_i \sum_j a_{\mathcal{R}_{n-1}}(i,j)x^i l^j$, where $a_\Omega(i,j) = \sum_{\mathcal{A}\in\Omega} a_{\mathcal{A}}(i,j)$. Since a terminal does not enter the silent mode $\mathcal{S}$ during a talk spurt, it suffices to consider the truncated mode diagram in Fig. 4.10 for calculating $\pi(\nu)$. The branch mode labels are modified with

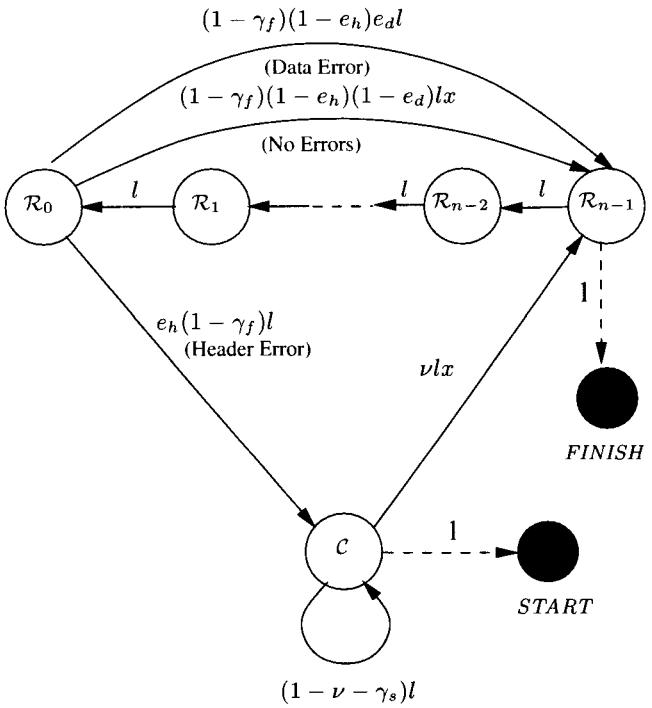

Figure 4.10: Terminal mode diagram, with packet errors, during a talk spurt

indeterminates l and x as described in Section 4.6.3.

Figure 4.10 satisfies the property, previously observed in Fig. 4.4 in Section 4.6.3, that the two loops with a factor x in their loop gains are non-touching. Hence, the generating functions are once again first order rational functions in x. They can be written as

$$g_\Omega(x,l) = \frac{\alpha_\Omega(l)+\beta_\Omega(l)x}{\gamma(l)+\delta(l)x}, \tag{4.29}$$

$$g_{\mathcal{R}_{n-1}}(x,l) = \frac{\alpha_{\mathcal{R}_{n-1}}(l)+\beta_{\mathcal{R}_{n-1}}(l)x}{\gamma(l)+\delta(l)x}. \tag{4.30}$$

The expressions for $\alpha_\Omega(l)$, $\beta_\Omega(l)$, $\alpha_{\mathcal{R}_{n-1}}(l)$, $\beta_{\mathcal{R}_{n-1}}(l)$, $\gamma(l)$, and $\delta(l)$ are derived in Appendix 4.B.1. Therefore, the recursion used in Section 4.6.4 directly applies in this case also. From Fig. 4.10, it can also be seen that the minimum length of a closed loop, whose loop gain has x as a factor, is n. Therefore, the simplifications carried out in Section 4.6.5 are also applicable. They are listed below briefly since the formulae are rather different.

Since the minimum length of a closed loop that has x as a factor in its loop gain is

n, $\delta(l)$ can be written as $l^n\delta_s(l)$, where $\delta_s(l)$ is itself a polynomial in l. We recursively evaluate the function $f_\Omega^{(i)}(l)$ defined by

$$\begin{aligned} f_\Omega^{(0)}(l) &= \tfrac{\alpha_\Omega(l)}{\gamma(l)} &&= a_\Omega^{(0)}(l), \\ f_\Omega^{(1)}(l) &= \tfrac{\beta_\Omega(l)\gamma(l)-\alpha_\Omega(l)\delta(l)}{\gamma(l)^2} &&= a_\Omega^{(1)}(l), \\ f_\Omega^{(i)}(l) &= -\tfrac{\delta_s(l)}{\gamma(l)} f_\Omega^{(i-1)}(l) &&= l^{-n(i-1)} a_\Omega^{(i)}(l),\ i \geq 2. \end{aligned} \tag{4.31}$$

The formula for $\pi(\nu)$, derived in (4.28), can then be simplified to

$$\begin{aligned} \pi(\nu) &= \gamma_f^2 \sum_{i=1}^{\infty}\sum_{t'=1}^{\infty} t' f_\Omega(i, n(t'+1))(1-\gamma_f)^{i+t'-1} \\ &\quad - \gamma_f^2(1-e_h)(1-e_d)\sum_{i=1}^{\infty}\sum_{t'=1}^{\infty}\left[\sum_{w=1}^{n} f_{\mathcal{R}_{n-1}}(i, nt'+w)\right](1-\gamma_f)^{i+t'-1} \\ &\quad + \gamma_f^2 \sum_{q=1}^{\infty} q f_\Omega(0, nq)(1-\gamma_f)^{q-1}, \end{aligned} \tag{4.32}$$

where $f_\Omega(i,j)$ ($f_{\mathcal{R}_{n-1}}(i,j)$) is the coefficient of l^j in the polynomial $f_\Omega^{(i)}(l)$ ($f_{\mathcal{R}_{n-1}}^{(i)}(l)$).

We now evaluate the steady state system probability distribution using EPA [70]. The relationship between the packet error rates, e_h and e_d, and the co-channel interference, for an interference-limited cellular system, is also arrived at in the following sections.

4.7.5 Equilibrium Point Analysis Equations

Let s, c, and r_i ($0 \leq i \leq n-1$) denote the equilibrium values of S, C, and R_i ($0 \leq i \leq n-1$), respectively. The EPA equations are given by

At mode $\mathcal{S}$

$$s\sigma = r_0\gamma_f + c\gamma_s, \tag{4.33}$$

At mode $\mathcal{C}$

$$c(\nu+\gamma_s) = \sigma s + r_0(1-\gamma_f)e_h, \tag{4.34}$$

At mode $\mathcal{R}_{n-1}$

$$r_{n-1} = c\nu + r_0(1-\gamma_f)(1-e_h), \tag{4.35}$$

At mode $\mathcal{R}_i$ $(0 \leq i \leq n-2)$

$$r_i = r_{i+1}. \tag{4.36}$$

The transition probability ν from $\mathcal{C}$ to $\mathcal{R}_{n-1}$ is given by

$$\nu = u(c)p(1-r_0)(1-e_h)(1-e_d)(1-\gamma_s), \tag{4.37}$$

where

$$u(c) = \begin{cases} (1-p)^{(c-1)}, & c \geq 1 \\ 1, & c < 1 \end{cases}.$$

We also have an additional equation because the number of terminals t per cell is a constant:

$$t = \sum_{i=0}^{n-1} r_i + s + c. \tag{4.38}$$

Solving (4.33)–(4.38) yields the equilibrium values of the system variables. We shall use these values in the following sections to arrive at the system probability distribution.

4.7.6 Interference Dependent Packet Error Rate Evaluation

The average packet error rate is a function of the SIR Γ seen by the BS's receiver. It also depends on the error correction capability of the coding used for the packet header and data.[16]

Fast fading channel The fading is constant over the duration of a symbol. The fading from one symbol to another is i. i. d. An upper bound on the probability of error e_h of a linear block code for a Rayleigh fading channel and maximum likelihood soft

[16]The formulae given in this section are union bounds that assume maximum likelihood decoding. Packet error expressions for hard decision decoding may also be used. For example, the probability of decoding error e_h for a t error correcting code is given by [42, Chp. 10]

$$e_h = 1 - \sum_{h=0}^{t} \binom{L}{h} \mathrm{BER}(\Gamma)^h (1 - \mathrm{BER}(\Gamma))^{L-h}, \tag{4.39}$$

where L is the block code length and $\mathrm{BER}(\Gamma)$ is the probability of bit error for SIR $= \Gamma$. Here the fading has been assumed to be i. i. d. from one bit to another.

decision decoding is given by [82]:

$$e_h \leq \frac{1}{2} \sum_{h=d_{\min}}^{L} W(h) \left(\frac{1}{1+\Gamma} \right)^h, \tag{4.40}$$

where $W(.)$ is the weight enumerator [44, Chp. 4] of the code with codewords of length L and minimum distance $d_{\min}$.

Slow fading channel The fading is constant over the duration of a packet. The fading from one packet transmission to another is i. i. d. For this case, the upper bound for maximum likelihood soft decision decoding becomes

$$e_h \leq \frac{1}{2} \sum_{h=d_{\min}}^{L} W(h) \left(\frac{1}{1+h\Gamma} \right). \tag{4.41}$$

The results in Section 4.7.8 are for the fast fading channel.

Evaluating SIR

A hexagonal cellular layout with a given reuse cluster size is assumed. Interference is taken to originate only from the first tier of co-channel interferers. The power fall-off with distance d from the transmitter is modeled as $P_r = P_t/d^g$, where g is the path loss exponent, P_r is the received power, and P_t is the transmitted power. Since the system is assumed to be interference-limited, the circum-radius of the hexagonal cells R is normalized to 1, and so is P_t.

The equilibrium values of the number of contending users and reserved slots in adjacent cells determine the average interference $\overline{\mathcal{I}}$ as follows. Let the distance of an interfering user (in the neighboring cell) from the BS of the reference cell be d_{av}.[17] Recall that this average distance approximation was also used previously in Section 3.4. If the current slot is reserved, the average interference from a neighboring cell is[18] $P_r(1-\gamma_f)$. Instead, if the

[17] d_{av} is the distance at which all the interfering users are assumed to be located from the receiving BS. d_{av} is given by $d_{\text{av}} = ((1/t)\sum_{i=1}^{t} 1/d_i^g)^{-1/g}$, where d_i is the distance of the i^{th} interfering user from the receiving BS.

[18] The factor $(1-\gamma_f)$ is there because a terminal in a reserved slot may end its talk spurt and return to mode $\mathcal{S}$ without transmitting a packet.

current slot is unreserved, the average interference is $P_r cp(1-\gamma_s)$, where c is the average number of contending users, who transmit in an unreserved slot with probability p, in a cell. The probability that a slot is reserved is r_0. Thus, $\overline{\mathcal{I}}$ is given by

$$\overline{\mathcal{I}} = 6\frac{P_t}{d_{\text{av}}^g}\left[(1-\gamma_f)r_0 + cp(1-\gamma_s)(1-r_0)\right], \tag{4.42}$$

where the factor 6 arises because there are 6 first tier interfering cell neighbors in a hexagonal cellular layout. As assumed previously in Chapter 3, the co-channel interference emanating from cells farther than the first tier cells is negligible. Hence the average SIR, assuming that the interference is Gaussian, is given by

$$\overline{\Gamma} = \frac{\frac{P_t}{X^g}}{6\frac{P_t}{d_{\text{av}}^g}\left[(1-\gamma_f)r_0 + cp(1-\gamma_s)(1-r_0)\right]}, \tag{4.43}$$

where X is the distance of the transmitting user from its base station.

Using the above equation in (4.40) or (4.41), as the case may be, and jointly solving for e_h with the other EPA equations (4.33)–(4.38) yields the equilibrium values of the variables s,c, $r_0, \ldots, r_{n-1}$, and e_h. Since the code used for a packet's header is stronger than that for a packet's data, we fix e_d to be λe_h, where $\lambda \geq 1$. From the above discussion, it is clear that the packet error rates and the equilibrium values of system variables are interrelated in an interference-limited environment, and cannot be determined or fixed independently of each other.

4.7.7 System Probability Distribution Estimate

The system state vector is given by $(S, C, R_0, \ldots, R_{n-1})$, where S is the number of terminals in a cell in silent mode $\mathcal{S}$, C is the number of terminals in contention mode $\mathcal{C}$, and R_i is number of terminals in reservation mode $\mathcal{R}_i$ $(0 \leq i \leq n-1)$. $\pi(\nu)$ is then evaluated for each of the states (each state has a different value of ν) and is subsequently unconditioned according to the probability distribution estimate. The average packet dropping probability

$\overline{\pi}$ is given by

$$\overline{\pi} = \sum_{R=0}^{n} \sum_{C=1}^{t-R} \pi(\nu(C,R))\theta_{C|R}(C)\theta_R(R), \tag{4.44}$$

where

$$\begin{aligned} \nu(C,R) &= (1-\theta_{R_0|R}(1))p(1-p)^C(1-e_h)(1-e_d), \\ R &= \sum_{i=0}^{n-1} R_i. \end{aligned}$$

The notation $\theta_{(.)}(.)$ is the same as that in Section 4.6.7.

Distribution of R_0 Given R, $\theta_{R_0|R}(.)$

We have $r_0 = r_1 = \ldots = r_{n-1}$ from (4.36). In other words R_0, $R_1, \ldots, R_{n-1}$ are i. i. d. binary random variables. Therefore, the probability that the current slot is reserved given that R slots, in a frame of n, are reserved is [70]

$$\theta_{R_0|R}(1) = \frac{R}{n}. \tag{4.45}$$

Distribution of R, $\theta_R(.)$

The average number of slots occupied in a frame is nr_0. Under the valid approximation that the slots are occupied independently of each other, $\theta_R(R)$ is a binomial probability distribution of the form

$$\theta_R(R) = \begin{pmatrix} n \\ R \end{pmatrix} r_0^R(1-r_0)^{n-R}. \tag{4.46}$$

We had used this approximation previously in Section 4.6.6.

Distribution of C Given R, $\theta_{C|R}(.)$

We use an M/M/1 geometric distribution approximation [70], similar to the one made in Section 4.6.6, for evaluating $\theta_{C|R}(C)$. We have

$$\theta_{C|R}(C) = \begin{cases} p_0(1-p_0)^C, & C < t - R \\ (1-p_0)^C, & C = t - R \\ 0, & \text{otherwise} \end{cases}, \tag{4.47}$$

where $p_0 = \frac{1}{(c+1)}$.

4.7.8 Results

We assume the following values [70, 71] for the system parameters shown in Fig. 4.9: average talk spurt duration $t_a = 1.00$ s, average silence duration $t_q = 1.35$ s, frame duration $t_f = 0.016$ s, number of slots in a frame $n = 20$, and delay constraint $d = 40$ slots. Therefore, the time slot duration t_s is $t_f/n = 0.0008$ s. Since the terminal model is Markov, the probability σ that a silent terminal enters the talk spurt mode is $1 - \exp(-t_s/t_q) = 0.000592$. The probability γ_s that a terminal in mode C returns to the silent mode is $1 - \exp(-t_s/t_a) = 0.000799$. The probability γ_f that a talk spurt has ended for a terminal in a reservation mode is then $1 - (1-\gamma_s)^n = 0.0158$.

We first present the results for the fixed error rate model. They are followed by the results for the SIR dependent error rate model. In both the cases, only voice users are considered in the system. These results obtained using analysis are also compared with those obtained using simulations.

Fixed Error Rates

Figure 4.11 plots the voice packet dropping probability as a function of the number of users per cell for $e_h = 0.01, 0.03$, and 0.05, when packet data error is ignored[19] ($e_d = 0$). The probability p that a contending user transmits in an unreserved slot is taken to be 0.3 in this

[19]This is the case studied in [71].

case. Notice the good match between the analysis and simulation results.

Next, we investigate the effect of packet data error on the voice packet dropping probability. Figure 4.12 plots the voice packet dropping probability for $p = 0.3$ assuming ideal header protection ($e_h = 0$) and a fixed packet data error probability. Figure 4.13 plots the voice packet dropping probability as a function of the number of users per cell for fixed e_h and e_d, with $e_d = e_h$ and $p = 0.3$. Note the close match between our analytical and simulation results in Figures 4.12 and 4.13. In view of this close match, simulation results are not plotted in the following figure. Figure 4.14 plots packet dropping probability *vs.* number of users for $e_d = 2e_h$.

Interference Dependent Error Rates

We now look at the case where the packet error rates are interference dependent. For this purpose, we use simple 1 error and 3 error correcting codes as explained below. Figure 4.15 plots the voice packet dropping probability *vs.* the number of users per cell when e_h is a function of the co-channel interference from the neighboring cells, as given by (4.40). The single error correcting $(63, 57, 3)$ Hamming code [44] was used for this plot with a path loss coefficient $g = 3.5$, $e_d = e_h$, and reuse cluster size 3. This is done for $p = 0.1, 0.2,$ and 0.3. Recall that the results are for the case where all users are at the same distance ($X = R/\sqrt{2}$) from their respective BSs.

The analytically obtained results for the SIR dependent case are compared with those obtained using simulations in Fig. 4.16. The comparison is made when $p = 0.3$, $g = 3.5$, $e_d = e_h$, and the reuse cluster size is 3. While the analysis tracks the simulation results well, the error is greater than that for the fixed error rate model considered earlier. This is because interference is a dynamic time varying quantity, while the analysis uses only its mean value in (4.43). Moreover, the system becomes unstable after $t > 36$, and is marked by a jump in the voice packet dropping probability. This is not captured by the analysis.

With a 3 error correcting code like the $(63, 45, 7)$ Golay code, the effect of co-channel interference is negligible (for $g \geq 3.0$), *i.e.*, the performance is the same as that for the no packet errors case. The voice packet dropping probability is unacceptably high for the uncoded packet header case, and is not shown here.

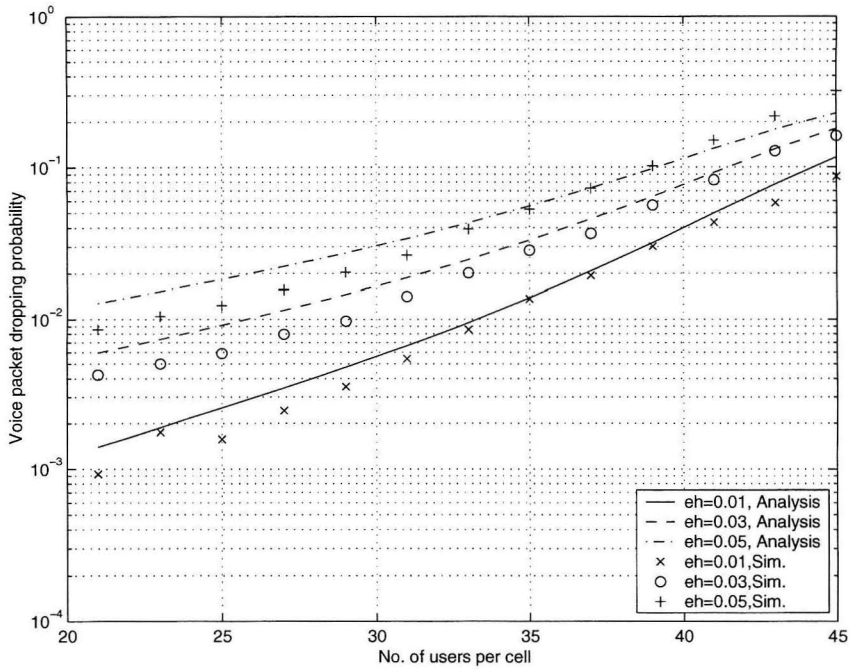

Figure 4.11: Voice packet dropping probability for $p = 0.3$, $e_d = 0$, and fixed e_h

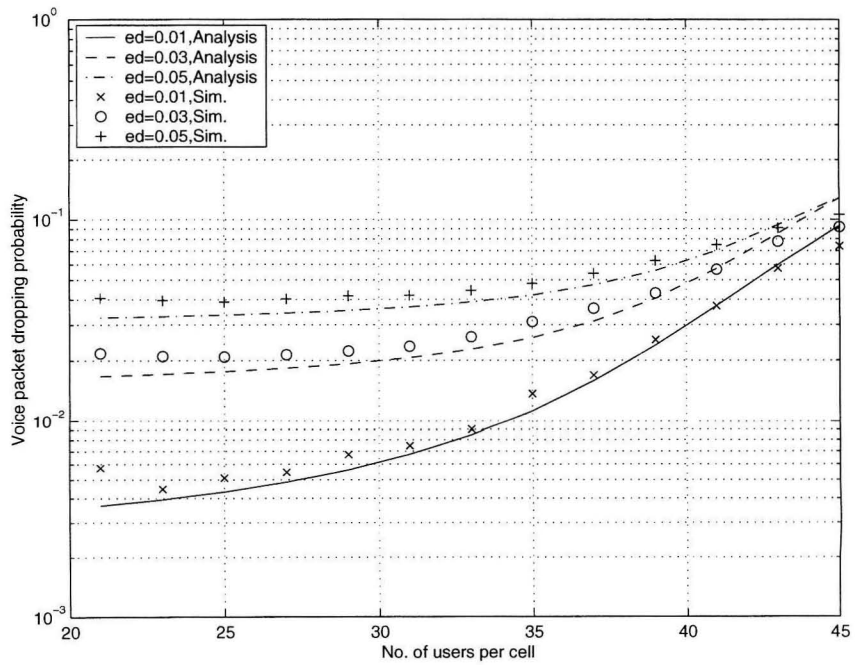

Figure 4.12: Voice Packet dropping probability for $p = 0.3$, $e_h = 0$, and fixed e_d

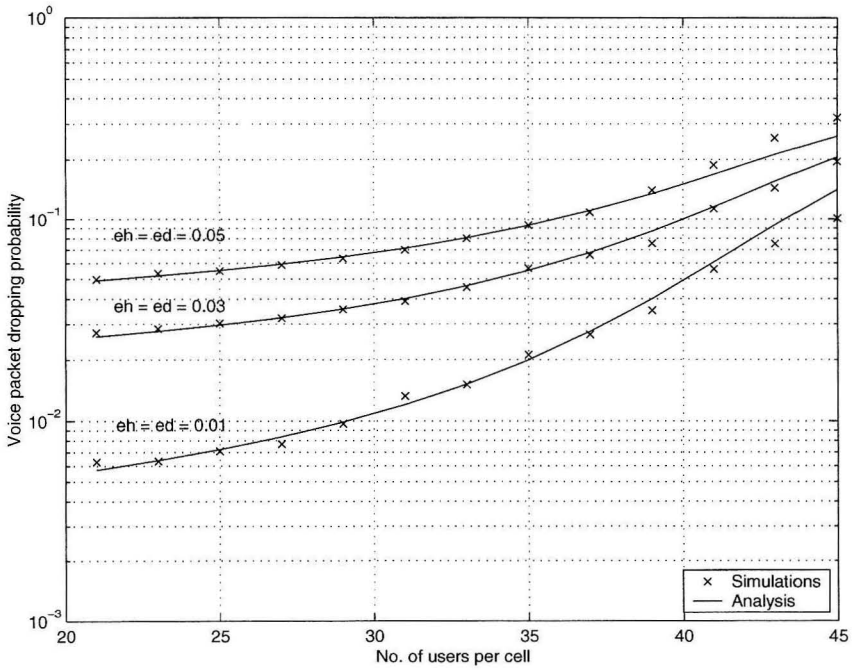

Figure 4.13: Voice Packet dropping probability for $p = 0.3$ and fixed $e_d = e_h$

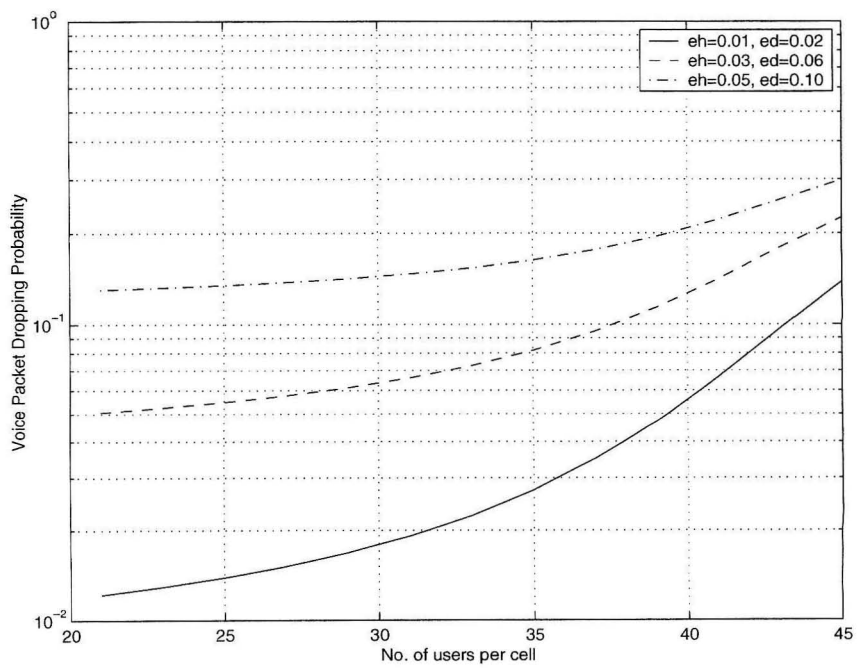

Figure 4.14: Voice packet dropping probability for $p = 0.3$ and fixed $e_d = 2e_h$

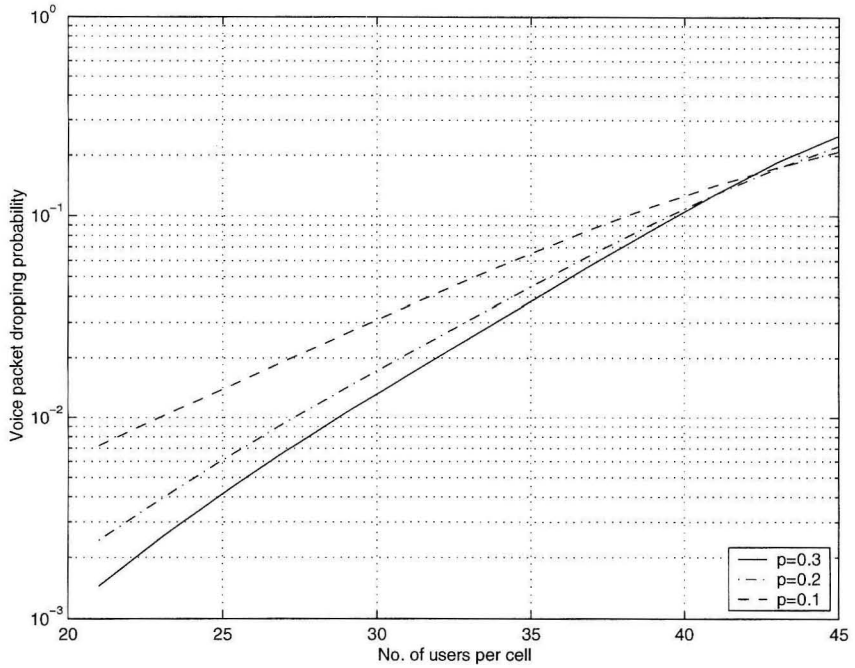

Figure 4.15: Voice packet dropping probability: single error correcting code with $g = 3.5$, reuse cluster size 3, and $e_d = e_h$

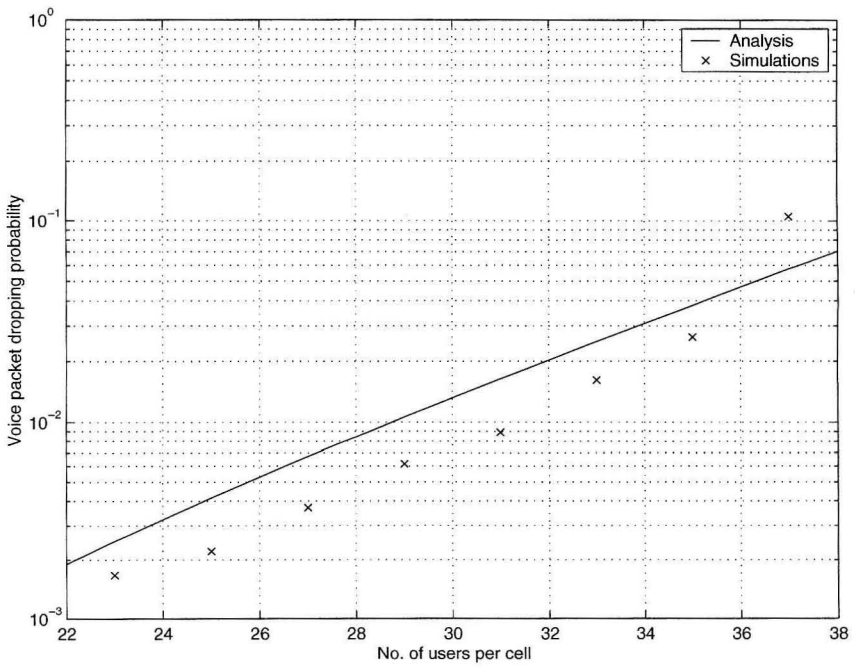

Figure 4.16: Voice packet dropping probability: analysis *vs.* simulations for interference dependent packet error rates

On Equilibrium Points

EPA is an approximate analysis technique that solves a set of mean flow equations, (4.33)–(4.38) in Section 4.6.7 and (4.19)–(4.20) in Section 4.7.5, about the multi-dimensional Markov process it is applied to. In Sections 4.6.7 and 4.7.5, we saw that the equilibrium values of the state variables are obtained by solving a non-linear equation. Multiple equilibrium solutions can exist[20] for these non-linear equations. Moreover, in the analysis of the packet error model, we found cases where no solutions existed for the EPA equations.

The presence of multiple equilibrium states can cause the results obtained using an EPA based analysis to deviate from the actual numbers since the system probability distribution estimate uses the equilibrium values for state variables from only one of the equilibrium states. This deviation is significant when the system spends a significant portion of its time in or near both the equilibrium states. Unfortunately, it is not possible to determine using EPA what fraction of time the system spends in each of its equilibrium states [77, Chp. 2].

4.8 Conclusions

We have looked at two issues that influence the performance of packet reservation multiple access for voice terminals, namely, terminal mobility and packet errors. In summary, a voice terminal that has a time slot reserved for transmitting its talk spurt in progress loses its reservation when it moves to another cell. Having to recontend again induces additional delays in transmitting voice packets. Likewise, an error in the header of a transmitted packet can cause the BS to prematurely declare the reserved slot, in which the voice terminal transmitted its packet, as unreserved. However, errors in the data part of the packet do not lead to any such loss of reservation. The packet in question just needs to be retransmitted, subject to it being able to meet the voice delay constraint.

In both the above cases, the voice packet dropping probability is first evaluated as a function of the successful packet transmission probability of a contending terminal. This is

[20]For example, in [70] it was shown for the basic no mobility and no packet error case that the number of equilibrium states is always odd. Multiple equilibrium states were shown to exist when the number of users per cell was large.

achieved by means of a signal flow graph technique based on an appropriate Markov model for a voice terminal. An approximate system probability distribution, arrived at using equilibrium point analysis, is then employed to compute the average voice packet dropping probability. The cases corresponding to packet error rates fixed a priori and packet error rates dependent on co-channel interference were analyzed. In the latter case, the interdependence between interference power levels and packet error rates was easily captured by our approximate analysis technique.

The good match between analysis and simulation results validates the approximate analysis. The margin for error could, possibly, be further reduced, and the problems associated with multiple equilibrium points circumvented, if the approximate steady state distribution based on Markov analysis [71] is employed instead of the EPA based probabilities that we use in this chapter.

The model we employ in this chapter implicitly assumes that all terminals face similar path loss and statistical fading models. In more realistic scenarios, the terminals are randomly distributed over the entire cell area. In such a case, terminals close to the base station will face packet error probabilities significantly lower than those faced by terminals closer to the cell boundary. Analysis of this general case is an open problem.

Our results show that while terminal mobility results in an increase in the voice packet dropping probability, the effect is marginal. Contention is the main reason for dropping most of the voice packets. However, this cannot be said when packet errors are present – packet errors do have a significant impact on the system performance.

Appendix

4.A Impact of Terminal Mobility

4.A.1 $b_w^{(q-i)}(h)$ Derivation

The last packet in a q packet long talk spurt has to be transmitted before time $n(q-1)+d$ or it gets dropped. Let $d = \tau n$. The terminal can only transmit in the time interval $nq+1, \ldots, n(q-1)+d$ (of length $d-n = (\tau-1)n$) at most $(\tau-1)$ voice packets. To transmit a packet the terminal in mode $\mathcal{R}_w$ must reach $\mathcal{R}_{n-1}$ without moving into another cell. The probability that the terminal remains in the same cell for w slots, the time after which it can transmit its reserved slot, is $(1-m)^{w+1}$. Therefore, the terminal in mode $\mathcal{R}_w$ can transmit at least one packet with probability $(1-m)^{w+1}$. Similarly, it can transmit at least two packets with probability $(1-m)^{n+w+1}$, and so on.

The probability that no packet is transmitted is $1-(1-m)^{w+1}$. The probability that exactly one packet is transmitted is $(1-m)^{w+1}-(1-m)^{n+w+1}$. In general, the probability that h packets $(1 \leq h < \min(q-i, \tau-1))$ are transmitted is $(1-m)^{w+1+n(h-1)}(1-(1-m)^n)$. The probability that the maximum possible number of packets are transmitted $(\min(q-i, \tau-1))$ is therefore $1-\sum_{h=0}^{\min(q-i,\tau-1)-1} b_w^{(q-i)}(h)$. Hence the formula given in (4.2).

4.A.2 Gain Formula

The various forward paths for different modes to which *FINISH* is appended, their respective forward path gains ($F_q(x,l)$) and co-factors ($\Delta_q(x,l)$) are given in Table 4.1.[21]

The graph determinant $\Delta(x,l)$ is given by

$$\Delta(x,l) = \left(\begin{array}{l} (1-(1-\nu-m)l - l^y(1-(1-m)^y) - m(1-m)^y l^{y+1} \\ +(1-\nu-m)l^{y+1}(1-(1-m)^y) \end{array} \right)$$

[21] We are being slack in our notation in Table 4.1 when we write $\mathcal{A} \rightarrow \mathcal{B}$. The corresponding forward path gains and co-factors are always evaluated from *START* to *FINISH*. *START* is appended to $\mathcal{A}$ and *FINISH* is appended to $\mathcal{B}$.

From $\rightarrow$ To	$F_q(x,l)$	$\Delta_q(x,l)$
$\mathcal{C} \rightarrow \mathcal{R}_{n-k-1}$ $(0 \leq k \leq n-1)$	$x\nu l^{k+1}(1-m)^k$	$1-(1-(1-m)^y)l^y$
$\mathcal{C} \rightarrow \mathcal{B}_k$ $(1 \leq k \leq y)$	ml^{y-k+1} $x\nu ml^{y-k+2}$ $x\nu ml^{y-k+2}(1-m)l$ $\vdots$ $x\nu ml^{y-k+2}(1-m)^{n-1}l^{n-1}$	$1-(1-m)^n l^n x$ 1 1 $\vdots$ 1
$\mathcal{C} \rightarrow \mathcal{C}$	1	$1-(1-(1-m)^y)l^y-(1-m)^n l^n x$ $+(1-m)^n(1-(1-m)^y)l^{n+y}x$

Table 4.1: Terminal model with mobility: forward path gains and corresponding co-factors

$$+x\left(\begin{array}{l} -(1-m)^n l^n - \nu m(1-m)^y l^{y+2}\left(\sum_{i=0}^{n-1}(1-m)^i l^i\right) \\ +(1-\nu-m)(1-m)^n l^{n+1}+(1-m)^n(1-(1-m)^y)l^{y+n} \\ -(1-\nu-m)(1-(1-m)^y)(1-m)^n l^{n+y+1} \\ +m(1-m)^{y+n}l^{n+y+1} \end{array}\right). \tag{4.48}$$

Using the representation for $g_\Omega(x,l)$ and $g_{\mathcal{R}_{n-1}}(x,l)$ in (4.5) and (4.6), respectively, $\alpha_\Omega(l)$, $\beta_\Omega(l)$, $\alpha_{\mathcal{R}_{n-1}}(l)$, $\beta_{\mathcal{R}_{n-1}}(l)$, $\gamma(l)$, and $\delta(l)$, all polynomials in l, are given by

$$\alpha_\Omega(l) = ml\left(\sum_{i=0}^{n-1} l^i\right)+1-(1-(1-m)^y)\,l^y, \tag{4.49}$$

$$\begin{aligned}\beta_\Omega(l) &= \nu l\,(1-(1-(1-m)^y)l^y)\left(\sum_{i=0}^{n-1}(1-m)^i l^i\right)-ml(1-m)u^n l^n\left(\sum_{i=0}^{n-1} l^i\right) \\ &\quad +\nu ml^2\left(\sum_{i=0}^{n-1}(1-m)^i l^i\right)\left(\sum_{i=0}^{n-1} l^i\right)-(1-m)^n l^n \\ &\quad +(1-m)^n\,(1-(1-m)^y)\,l^{n+y},\end{aligned} \tag{4.50}$$

$$\alpha_{\mathcal{R}_{n-1}}(l) = 0, \tag{4.51}$$

$$\beta_{\mathcal{R}_{n-1}}(l) = \nu l\,(1-(1-(1-m)^y)l^y)\,, \tag{4.52}$$

$$\begin{aligned}\gamma(l) &= 1-(1-\nu-m)l-l^y\,(1-(1-m)^y)-m(1-m)^y l^{y+1} \\ &\quad +(1-\nu-m)\,(1-(1-m)^y)\,l^{y+1},\end{aligned} \tag{4.53}$$

$$\begin{aligned}\delta(l) &= -(1-m)^n l^n - \nu m(1-m)^y l^{y+2}\left(\sum_{i=0}^{n-1}(1-m)^i l^i\right) \\ &\quad + (1-\nu-m)(1-m)^n l^{n+1} + (1-m)^n l^{n+1} \\ &\quad + (1-m)^n\left(1-(1-m)^y\right) l^{y+n} + m(1-m)^{y+n} l^{n+y+1} \\ &\quad - (1-\nu-m)\left(1-(1-m)^y\right)(1-m)^n l^{n+y+1}. \end{aligned} \tag{4.54}$$

4.A.3 The No Terminal Mobility Case

We show that the expressions derived[22] for $\pi(\nu)$ are equivalent to those derived in [70] for $m = 0$. For $m = 0$, we have

$$\begin{aligned} g_\Omega(x,l) &= \frac{\nu x l \sum_{i=0}^{n-1} l^i + 1 - l^n x}{(1-(1-\nu)l)(1-l^n x)}, \\ g_{\mathcal{R}_{n-1}}(x,l) &= \frac{x\nu l}{(1-(1-\nu)l)(1-l^n x)}. \end{aligned}$$

Hence, $\left(-\frac{\delta_s(l)}{\gamma(l)}\right) = 1$. As a result, $f_\Omega^{(.)}(l)$ is given by

$$f_\Omega^{(i)}(l) = \begin{cases} \frac{1}{1-(1-\nu)l}, & i = 0 \\ \frac{\nu l \sum_{i=0}^{n-1} l^i}{1-(1-\nu)l}, & i \geq 1 \end{cases}. \tag{4.55}$$

It follows that

$$f_\Omega(i,j) = \begin{cases} (1-\nu)^j, & i = 0 \\ 1-(1-\nu)^j, & i \geq 1,\ 1 \leq j < n \\ (1-\nu)^{j-n}(1-(1-\nu)^n), & i \geq 1,\ j \geq n \end{cases}. \tag{4.56}$$

Similarly, $f_{\mathcal{R}_{n-1}}(i,j)$ is given by

$$f_{\mathcal{R}_{n-1}}(i,j) = \begin{cases} 0, & i = 0 \\ \nu(1-\nu)^{j-1}, & i \geq 1, j \geq 1 \end{cases}. \tag{4.57}$$

[22] To make the expressions derived in this chapter tally with those derived in [70], ν must be replaced with $(1-\nu)$.

Substituting the above expressions in (4.11) we get,

$$\begin{aligned}
\pi(\nu) &= \gamma_f \sum_{t'=1}^{\infty} t' f_\Omega(1, n(t'+1))(1-\gamma_f)^{t'} - \gamma_f \sum_{t'=1}^{\infty} \left[\sum_{w=1}^{n} f_{\mathcal{R}_{n-1}}(1, nt'+w)\right](1-\gamma_f)^{t'} \\
&\quad + \gamma_f^2 \sum_{q=1}^{\infty} q f_\Omega(0, nq)(1-\gamma_f)^{q-1}, \\
&= \gamma_f \sum_{t'=1}^{\infty} t'(1-\nu)^{nt'}(1-(1-\nu)^n)(1-\gamma_f)^{t'} \\
&\quad - \gamma_f \sum_{t'=1}^{\infty} \left[\sum_{w=1}^{n} \nu(1-\nu)^{nt'+w-1}\right](1-\gamma_f)^{t'} \\
&\quad + \gamma_f^2 \sum_{q=1}^{\infty} q(1-\nu)^{nq}(1-\gamma_f)^{q-1}, \\
&= \frac{\gamma_f(1-\gamma_f)(1-(1-\nu)^n)(1-\nu)^n}{(1-(1-\nu)^n(1-\gamma_f))^2} - \frac{\gamma_f(1-\gamma_f)(1-\nu)^n(1-(1-\nu)^n)}{1-(1-\nu)^n(1-\gamma_f)} \\
&\quad + \frac{\gamma_f^2(1-\nu)^n}{(1-(1-\nu)^n(1-\gamma_f))^2}, \\
&= \frac{\gamma_f(1-\gamma_f)^2(1-(1-\nu)^n)(1-\nu)^{2n}}{(1-(1-\gamma_f)(1-\nu)^n)^2} + \frac{\gamma_f^2(1-\nu)^n}{(1-(1-\gamma_f)(1-\nu)^n)^2}. \qquad (4.58)
\end{aligned}$$

This is the expression obtained by substituting $d = 2n$ in (37) of [70].

4.B Impact of Packet Errors

4.B.1 Gain Formula

As explained in Section 4.6.3, we append two additional dummy modes *START* and *FINISH*, and evaluate the gain from *START* to *FINISH*. *START* is connected to $\mathcal{C}$ by a branch with gain 1. Similarly, *FINISH* is connected to $\mathcal{A}$ by a branch of unity gain. Fig. 4.4 shows the technique for $\mathcal{A} = \mathcal{R}_{n-1}$. The various forward paths for different modes to which *FINISH* is appended, their respective forward path gains $F_q(x, l)$ and co-factors $\Delta_q(x, l)$ are given in Table 4.2.

From $\rightarrow$ To	$F_q(x,l)$	$\Delta_q(x,l)$
$\mathcal{C} \rightarrow \mathcal{R}_{n-k-1}$ $(0 \leq k \leq n-1)$	$x\nu l^{k+1}$	1
$\mathcal{C} \rightarrow \mathcal{C}$	1	$1-(1-\gamma_f)(1-e_h)((1-e_d)x+e_d)l^n$

Table 4.2: Terminal model with packet errors: forward path gains and corresponding cofactors

The graph determinant $\Delta(x,l)$ is

$$\begin{aligned}\Delta(x,l) &= 1-\nu e_h(1-\gamma_f)l^{n+1}x-(1-\gamma_f)(1-e_h)((1-e_d)x+e_d)l^n-(1-\nu-\gamma_s)l \\ &\quad +(1-\nu-\gamma_s)((1-\gamma_f)(1-e_h)((1-e_d)x+e_d)l^n). \qquad (4.59)\end{aligned}$$

Using the representation for $g_\Omega(x,l)$ and $g_{\mathcal{R}_{n-1}}(x,l)$ in (4.29) and (4.30), respectively, the polynomials $\alpha_\Omega(l)$, $\beta_\Omega(l)$, $\alpha_{\mathcal{R}_{n-1}}(l)$, $\beta_{\mathcal{R}_{n-1}}(l)$, $\gamma(l)$, and $\delta(l)$ are given by

$$\begin{aligned}\alpha_\Omega(l) &= 1-(1-\gamma_f)(1-e_h)e_d l^n, \qquad (4.60)\\ \beta_\Omega(l) &= \sum_{k=0}^{n-1}\nu l^{k+1}-(1-\gamma_f)(1-e_h)(1-e_d)l^n, \qquad (4.61)\\ \alpha_{\mathcal{R}_{n-1}}(l) &= 0, \qquad (4.62)\\ \beta_{\mathcal{R}_{n-1}}(l) &= \nu l, \qquad (4.63)\\ \gamma(l) &= 1-(1-\gamma_f)(1-e_h)e_d l^n-(1-\nu-\gamma_s)l \\ &\quad +(1-\nu-\gamma_s)((1-\gamma_f)(1-e_h)e_d l^n, \qquad (4.64)\\ \delta(l) &= -\nu e_h(1-\gamma_f)l^{n+1}-(\nu+\gamma_s)(1-\gamma_f)(1-e_h)(1-e_d)l^n. \qquad (4.65)\end{aligned}$$

Chapter 5 New Call Access Control

5.1 Need for Call Admission Control

The previous chapters dealt with cellular system scenarios where a given number of users were contending for resources. The users that contend among themselves for available resources need to be first admitted into the system. The admission control mechanism that does so has a significant impact on the quality of service seen by the users in the system. For example, an excessive number of users participating in the contention phase of multiple access protocols like PRMA would lead to a severe degradation in the voice packet dropping probability in Chapter 4. Similar observations can also be made for the link adaptation problem considered in Chapter 3. The average packet waiting time for the stable queue scenario in Section 3.5 increases with an increase in the number of users wanting to transmit in a slot. In this chapter we study an aspect related to hand-offs of this access control problem for cellular systems.

In cellular systems with a FDMA/TDMA air interface, the channels[1] assigned to users are orthogonal. The channel assignment scheme is responsible for assigning a channel to a user. This is done for the duration of a user's stay inside a cell, or until his call terminates.[2] When a user moves into another cell, a new call may be assigned to the user in the new cell. Therefore, channels need to be assigned to two different kinds of requests: new call requests and hand-off call requests. If the system does not (or cannot) assign a channel to the call that wants to hand-off to the cell it has moved into, the call request is rejected. This can lead to the call getting dropped in the middle of a conversation. Thus, hand-off has a significant impact on the quality of service perceived by mobile users currently using the system. The perceived drop in quality when a request from a new call, which is trying to initiate a conversation, is rejected is generally much less.

[1] Here, a channel can mean either a time slot or a frequency band.

[2] This is not to imply that the same channel is assigned to the user for the duration of his call.

Whether resources are available when hand-off call requests arise depends on the new call access control strategy in use. For example, a new call access control strategy that is designed to maximize the number of users per cell[3] may not do well in terms of handling hand-off traffic. On the other hand, an extremely conservative new call access strategy would reject almost all new call requests so as to always be able to handle hand-off call requests. Both these scenarios may be unacceptable to the cellular system designer. It can be seen that the new call access control strategy can be used to implement a trade-off in allocating resources to new calls and hand-off calls. This is relevant especially when hand-offs are more expensive to reject than new call requests. Therefore, the access control procedure needs to be different for new call requests and hand-off call requests.

In this chapter we propose several techniques for modifying the new call access control strategies used in existing unprioritized channel assignment schemes (CAS) that treat new call requests and hand-off call requests the same. The techniques we propose control new call access based on whether their estimates of the future system configuration(s) meet certain required criteria. These criteria are elaborated upon in the following sections. Using prediction based rules enables the techniques to handle traffic with non-uniform mobility. However, this approach is not without its disadvantages. The first issue, which we shall see later, is the computational complexity involved in estimating the possible future system configurations. This makes the new call access control techniques centralized. Second, the optimal criteria for judging the various possible future system configurations, to ensure an optimal allocation of resources between hand-off call and new call requests, are as yet unknown. The techniques we study are therefore heuristic.

5.2 Chapter Organization

The various CASs proposed in literature are described in Section 5.3. The CAS model we use is given in Section 5.4. The prediction based new call access control techniques are

[3]The new call access strategy that maximizes the number of users per cell is as yet unknown. The solution to this problem is known only for the asymptotic case when the number of channels tends to infinity and there is no user mobility. This solution is called the Timid Dynamic Channel Assignment Algorithm (TDCAA) [83]. We briefly describe this scheme in Section 5.3.

described in Sections 5.5 and 5.6. Section 5.7 describes the system model and analysis. The results are presented in Section 5.8. A simpler new call access control criterion is investigated in Section 5.9. We conclude in Section 5.10.

5.3 Channel Assignment Scheme Literature

Many channel assignment schemes have been proposed and studied in the literature. The schemes differ in the complexity of implementation, the centrality of the process used for resource assignment, and the overhead incurred in handling channel assignment requests. A comprehensive survey of most of these schemes can be found in [9, 84]. In Fixed Channel Assignment (FCA) [20, Chp. 2] [22, Chp. 1], a fixed set of channels is assigned to every cell. The channel sets assigned to cells constituting a reuse cluster are disjoint. The reuse cluster size is determined by the amount of co-channel interference that is tolerable. In order to make a cellular network handle variations in traffic (which may occur over space or time), channel borrowing schemes like Borrowing with Channel Ordering (BCO) [85] and Borrowing with Directional Channel Locking (BDCL) [86] have been proposed. In these schemes unused channels from neighboring cells (also called nominal cells) are borrowed to accommodate more calls. The channels so borrowed are 'locked', *i.e.*, they are unavailable for use in the adjacent nominal cells.

Dynamic Channel Assignment (DCA) schemes, on the other hand, put all the channels into a central pool from which they are assigned on demand to each cell. In Maximal Packing (MP) [87], a call request is blocked only if there is no possible reallocation of channels to calls in progress that can result in the request being accepted. For a two-dimensional cellular network, MP is an idealization because of the large number of reassignments over the entire cellular network that it might need to make. However, for a linear array of cells, Raymond [88] has shown that it is possible to implement MP without ever making more than two reassignments upon the arrival of a call request. Note that this is true independent of the size of the network. Locally Optimized Dynamic Assignment (LODA) [86] is another DCA scheme. It assigns a cost to each candidate channel based on the state of neighboring cells, and chooses the channel with the least cost. The cost function is de-

signed to prefer channels with a higher usage frequency. It also prefers channels that are used by cells spaced close together. The Timid Dynamic Channel Assignment Algorithm (TDCAA) [83] is a DCA scheme that does not reassign channels of calls in progress. It chooses a channel at random from the set of available channels and assigns it to a call request. Deora [83] showed that TDCAA is "asymptotically optimal in the sense that as the number of channels increases, the carried traffic per channel is at least as large as can be achieved by any other algorithm."

5.3.1 Hand-offs in Channel Assignment Schemes

The schemes described so far do not distinguish between new call and hand-off call requests. The guard channel scheme, analyzed by Oh and Tcha [36], prioritizes hand-offs by means of guard channels. Guard channels are reserved solely for hand-off calls, and are not available for new call requests. The sub-rating scheme proposed by Lin, Noerpel, and Harasty [89] is another strategy for accommodating hand-off call requests. In sub-rating, an occupied full rate channel is temporarily divided into two half-rate channels – one to serve the existing call, and the other to serve the hand-off call request. If the channel splitting only occurs for relatively short periods during an on-going conversation, its impact on voice quality is, arguably, negligible.

Queuing hand-off requests when channels are unavailable is also a technique for reducing hand-off call rejection probability [84]. The queued hand-off call requests are given priority over new call requests. Queuing new call requests instead of hand-off call requests has also been explored in the literature [84]. Naghshineh and Schwartz [90] proposed a distributed call admission control algorithm that takes into account the number of calls in adjacent cells in addition to the number of calls in the cell where a new call request arises. Their call admission algorithm was based on approximations for overload probability in a cellular system with FCA. It tries to ensure that the overload probability, due to hand-offs from adjacent cells, for a cell accepting a new call and for its neighboring cells, is less than an acceptable threshold.

The new call blocking probability and the hand-off blocking probability are the perfor-

mance metrics used for analyzing the performance of all of the above schemes.

5.3.2 Our Work

As mentioned earlier, we investigate techniques for modifying new call access criteria in unprioritized channel assignment schemes to introduce hand-off prioritization. This is implemented by generalizing the concept of guard channels to that of reject states. The system in a reject state does not accept new calls. Hand-off calls are accepted as allowed by the parent unprioritized CAS.

Analysis of channel assignment schemes has traditionally been done by considering an isolated cell, and modeling the new call and hand-off call requests as two independent Poisson arrival processes [31, 38, 36, 91]. This model does not take into account the fact that a hand-off involves multiple cells. Sidi and Starobinski [35] have observed that new call blocking and hand-off call blocking are fundamentally different. While the hand-off blocking probability goes to 0 as the hand-off rate increases, the new call blocking probability has a non-zero lower bound. We use a complete system state description technique to track the system, and avoid the simplifying approximations made in previous analyses. However, our approach does significantly increase the complexity of system analysis and prediction.

5.4 Channel Assignment Scheme Description

A channel assignment scheme (CAS) assigns channels to call requests without violating the reuse constraints of the system. The following classification of states can be defined for any CAS.

Reject state with respect to a cell i For a given channel assignment scheme, a state $\underline{n}$ is said to be a *reject state with respect to cell* i if no new call requests in cell i are accepted by the scheme when the system state is $\underline{n}$.

Accept state with respect to a cell i A state that is not a reject state with respect to cell i is an *accept state*.

Let G_i denote the set of accept states with respect to cell i. The call acceptance policy for a CAS can then be written as follows:

New call request in cell i A new call request in cell i is accepted if the current system state $\underline{n}$ is an accept state, *i.e.*, $\underline{n} \in G_i$.

Hand-off call request A hand-off call request from cell j to cell i is accepted only if the resultant state vector is a valid state.

Hand-off prioritization is implemented by restricting G_i to a subset of the unprioritized scheme's set of accept states Ω_i. The hand-off prioritization problem then reduces to determining whether a given state in Ω, the set of all system states, should be classified as reject or accept. The guard channel scheme is clearly a special case of the scheme based on accept/reject states since it can be described in terms of the latter. However, more sophisticated criteria can now be employed for determining the appropriate call acceptance criteria.

5.4.1 State Description

A channel assignment scheme also imposes its own set of constraints, in addition to the reuse constraints. For example, in FCA the set of channels allocated to a cell cannot change with time. The state of channel allocation (system state) in the system is defined as the allocation state of each channel, where the allocation state of a channel is the set of cells in which the channel is currently in use. The parameters that characterize the system state depend on the channel allocation scheme under consideration. For schemes such as MP, which can potentially reassign channels at every call request, and FCA, which has a static channel allocation, the occupancy vector suffices as the system state [87, 88, 92, 93]. The occupancy vector $\underline{n}$ is defined as

$$\underline{n} = \begin{bmatrix} n(1) \\ \vdots \\ n(C) \end{bmatrix},$$

where $n(i)$ is the number of active mobiles in cell i and C is the total number of cells. For schemes that, in addition, use user location information for determining channel allocation [41], more parameters are required to define the system state. This is also true for TDCAA and BCO, which do not reassign channels allocated to calls in progress – the instantaneous occupancy vector does not suffice to describe the system state.

For the sake of simplicity, we restrict our attention to schemes wherein the occupancy vector is the system state. The global state information to be collected is then minimal. We refer to the occupancy vector as the system state from now on. It must be noted that these concepts are applicable in general to any unprioritized channel assignment scheme for which the state can be suitably defined. The channel reuse constraints, which specify how close two cells that use the same channel can be to each other, and the CAS's constraints are given by a set of linear inequalities of the form [93]

$$A\underline{n} \leq \underline{c}. \tag{5.1}$$

An occupancy vector $\underline{n}$ is valid if and only if it satisfies (5.1). The CAS is uniquely characterized by the matrix A and vector $\underline{c}$. The matrix A is determined by the channel reuse and CAS constraints, and the vector $\underline{c}$ is proportional to the total number of channels.

5.4.2 Evaluating A and $\underline{c}$

We now describe A and $\underline{c}$ for FCA and MP.[4] Let m channels be available to the system, and let the reuse distance r be 2. We consider a circular cellular array, shown in Fig. 5.1, to illustrate the schemes. Making it a circular array avoids unwanted effects that may arise in the boundary cells.

FCA A system state $\underline{n}$ is valid for FCA if it satisfies[5]

$$n(i) \leq \frac{m}{2}, \ \forall\, i \in \{1, \ldots, C\}. \tag{5.2}$$

[4]Not all cellular system configurations are covered by this representation. For example, for a circular cellular array, the number of cells has to be even.

[5]For reuse distance r, a system state $\underline{n}$ is valid if it satisfies $n(i) \leq \frac{m}{r}$, $\forall\, i \in \{1, \ldots, C\}$.

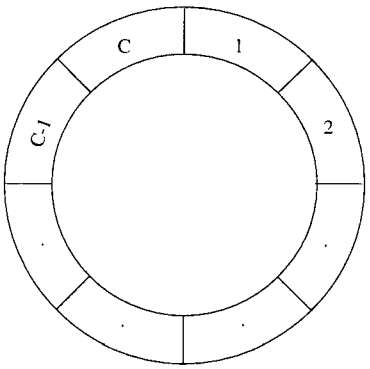

Figure 5.1: Circular cellular array

This leads to the following values for A and $\underline{c}$:

$$\begin{aligned} A &= \begin{bmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 0 & & & & 1 \end{bmatrix}_{C\times C}, \\ \underline{c} &= \frac{m}{2}\underline{J}, \end{aligned} \tag{5.3}$$

where

$$\underline{J} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{C\times 1}.$$

MP A system state $\underline{n}$ is valid for MP if it satisfies[6]

$$\begin{aligned} n(i) + n(i+1) &\leq m, \forall\, i \in \{1, \ldots, C-1\}, \\ n(C) + n(1) &\leq m. \end{aligned} \tag{5.4}$$

[6]For reuse distance r, a system state $\underline{n}$ is valid if it satisfies

$$\begin{aligned} n(1) + n(2) + \cdots + n(r) &\leq m, \\ n(2) + n(3) + \cdots + n(r+1) &\leq m, \text{ and so on.} \end{aligned}$$

This leads to

$$A = \begin{bmatrix} 1 & 1 & 0 & \cdots & \\ 0 & 1 & 1 & 0 & \cdots \\ \vdots & & \ddots & & \\ 0 & \cdots & 0 & 1 & 1 \\ 1 & 0 & \cdots & 0 & 1 \end{bmatrix}_{C \times C},$$

$$\underline{c} = m\underline{J}. \tag{5.5}$$

5.4.3 Unprioritized Scheme's Call Acceptance Criteria

Eqn. (5.1) also serves as the defining criterion for accepting or rejecting new call and hand-off call requests. Specifically, consider a CAS described by (5.1). Let $\underline{e}_i$, the increment vector, be defined as

$$\underline{e}_i(j) = \begin{cases} 0, & j \neq i \\ 1, & j = i \end{cases}.$$

If the present state is $\underline{n}$, a new call arriving in cell i is accepted only if the resultant state $(\underline{n} + \underline{e}_i)$ is valid, *i.e.*, it satisfies

$$A(\underline{n} + \underline{e}_i) \leq \underline{c}. \tag{5.6}$$

Similarly, a hand-off call request from cell j to cell i is accepted if the resultant state $(\underline{n} + \underline{e}_i - \underline{e}_j)$ satisfies

$$A(\underline{n} + \underline{e}_i - \underline{e}_j) \leq \underline{c}. \tag{5.7}$$

5.4.4 Hand-off Failure Scenarios

When a hand-off call request is rejected by the channel assignment scheme, two scenarios are possible for dealing with the request:

Failed hand-off dropped The failed hand-off call is dropped from the system. In other words, the failed requests are not queued to be served whenever resources become available later.

Failed hand-off returned The failed hand-off call is not dropped, but is instead returned to the cell from which its request originated. This model is appropriate when adjacent cells overlap. So long as the mobile remains in the region of overlap between the two cells, it can try hand-offs again and again.

We evaluate the hand-off failure probability (b_h) and the new call blocking probability (b_n) for both these scenarios. Pallant and Taylor [37] have shown that closed form solutions exist for the new call blocking probability and hand-off failure probability for certain values of system parameters, when failed hand-offs are returned to the system. We shall use their formulae to cross-check our results whenever possible. The conditions under which closed form solutions exist are elaborated upon later.

We now describe two prediction based criteria that are studied in this chapter. Both the criteria classify a state as accept or reject based on the configurations the system is likely to be in after time τ. It must be pointed out here that the evolution of the system over the interval τ, and not just the resulting configurations, should ideally be used by the decision criteria. However, the large number of possible evolutions imposes a large computational complexity burden that prevents us from basing the state classification on the entire evolution.

5.5 Cost Function Criterion (CFC)

In the cost function criterion, the classification of states as accept or reject is based on the expected cost function at some future time τ. The cost function is taken to be a weighted average of the hand-off failure and new call blocking probabilities. It is of the form $(1-\alpha)b_n + \alpha b_h$, with $\alpha > 0.5$ to reflect the higher cost of rejecting hand-off call requests.

After time τ, the system can be in one of many possible configurations. Not all of these configurations will be valid CAS compatible states. An invalid future configuration indicates a violation of the CAS constraints; it is therefore penalized. Specifically, a maximum penalty of 1 is assigned to the new call blocking and hand-off failure probabilities for such configurations in the cost function calculation. Hand-off and new call blocking

probabilities can be calculated for each of the other remaining valid states as follows.

$$b_n(\underline{n_\tau}) = \begin{cases} \frac{\sum_{i=1}^{C} \lambda_i I[\underline{n_\tau}+\underline{e}_i \notin \Omega]}{\sum_{i=1}^{C} \lambda_i}, & \forall\, \underline{n_\tau} \in \Omega \\ 1, & \forall\, \underline{n_\tau} \notin \Omega \end{cases},$$

$$b_h(\underline{n_\tau}) = \begin{cases} \frac{\sum_{i=1}^{C} \sum_{k=1}^{C} n_\tau(i)\gamma_i p(i,k) I[\underline{n_\tau}-\underline{e}_i+\underline{e}_k \notin \Omega]}{\sum_{i=1}^{C} n_\tau(i)\gamma_i}, & \forall\, \underline{n_\tau} \in \Omega \\ 1, & \forall\, \underline{n_\tau} \notin \Omega \end{cases}, \tag{5.8}$$

where the indicator function $I[.]$ is defined as

$$I[x] = \begin{cases} 1, & \text{if } x \text{ is true} \\ 0, & \text{if } x \text{ is false} \end{cases}.$$

Evaluating Expected Cost Function

If $A(\underline{n}+\underline{e}_i) \not\leq \underline{c}$, the new call request is not (cannot be) accepted. Otherwise the following calculations are performed to evaluate the expected cost function. If the new call is accepted, the present state of the system will be $(\underline{n}+\underline{e}_i) \in \Omega$. For all the possible states in the future $\underline{n_\tau}$, the probabilities $p_\tau(\underline{n_\tau}|(\underline{n}+\underline{e}_i))$ of the system being in state $\underline{n_\tau}$ after time τ given that the present state is $(\underline{n}+\underline{e}_i)$ are calculated. The expected new call blocking probability $\widehat{b}_{n\tau}(\underline{n}+\underline{e}_i)$ and the expected hand-off failure probability $\widehat{b}_{h\tau}(\underline{n}+\underline{e}_i)$ are obtained from

$$\widehat{b}_{n\tau}(\underline{n}+\underline{e}_i) = \sum_{\underline{n_\tau}} p_\tau(\underline{n_\tau}|(\underline{n}+\underline{e}_i)) b_n(\underline{n_\tau}),$$

$$\widehat{b}_{h\tau}(\underline{n}+\underline{e}_i) = \sum_{\underline{n_\tau}} p_\tau(\underline{n_\tau}|(\underline{n}+\underline{e}_i)) b_h(\underline{n_\tau}). \tag{5.9}$$

If the expected cost function exceeds the threshold T, a new call request in cell i is not accepted. Hence, the new call acceptance criterion is as given below.

New Call Acceptance Criterion

A new call request in cell i, when the system state is $\underline{n}$, is accepted if

$$A(\underline{n}+\underline{e}_i) \leq \underline{c}, \text{ and}$$

$$(1-\alpha)\widehat{b}_{n\tau}(\underline{n}+\underline{e}_i)+\alpha\widehat{b}_{h\tau}(\underline{n}+\underline{e}_i) \;\leq\; T. \tag{5.10}$$

The threshold T and look-ahead time τ need to be chosen in CFC.

5.6 Maximum Likelihood State (MLS) Scheme

A simple extension of the guard channel scheme is to apply its criterion to the most likely future system state instead of just the present state. Therefore, the criterion we investigate in this section is the guard channel criterion. A new call is accepted in cell i if its acceptance doesn't violate the unprioritized scheme's validity constraint (5.1) and if, after accepting the new call, the most likely future state at a given future time τ meets the guard channel criterion. First we characterize the guard channel criterion.

5.6.1 Guard Channel Criterion Characterization

As per the guard channel criterion, a new call request in cell i is accepted if the resultant state $(\underline{n}+\underline{e}_i)$ satisfies

$$A(\underline{n}+\underline{e}_i) \leq \underline{c} - g\underline{J}_i, \tag{5.11}$$

where $\underline{J}_i$ is a vector defined by $\underline{J}_i(j) = 1$ whenever the j^{th} inequality in $A\underline{n} \leq \underline{c}$ is a function of $n(i)$; it is 0 otherwise. For example, in a 4 cell circular array with MP as the CAS, we have

$$\underline{J}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \; \underline{J}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \; \underline{J}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \underline{J}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}. \tag{5.12}$$

5.6.2 MLS New Call Acceptance Criterion

For a new call arrival in cell i when the system state is $\underline{n} \in \Omega$, let $\widehat{(\underline{n}+\underline{e}_i)}_\tau$ denote the most likely future state after time τ, given that the new call request is accepted. As per MLS, the

new call request in cell i is accepted when the system state is $\underline{n}$ if $\widehat{(\underline{n}+\underline{e}_i)}_\tau$ satisfies

$$\begin{aligned} A(\underline{n}+\underline{e}_i) &\leq \underline{c}, \text{ and} \\ A\widehat{(\underline{n}+\underline{e}_i)}_\tau &\leq \underline{c} - g\underline{J}_i. \end{aligned} \tag{5.13}$$

The MLS scheme is parameterized by the look-ahead time τ and the number of guard channels g.

5.7 System Model Details

The Maximum Packing (MP) CAS was chosen as the parent unprioritized scheme since it has the largest valid state space Ω among channel assignment schemes. We study a 4 cell, 3 channel circular cellular system to maintain a reasonable complexity in our analysis. The matrices A and $\underline{c}$ are given by

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \underline{c} = \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix}. \tag{5.14}$$

The new call arrival process is assumed to be Poisson, and the call duration is assumed to be exponentially distributed with mean $1/\mu$. The cell dwell time, discussed in Section 2.5.2, is an exponential random variable with mean $1/\gamma_i$ in cell i. Let $p(i,j)$ denote the probability that a user in cell i moves to cell j, given that he is moving out of cell i. The call termination rate μ is normalized to 1. For a 4 cell, 3 channel system, the number of valid states for the MP scheme is $|\Omega| = 70$. The number of accept states, with respect to any cell i, for the unprioritized scheme is $|\Omega_i| = 40$.

Two different types of system traffic – uniform traffic and hot spot traffic – are analyzed. In uniform traffic, all the cells have the same traffic parameters. In hot spot (HS) traffic, the mobility traffic statistics vary from one cell to another. The HS cell is defined to be the point of traffic congestion. The mobiles in the cells neighboring the HS cell exhibit

a greater tendency to move into the HS cell than move away from it. The parameters for these two traffic models are given below.

Uniform traffic (UT) The cells have the same traffic parameters. The traffic parameters take the following values.

- *New call arrival rates:*

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda. \tag{5.15}$$

- *Hand-off rates:*

$$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma. \tag{5.16}$$

- *Directional probabilities:*

$$\begin{aligned} p(1,2) &= p(1,4) = 0.5, \\ p(2,3) &= p(2,1) = 0.5, \\ p(3,4) &= p(3,2) = 0.5, \\ p(4,1) &= p(4,3) = 0.5. \end{aligned} \tag{5.17}$$

Hot Spot (HS) traffic Cell 2 is taken to be the HS cell. There is a tendency for mobiles in cells 1 and 3 to move into the HS cell. Also the mobiles in the HS cell, as compared to other cells, exhibit a lesser tendency to move out of their cell. The traffic parameters, illustrated in Fig. 5.2, are as follows.

- *New call arrival rates:*

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda. \tag{5.18}$$

- *Hand-off rates:*

$$\begin{aligned} \gamma_1 &= \gamma_3 = \gamma_4 = \gamma, \\ \gamma_2 &= \frac{\gamma}{2}. \end{aligned} \tag{5.19}$$

- *Directional probabilities:*

$$p(1,2) = 0.9,\ p(1,4) = 0.1,$$
$$p(3,4) = 0.1,\ p(3,2) = 0.9,$$
$$p(2,3) = p(2,1) = 0.5,$$
$$p(4,1) = p(4,3) = 0.5. \tag{5.20}$$

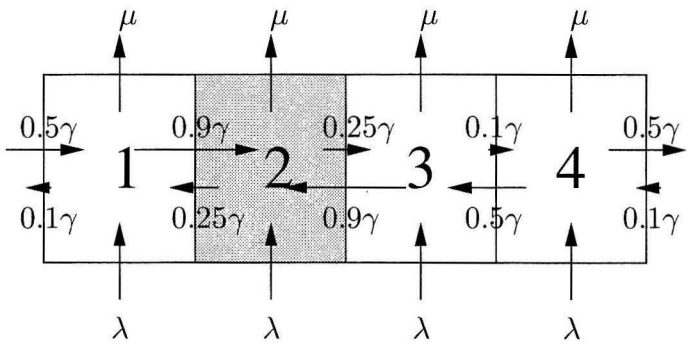

Figure 5.2: Hot spot traffic parameters

The above models do not consider traffic that has multiple mobility rates, like pedestrian traffic along with vehicular traffic. This simplification in the model is necessary to maintain analytical tractability.

5.7.1 Evaluation of Transition Probabilities

The calculation of the transition probabilities, which needs to be done for the prediction based criteria that we consider, is explained below. The state $\underline{n}$ of the cellular system executes a continuous time Markov process. Since the movement of users is independent of the CAS, the state evolution is also independent of the CAS. The impact of new call arrivals is not taken into consideration in determining the transition probabilities since the decision on the new call request under consideration is itself not known. Consequently, the system evolution, which is calculated for determining the expected cost function or the most likely state, is governed by the mobility and the call termination processes. This

implies that the number of users in the system decreases due to call termination as the prediction interval of length τ increases.

Notation: Let F denote the transition rate matrix for this process. The matrix element $F(\underline{s}, \underline{n})$ denotes the transition rate from state $\underline{n}$ to $\underline{s}$ ($\underline{n}, \underline{s} \in \Gamma$), where Γ is the set of all possible future system configurations. F is given by

$$F(\underline{s}, \underline{n}) = \begin{cases} n(i)\gamma_i p(i,j), & \text{if } \underline{s} = \underline{n} - \underline{e}_i + \underline{e}_j, \text{ for some } i, j \ (i \neq j) \\ n(i)\mu_i, & \text{if } \underline{s} = \underline{n} - \underline{e}_i, \text{ for some } i \\ 0, & \text{otherwise} \end{cases},$$

$$F(\underline{n}, \underline{n}) \triangleq -\sum_{\underline{s} \in \Gamma \setminus \{\underline{n}\}} F(\underline{s}, \underline{n}). \tag{5.21}$$

The transition probability matrix $\mathcal{T}_\tau$ for time interval of length τ, in which $\mathcal{T}_\tau(\underline{s}, \underline{n}) = p_\tau(\underline{s}|\underline{n})$, is then given by [55]

$$\mathcal{T}_\tau = \exp(\tau F). \tag{5.22}$$

$p_\tau(\underline{s}|\underline{n})$ can thereby be calculated.

5.7.2 Analysis

The new call blocking probability (b_n) and hand-off failure probability (b_h) of a CAS can be calculated from the steady state probabilities of system states. b_n and b_h for any CAS, with set of accept states G_i with respect to cell i, are given by

$$\begin{aligned} b_n &= \frac{\sum_{\underline{n} \in \Omega} (\sum_{i=1}^{C} \lambda_i I\left[\underline{n} \notin G_i\right]) \pi(\underline{n})}{\sum_{\underline{n} \in \Omega} (\sum_{i=1}^{C} \lambda_i) \pi(\underline{n})}, \\ b_h &= \frac{\sum_{\underline{n} \in \Omega} (\sum_{i=1}^{C} \sum_{j=1}^{C} n(i)\gamma_i p(i,j) I\left[(\underline{n} - \underline{e}_i + \underline{e}_j) \notin \Omega\right]) \pi(\underline{n})}{\sum_{\underline{n} \in \Omega} (\sum_{i=1}^{C} n(i)\gamma_i) \pi(\underline{n})}, \end{aligned} \tag{5.23}$$

where $\pi(\underline{n})$ denotes the steady state probability that the system is in state $\underline{n}$ and the indicator function $I[.]$ is defined as

$$I[x] = \begin{cases} 1, & \text{if } x \text{ is true} \\ 0, & \text{if } x \text{ is false} \end{cases}. \tag{5.24}$$

The evaluation of the steady state probabilities $\pi(\underline{n})$ is given in Appendix 5.A.

For the class of schemes under consideration, no general product form [94] solution for the steady state probabilities is known to exist. However, for the failed hand-off call returned scenario, the conditions specified in [37] for a product form solution to exist are fulfilled for the parent unprioritized scheme with uniform traffic. Closed form solutions for b_h and b_n can then be found for this specific case. They are of the form

$$\begin{aligned} b_h &= 1 - \frac{G(A, \underline{c} - A\underline{e}_1)}{G(A, \min(\underline{c} - A\underline{e}_1, \underline{c} - A\underline{e}_2))}, \\ b_n &= 1 - \frac{G(A, \underline{c})}{G(A, \underline{c} - A\underline{e}_1)}, \end{aligned} \tag{5.25}$$

where $G(A, \underline{c})$ is the probability that the system state is the all-zero state. We will see shortly that setting τ asymptotically large in the prioritized scheme effectively removes the prioritization. Thus, for large τ, our results approach these closed form solutions.

5.8 Results

The set of accept states completely defines a scheme. Therefore, to understand the behavior of the schemes, we first plot the number of accept states as a function of τ. We then plot the resulting b_h, b_n, and net cost. We first study the MLS scheme followed by the CFC scheme. In the MLS scheme, we first investigate the hot spot traffic scenario followed by the uniform traffic scenario. High mobility ($\gamma = 10$) and low mobility ($\gamma = 1$) cases are investigated for each of these scenarios. We follow the same format for the CFC scheme.

5.8.1 MLS Scheme

Hot Spot (HS) Traffic

The HS scenario in the high mobility case brings out certain interesting features of the MLS scheme. The number of accept states as a function of the look-ahead time τ is plotted in Fig. 5.3. Notice that the number of accept states with respect to each of the cells is different. This can be explained as follows. Cells 1 and 3 have similar traffic parameters in (5.18)–(5.20). Therefore, the number of accept states assigned to these two cells is the same for all τ, *i.e.*, $|G_1| = |G_3|$. The HS cell 2, which attracts mobiles from its neighboring cells, typically has the smallest number of accept states assigned to it. This way, the MLS scheme ensures that a HS cell does not accept as many new call requests as the other cells. A larger number of hand-off call requests can then be accommodated in the HS cell. The cells adjacent to the HS cell also have a low accept state count to block almost as many new calls, since these new calls are likely to hand-off into the HS cell. Cell 4, which is farthest from the HS cell, has the largest number of accept states assigned to it by the MLS scheme.

Observe that in Fig. 5.3 the number of accept states assigned to the HS cell and its two neighbors decreases in the look-ahead interval $\tau \in (0.05, 0.22)$. It is in this interval that the effect of non-uniform traffic is pronounced. For small τ ($\tau \leq 0.05$), the movement of mobiles is not significant enough to have an effect. For large τ ($\tau \geq 0.22$), the call termination of mobiles diminishes the clustering effect of the unequal mobility parameters. Since subsequent new call arrivals are not considered in Section 5.7.1 for classifying states as accept or reject, the number of users in the system cannot increase during the prediction time interval of length τ. As a result, an increase in τ makes future states with fewer mobiles more probable. It can be seen that for $\tau \geq 0.65$ the set of accept states is 40, the same as that of the parent unprioritized CAS. This is a common feature of all the results presented about the MLS criterion. For $\tau = 0$, the MLS criterion is the same as the guard channel criterion, while for large τ, it is the same as the parent unprioritized CAS.

Figure 5.4 plots b_h, b_n, and cost $((1-\alpha)b_n + \alpha b_h)$, as a function of τ when failed hand-off calls are returned to the system. The weight factor[7] α is taken to be 10/11, making

[7]Note that the weight factor did not play any role in determining the new call acceptance criteria in MLS.

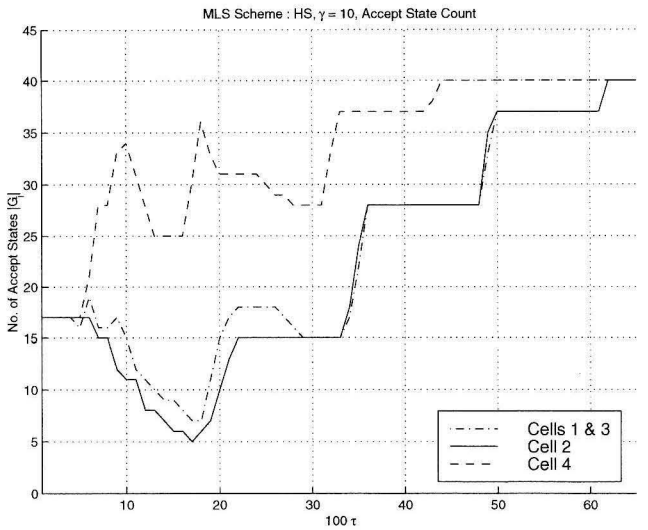

Figure 5.3: MLS: accept state count ($\gamma = 10$, HS)

hand-offs 10 times more expensive to reject than new calls. In the interval $(0.05, 0.22)$, where the number of accept states assigned to the HS cell and its neighbors decreases, MLS reduces the hand-off failure probability to nearly 0. However, this is at the expense of an increase in the new call blocking probability.

The low mobility case is studied in Fig. 5.5, which plots the number of accept states assigned to the 4 cells as a function of the look-ahead time τ. In contrast to the high mobility case, we now see that for each of the 4 cells, the number of accept states increases with τ. Since the speed of the mobiles in the system is low, their tendency to cluster in the HS cell takes longer to have an effect. Over this time period, the call termination process comes into play to compensate for this clustering tendency. The corresponding values of b_h, b_n, and cost are plotted in Fig. 5.6.

Figure 5.7 plots b_h and b_n *vs.* τ when rejected hand-offs are dropped from the system for the high mobility case. The failed hand-off dropped scenario shows the same trends but with lower values for the new call blocking and hand-off failure probabilities. This is because dropping rejected call requests makes the system spend less time in the boundary states that are responsible for blocking new call and hand-off call requests. For the rest of our performance results, we only plot the results for the case in which failed hand-off call requests are returned to the system.

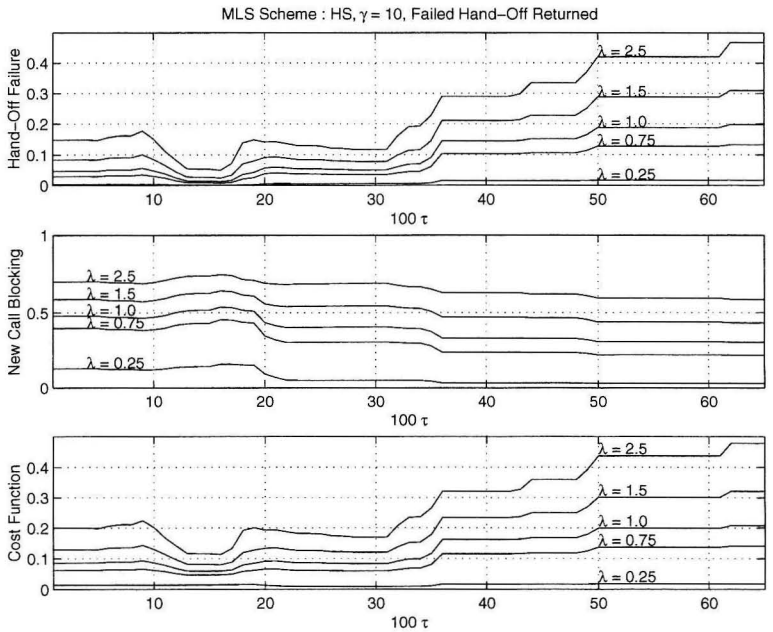

Figure 5.4: MLS: b_h, b_n, and cost (failed hand-off returned, $\gamma = 10$, HS)

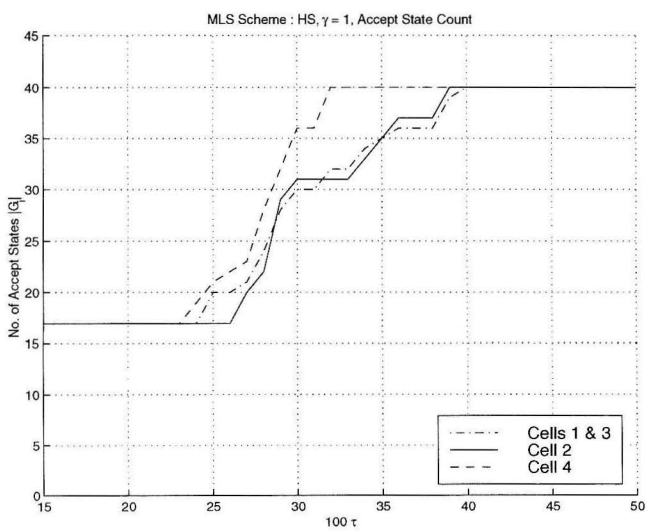

Figure 5.5: MLS: accept state count (failed hand-off returned, $\gamma = 1$, HS)

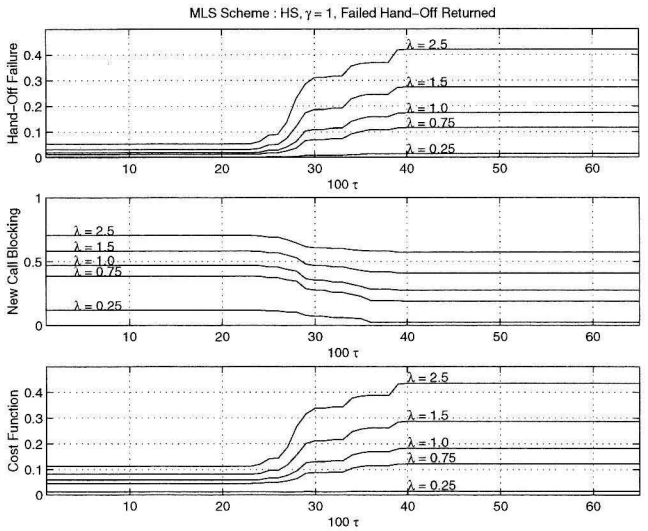

Figure 5.6: MLS: b_h, b_n, and cost (failed hand-off returned, $\gamma = 1$, HS)

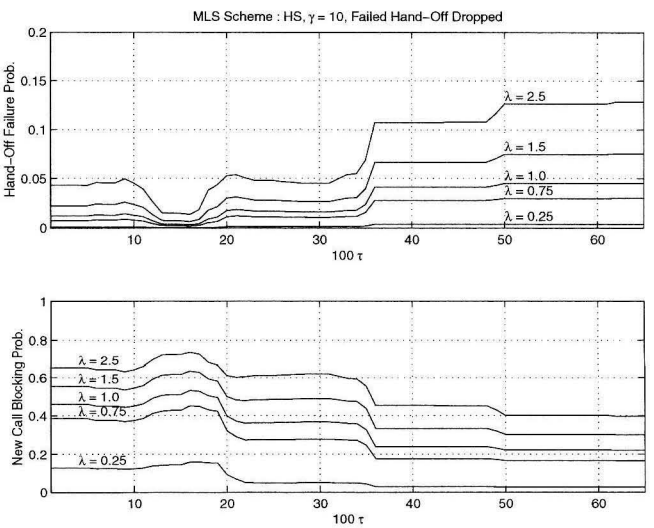

Figure 5.7: MLS: b_h and b_n(failed hand-off dropped, $\gamma = 10$, HS)

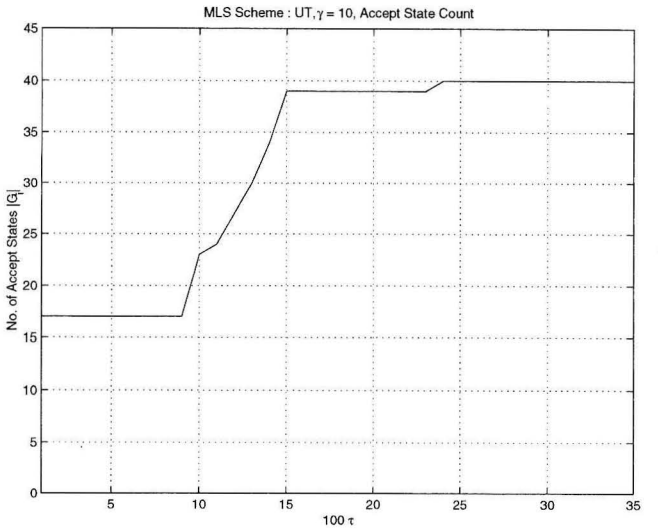

Figure 5.8: MLS: accept state count ($\gamma = 10$, UT)

Uniform Traffic

Figure 5.8 plots the number of accept states as a function of the look-ahead time τ for the high mobility case. Since the 4 cells have the same traffic parameters in (5.15)–(5.17), the number of accept states for each cell is now the same. The number of accept states increases with τ in this case. In Fig. 5.8, we see that the accept state space is the same as that of the parent scheme for $\tau \geq 0.24$. The variation in the number of accept states as a function of the look-ahead time τ shows similar trends for the low mobility case. It is therefore not plotted here.

Figure 5.9 plots b_h, b_n, and cost for the high mobility case as a function of the look-ahead time τ when failed hand-off calls are returned to the system. As the accept state count increases with increasing τ, the new call blocking probability decreases and the hand-off failure probability increases.

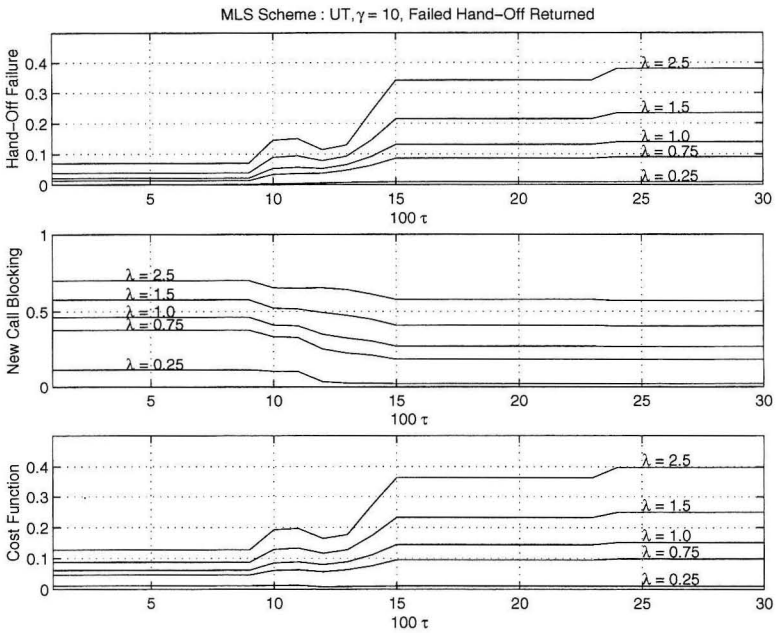

Figure 5.9: MLS: b_h, b_n, and cost (failed hand-off returned, $\gamma = 10$, UT)

5.8.2 CFC Scheme

Hot Spot Traffic

Figure 5.10 plots the number of accept states as a function of the look-ahead time τ for the high mobility case for different thresholds[8] T. The variation in the number of accept states for different cells is not as noticeable as in MLS for most thresholds. The averaging of expected cost function over various states in the future is partly responsible for this. As the threshold decreases, the number of states that accept new calls expectedly decreases. As in the MLS scheme we see that as τ increases the number of accept states increases and approaches that of the parent scheme. For $\tau \geq 0.82$, the set of accept states is the same as that of MP, the parent CAS. The CFC criterion is therefore the same as the parent unprioritized scheme for large τ.

The corresponding curves for b_h, b_n, and cost as a function of τ are plotted in Fig. 5.11. This is done for three different threshold values $T = 0.8, 0.6$, and 0.4. It is interesting to note that the cost function decreases as T decreases. Thus rejecting more new calls leads

[8]The case $T = 1$ is not shown since the set of accept states is then the same as that for the parent unprioritized scheme because the cost function cannot exceed 1.

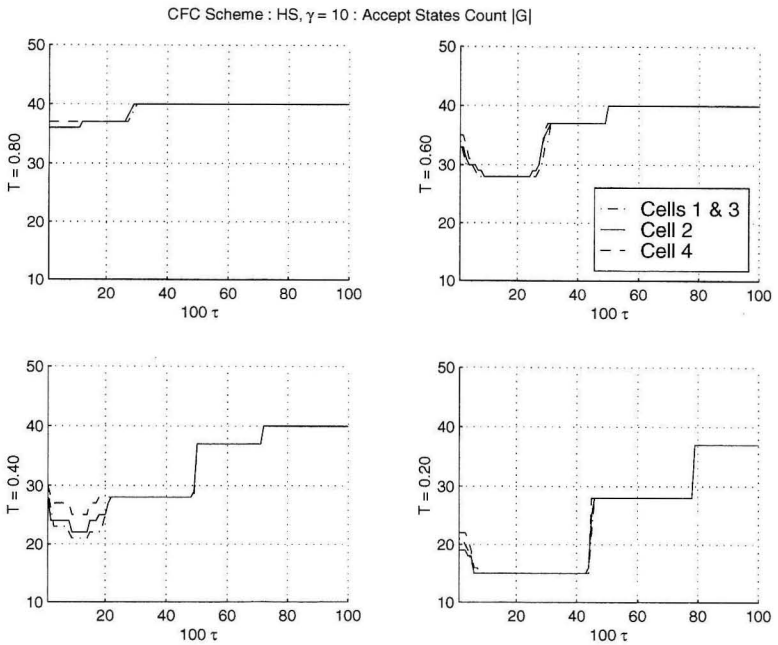

Figure 5.10: CFC: accept state count ($\gamma = 10$, HS)

to a lower cost function. This is primarily on account of the high weight factor of 10 for hand-off failure probability in the cost function. The results for the low mobility case are similar; so they are not shown here.

Uniform Traffic

Figure 5.12 plots the number of accept states for different cells, as classified by the CFC scheme, *vs.* the look-ahead time τ for the high mobility case. This is done for different values of the threshold T. The number of accept states decreases as the threshold T is reduced. The corresponding b_h, b_n, and cost *vs.* τ plots for the high mobility case with failed hand-off calls returned to the system are shown in Fig. 5.13.

5.8.3 Comparison

Table 5.1 and Table 5.2 compare the performance of the MLS, CFC, Guard channel, and parent unprioritized (MP) schemes. We first describe the MLS and CFC comparison. For

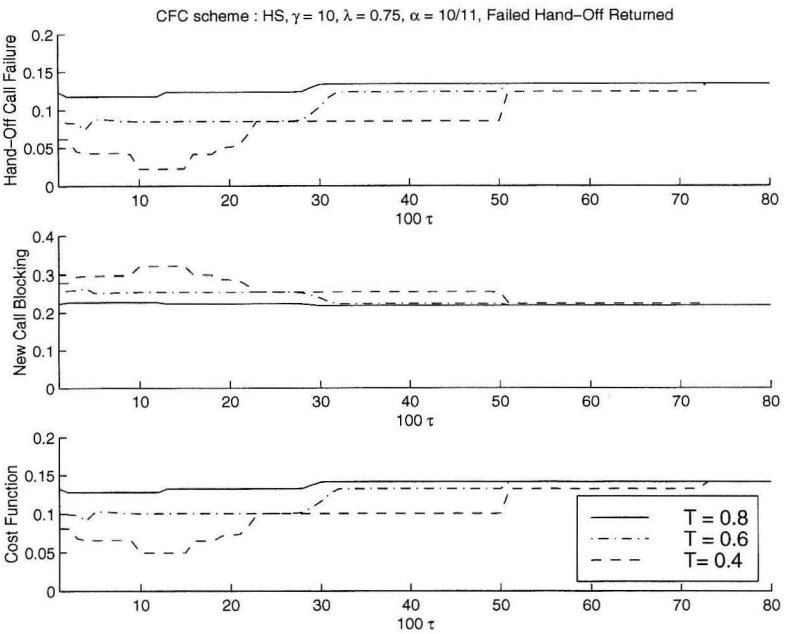

Figure 5.11: CFC: b_h , b_n, and cost (failed hand-off returned, $\gamma = 10$, HS)

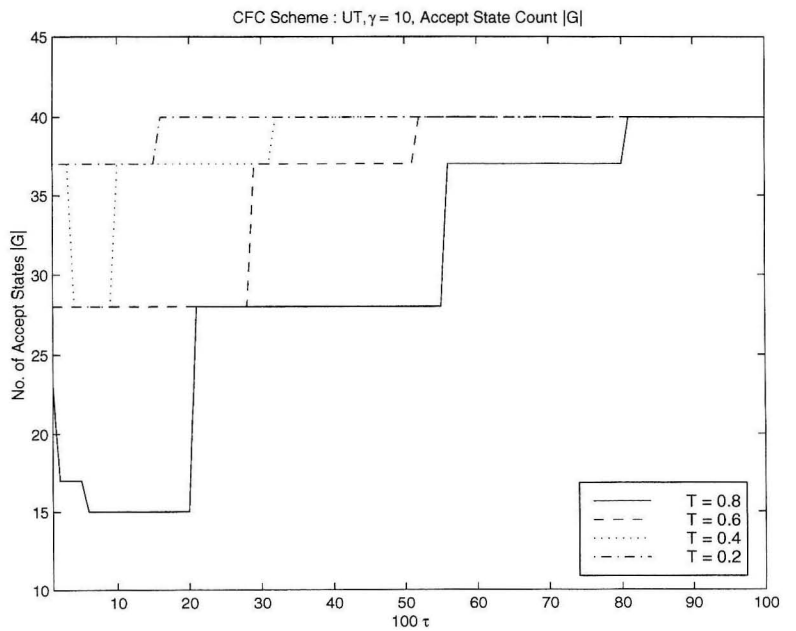

Figure 5.12: CFC: accept state count ($\gamma = 10$, UT)

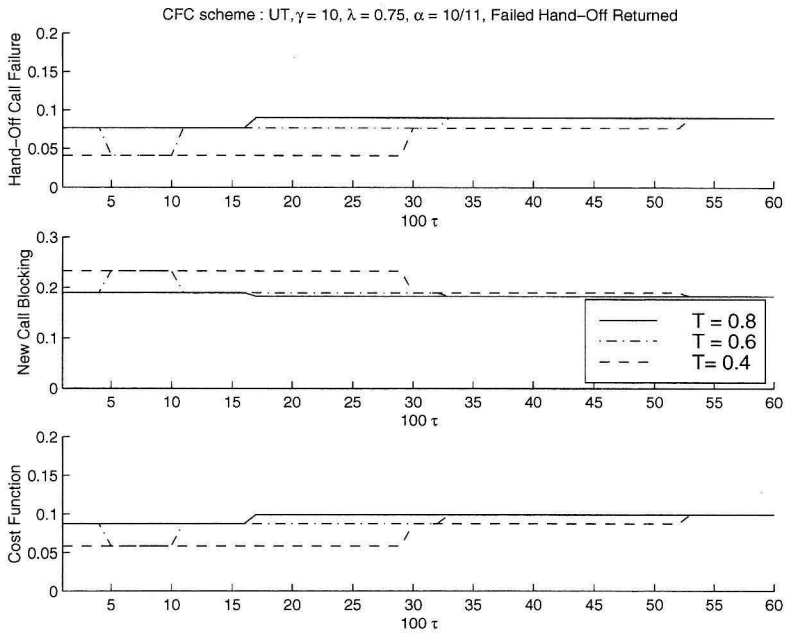

Figure 5.13: CFC: b_h, b_n, and cost (failed hand-off returned, $\gamma = 10$, UT)

this, we fix b_h to be the same for all schemes[9] and compute b_n for each scheme. Table 5.1 compares the schemes for the UT, $\gamma = 10$ case with $\lambda = 0.75$. Table 5.2 does the same for the HS, $\gamma = 10$ case with $\lambda = 0.75$. We find that in these tables the performance of MLS and CFC is roughly the same. The b_h and b_n achieved by the parent unprioritized scheme and the guard channel scheme are invariable since they have no freedom to trade off between b_h and b_n. However, in both MLS and CFC we can lower b_h (b_n) at the expense of increasing b_n (b_h) through appropriate adjustment of the schemes' parameters. We can also achieve the same performance as the guard channel or the parent schemes by appropriate choice of the parameters. While these results are for the failed hand-off returned case, the same trends also appear for the failed hand-off dropped case.

5.9 Total User Count (TUC) Based Scheme

Table 5.3 illustrates the increase in the size of the state space for the MP scheme as the number of channels or cells increases. This large state space makes calculation of the max-

[9]The b_h's of the schemes can only be made approximately equal since they use different state classification criteria.

	b_h	b_n	Parameters
MLS	0.0471	0.2263	$\tau = 0.13$
CFC	0.0412	0.2335	$\tau = 0.07, T = 0.4, \alpha = 0.9$
Guard Ch.	0.0135	0.3771	$g = 1$
MP	0.0904	0.1829	unprioritized

Table 5.1: Comparison between schemes (UT, $\gamma = 10, \lambda = 0.75$)

	b_h	b_n	Parameters
MLS	0.0457	0.2976	$\tau = 0.34$
CFC	0.0451	0.2944	$\tau = 0.19, T = 0.6, \alpha = 0.9$
Guard Ch.	0.0288	0.3969	$g = 1$
MP	0.1334	0.2183	unprioritized

Table 5.2: Comparison between schemes (HS, $\gamma = 10, \lambda = 0.75$)

imum likelihood state or the expected hand-off failure and new call blocking probabilities computationally expensive for a system with a large number of channels or cells.

For high mobility scenarios, the number of accept states, as a function of the look-ahead time τ, changes mainly due to valid states containing more and more mobiles. The classification of valid states as accept or reject can be simplified by basing the classification on the the total number of users in the state. The set of accept states with respect to cell i is

C	m	$\lvert\Omega\rvert$	$\lvert\Gamma\rvert$
2	2	6	6
2	3	10	10
2	4	15	15
4	2	26	70
4	3	70	210
4	4	155	495

Table 5.3: Increase in MP State Space

b_h	b_n	TUC	CFC ($\alpha = 0.9$)
0.0904	0.1829	$n_{\max} = 6$	$T = 1.0, \tau = 0.00$
0.0769	0.1900	$n_{\max} = 4$	$T = 0.5, \tau = 0.05$
0.0412	0.2335	$n_{\max} = 3$	$T = 0.5, \tau = 0.12$

Table 5.4: TUC and CFC comparison ($\lambda = 0.75, \gamma = 10$, UT)

then given as

$$G_i = \left\{ \underline{n} \in \Omega_i : \sum_{j=1}^{C} n(j) \leq n_{\max} \right\}, \ (0 \leq n_{\max} \leq N_{\max}). \tag{5.26}$$

The threshold value $n_{\max} = N_{\max}$ corresponds to the parent unprioritized CAS. This scheme cannot assign a different number of accept states to cells with different cell parameters, so

$$|G_1| = |G_2| = |G_3| = |G_4| \text{ always.}$$

TUC is parameterized by $n_{\max}$ only; it therefore cannot yield all the state classifications that are possible with the CFC scheme.

5.9.1 Performance of TUC

Figure 5.14 plots the number of accept states with respect to $n_{\max}$.

Uniform Traffic

Figure 5.15 plots b_h, b_n, and cost for the high mobility ($\gamma = 10$) case with failed hand-off calls returned to the system. As $n_{\max}$ increases, the hand-off failure probability increases and the new call blocking probability decreases, since a greater fraction of new call requests are accepted by the CAS. Closed form solutions for the hand-off failure probability and new call blocking probability exist for the uniform traffic case when rejected hand-off requests are returned to the system, as shown in Appendix 5.B. Table 5.4 lists the parameters of the CFC scheme (T, τ) and the TUC scheme $(n_{\max})$ for which their performance is the same.

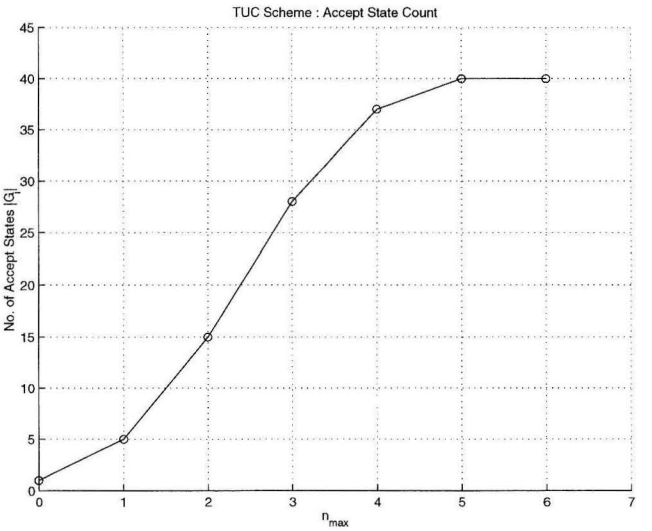

Figure 5.14: TUC: accept state count

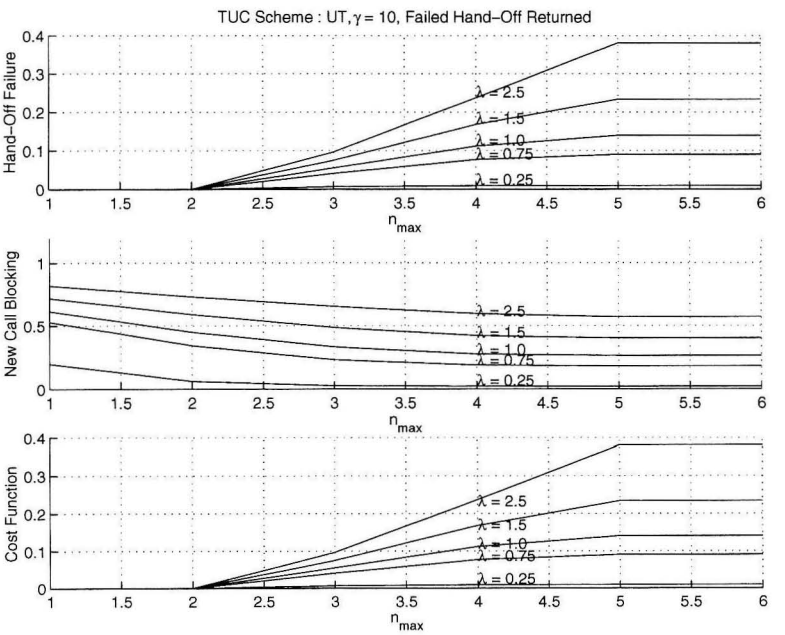

Figure 5.15: TUC: b_h and b_n (failed hand-off returned, $\gamma = 10$, UT)

b_h	b_n	TUC	CFC ($\alpha = 0.9$)
0.1334	0.2183	$n_{\max} = 6$	$T = 1.0, \tau = 0.00$
0.1231	0.2234	$n_{\max} = 4$	$T = 0.4, \tau = 0.42$
0.0849	0.2539	$n_{\max} = 3$	$T = 0.4, \tau = 0.16$

Table 5.5: TUC and CFC comparison ($\lambda = 0.75, \gamma = 10$, HS)

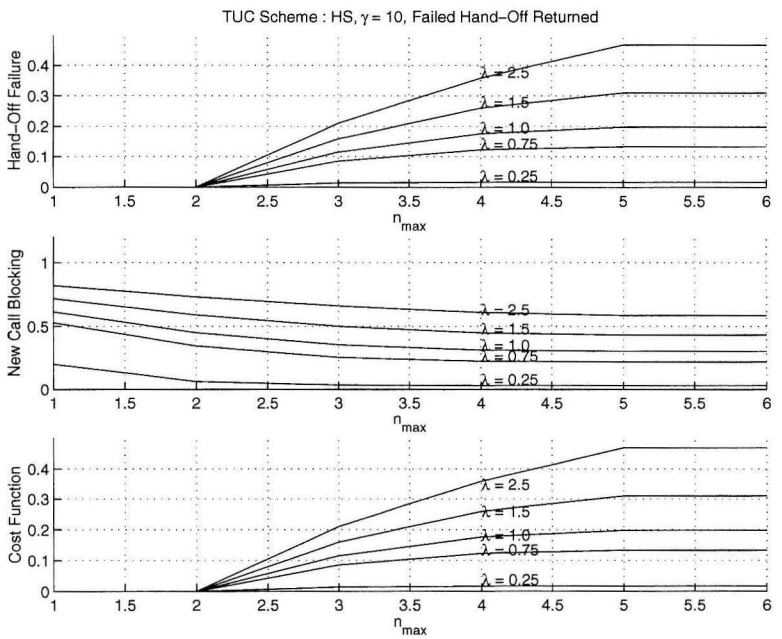

Figure 5.16: TUC: b_h and b_n (failed hand-off returned, $\gamma = 10$, HS)

Hot Spot Traffic

Figure 5.16 plots b_h, b_n, and cost for the high mobility ($\gamma = 10$) case with failed hand-off calls returned to the system. Since the behavior of the low mobility case is similar, its plots are not shown here. Table 5.5 lists the parameters of the CFC scheme (T, τ) and the TUC scheme ($n_{\max}$) for which their performance (b_h and b_n) is the same.

5.10 Conclusions

The new call access control criteria affect the quality of service perceived by users in the cellular system. We measure this quality of service by the hand-off failure probability and the new call blocking probabiilty. In general, it is better to block a new call than to drop

an existing call that requests a hand-off. This should be taken into account by the channel assignment scheme.

We investigated two criteria for modifying the new call access criteria in unprioritized channel assignment schemes to prioritize the acceptance of hand-off calls. In these criteria, the set of accept states, with respect to the cell in which the new call request occurs, is a function of the user mobility rates in the cells. The MLS criterion adapts to non-uniform mobility traffic by assigning different sets of accept states to cells with different traffic parameters. The look-ahead parameter τ needs to be chosen for implementing the required trade-off between the hand-off failure probability and the new call blocking probability. High weighting factors like $\alpha = 10/11$ in the cost function, which penalize rejecting hand-off requests much more than rejecting new call requests, lead to conservative new call acceptance criteria. The CFC scheme offers a way of directly incorporating the cost function that is to be minimized into the design of the accept state space. This scheme can be easily modified to incorporate more complex cost functions which accommodate, for example, different user priorities.

The increase in the number of possible future states of the system, which makes the calculation of the maximum likelihood state or the likely future cost function computationally expensive, is a drawback of both MLS and CFC schemes.

Appendix

5.A Steady State Probability Evaluation

Given the set of accept states G_i for a CAS, the transition rate from one system state to another is as given below. The transition rate depends on whether a failed hand-off call request is dropped from the system or is just returned to the system.

Failed hand-off returned For states $\underline{n}, \underline{s} \in \Omega$ $(\underline{n} \neq \underline{s})$, the transition rate from $\underline{n}$ to $\underline{s}$ is given by

$$R(\underline{s},\underline{n}) = \begin{cases} \lambda_i, & \text{if } \underline{s} = \underline{n} + \underline{e}_i,\ \underline{n} \in G_i \text{ for some } i \\ n(i)\gamma_i p(i,j), & \text{if } \underline{s} = \underline{n} - \underline{e}_i + \underline{e}_j \text{ for some } i,j\ (i \neq j) \\ n(i)\mu_i, & \text{if } \underline{s} = \underline{n} - \underline{e}_i \text{ for some } i \\ 0, & \text{otherwise} \end{cases} . \quad (5.27)$$

Failed hand-off dropped For states $\underline{n}, \underline{s} \in \Omega$ $(\underline{n} \neq \underline{s})$, the transition rate takes the form

$$R(\underline{s},\underline{n}) = \begin{cases} \lambda_i, & \text{if } \underline{s} = \underline{n} + \underline{e}_i,\ \underline{n} \in G_i \text{ for some } i \\ n(i)\gamma_i p(i,j), & \text{if } \underline{s} = \underline{n} - \underline{e}_i + \underline{e}_j \text{ for some } i \neq j \\ n(i)\mu_i + D_i(\underline{n}), & \text{if } \underline{s} = \underline{n} - \underline{e}_i \text{ for some } i \\ 0, & \text{otherwise} \end{cases} , \quad (5.28)$$

where $D_i(\underline{n})$ is an additional transition rate from $\underline{n}$ to $\underline{n} - \underline{e}_i$ due to rejected hand-off calls getting dropped from the system. It is given by

$$D_i(\underline{n}) = \sum_{k=1}^{C} n(i)\gamma_i p(i,k) I\left[\underline{n} - \underline{e}_i + \underline{e}_k \notin \Omega\right] . \quad (5.29)$$

The indicator function $I\left[.\right]$ has been defined earlier in (5.24).

The steady-state flow equations for a state $\underline{n} \in \Omega$ are then given by

$$\begin{aligned} \sum_{\underline{s} \in \Omega \setminus \{\underline{n}\}} \pi(\underline{s}) R(\underline{n}, \underline{s}) &= \pi(\underline{n}) \sum_{\underline{s} \in \Omega \setminus \{\underline{n}\}} R(\underline{s}, \underline{n}), \\ \sum_{\underline{s} \in \Omega} \pi(\underline{s}) &= 1, \end{aligned} \tag{5.30}$$

where $\pi(\underline{n})$ is the steady state probability of state $\underline{n}$. Solving the above linear equations yields $\pi(\underline{n})$, for all $\underline{n} \in \Omega$.

5.B Total User Count Scheme Analysis

Closed-form solutions for the hand-off failure and new call blocking probabilities exist for the uniform traffic case, as we show below. These are obtained by using modified A and $\underline{c}$ matrices.

The TUC scheme does not accept a new call if

$$\sum_{i=1}^{C} n(i) > n_{\max}. \tag{5.31}$$

Since hand-off call requests from one cell to the other do not increase the total number of mobiles in the system, the above criterion is satisfied automatically by the state reached by the system after a hand-off call request is accepted or rejected. Thus the same criterion can be used for both hand-off and new call requests. The TUC scheme can thus be described completely by means of augmented matrices A' and $\underline{c}'$ which contain (5.31) in addition to the CAS constraints given in (5.1). The new call blocking and hand-off failure probabilities can then be obtained by applying the formulae in (5.25) as follows:

$$\begin{aligned} b_h &= 1 - \frac{G(A', \underline{c}' - A'\underline{e}_1)}{G(A', \min(\underline{c}' - A'\underline{e}_1, \underline{c}' - A'\underline{e}_2))}, \\ b_n &= 1 - \frac{G(A', \underline{c}')}{G(A', \underline{c}' - A'\underline{e}_1)}, \end{aligned} \tag{5.32}$$

where $G(A', \underline{c}')$ is the probability of the system being in the all-zero state $\underline{n} = \underline{0}$.

5.C Simulations

Simulations were carried out for the Markov model for traffic in the cellular system described in Section 5.7. The system, which executes a first order continuous time Markov process, remains in a state $\underline{n}$ for a length of time that has an exponential distribution with parameter $q(\underline{n}) = \sum_{\underline{s} \in \Omega \setminus \{\underline{n}\}} R(\underline{s}, \underline{n})$. Here, $q(\underline{n})$ is the net rate at which the system departs from its present state. Given that the system leaves state $\underline{n}$, its next state will be $\underline{s}$ with probability

$$p(\underline{s}|\underline{n}) = R(\underline{s}, \underline{n})/q(\underline{n}). \tag{5.33}$$

Since only the time instants where a new call request, a call termination, or a hand-off call request occur are of interest, the continuous time Markov process can be converted into a discrete time Markov chain using the Jump Chain concept [94]. The simulation results so obtained match, as expected, with the results obtained from the numerical analysis described in Section 5.7.2.

Chapter 6 Conclusions

Future generations of wireless cellular mobile systems are being designed to facilitate a range of voice and data services while remaining highly spectrally efficient. The ability to provide mobile communication services is an important reason for the popularity and bright outlook for wireless systems today. However, user mobility also gives rise to a host of interesting technical challenges that need to be tackled. It must be realized that the impact of mobility is a function of the type of resource allocation scheme under consideration. This dissertation has focussed on analyzing the impact of user mobility on different resource allocation schemes. In particular, we have looked at the following issues.

In Chapter 3 we studied a link adaptation scheme in which a user chooses the modulation and error correction coding for transmitting his packet(s) based on the SIR estimate of the wireless link between him and his serving BS. A user may even choose not to transmit in case he estimates that the channel condition is below an acceptable threshold. The performance of such schemes is dependent on the adaptation thresholds and the accuracy of the estimate itself. Link estimate accuracy is a function of the user speed and the delay in feeding back estimates. For the scenario in which the users' queues are stable, we derived expressions for the average packet waiting time in terms of the basic system parameters like packet arrival statistics, channel fading statistics, adaptation thresholds, etc. We also analyzed a simpler saturated queues scenario in which users always have packets to transmit. For this scenario, we derived expressions for the system throughput, measured in terms of the number of packets successfully transmitted per slot, as a function of the given basic system parameters. The results show that the optimum SIR threshold below which a user should not transmit is a function of the channel correlation for the stable queues scenario. While the no-transmission mode succeeded in reducing the average packet delay for a channel with high correlation, it provided no such gains for a channel with low correlation. This is not so for the saturated queues scenario where the optimum SIR thresholds were found to be insensitive to the channel correlation and the Nakagami fading parameter. An analy-

sis of link adaptation schemes which can take into account unequal scenarios where some users face more favorable link conditions than others is an interesting and untackled problem. The issue of ensuring fairness among different users when the no-transmission mode is used in such unequal scenarios is another pertinent issue.

In Chapter 4, we looked at a multiple access scheme where the base station does not play as active a role as in link adaptation. In particular, we looked at the Packet Reservation Multiple Access (PRMA) protocol which was proposed to simultaneously handle periodic, delay intolerant (voice) traffic as well as aperiodic, delay tolerant (data) traffic. We developed an approximate analytical technique, based on signal flow graphs, to compute the voice packet dropping probability as a function of the user mobility rate (measured in terms of the cell hopping probability). We also used the technique to analyze the effect of packet errors on PRMA. This approximate analysis can be applied to the case where the error rates are fixed a priori as well as to the case where the error rates and the co-channel interference are inter-dependent. We showed that contention, and not user mobility, is the main reason for dropping voice packets. On the other hand, packet errors do have a significant impact on the system performance. The signal flow graph based technique avoids many of the approximations made in previous approaches, and was found to match well with the simulation results. Analysis of the general case where users are randomly distributed over the entire cell area, and thereby encounter significantly different channels, is an open research problem. Another interesting issue to study is when link adaptation, in which terminals use assigned channels for transmitting data and feed-back information, is combined with PRMA (or another access protocol), which assigns the channel(s) to users.

In Chapter 5, we investigated new call access control techniques that utilize statistical information about user mobility to introduce hand-off prioritization in unprioritized channel assignment schemes. We analyzed two heuristic prediction based criteria: cost function criterion and maximum likelihood state criterion, in which the new call access control is based on estimates of the future system configuration. Using mobility based statistics enables them to handle geographically non-uniform mobility traffic. The complexity of using the entire system state, even in its simplest form where it is completely determined by the number of users in each cell, is the main drawback of these criteria. An explicit characteri-

zation of the optimal call admission control criteria remains an open problem for investigation. Integrating call access criteria with the resource allocation schemes discussed in the previous chapters is another potentially rewarding problem.

Bibliography

[1] H. Holma and A. Toskala, *WCDMA for UMTS*. John Wiley, 2000.

[2] D. C. Cox, "Wireless personal communications: What is it?," *IEEE Pers. Commun.*, pp. 20–35, Apr. 1995.

[3] A. J. Goldsmith, *Lecture Notes on Wireless Communications*. California Institute of Technology, 1998.

[4] B. Crow, I. Widjaja, L. Kim, and P. Sakai, "IEEE 802.11 wireless local area networks," *IEEE Commun. Mag.*, pp. 116–126, Sep. 1997.

[5] A. C. Gummalla, *Wireless Home Networks: Architecture and Access Protocols*. PhD thesis, Georgia Institute of Technology, Atlanta, Georgia, U.S.A., Jul. 2000.

[6] J. C. Haarsten, "The bluetooth radio system," *IEEE Pers. Commun.*, vol. 71, pp. 28–36, Feb. 2000.

[7] T. Ojanperä and R. Prasad, *Wideband CDMA for third generation mobile communications*. Artech House Universal Personal Communications Series, 1998.

[8] D. N. Knisely, S. Kumar, S. Laha, and S. Nanda, "Evolution of wireless data services: IS-95 to cdma2000," *IEEE Commun. Mag.*, pp. 140–149, Oct. 1998.

[9] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *IEEE Pers. Commun.*, pp. 10–31, Jun. 1996.

[10] A. S. Tanenbaum, *Computer Networks*. Prentice Hall, 1995.

[11] C. Perkins, "Mobile IP," *IEEE Commun. Mag.*, pp. 84–99, May 1997.

[12] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi, "Packet reservation multiple access for local wireless communications," *IEEE Trans. Commun.*, vol. 37, pp. 885–890, Aug. 1989.

[13] J. Padgett, C. Gunther, and T. Hattori, "Overview of wireless personal communications," *IEEE Commun. Mag.*, pp. 28–41, Jan. 1995.

[14] M. Zeng, A. Annamalai, and V. K. Bhargava, "Recent advances in cellular wireless communications," *IEEE Commun. Mag.*, pp. 128–138, Sep. 1999.

[15] N. R. Sollenberger, N. Seshadri, and R. Cox, "The evolution of IS-136 TDMA for third-generation wireless services," *IEEE Pers. Commun.*, pp. 8–18, Jun. 1999.

[16] M. Mouly and M.-B. Pautet, "Current evolution of the GSM systems," *IEEE Pers. Commun.*, vol. 25, pp. 9–19, Oct. 1995.

[17] P. Dupuis, "A European view on the transition path toward advanced mobile systems," *IEEE Pers. Commun.*, pp. 60–63, Feb. 1995.

[18] A. Furuskär, S. Mazur, F. Müller, and H. Olofsson, "EDGE: Enhanced data rates for GSM and TDMA/136 evolution," *IEEE Pers. Commun.*, vol. 63, pp. 56–66, Jun. 1999.

[19] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Commun. Mag.*, pp. 70–77, Jul. 2000.

[20] T. S. Rappaport, *Wireless Communications: principles and practice.* Prentice Hall, 1996.

[21] A. J. Goldsmith and L. Greenstein, "A measurement-based model for predicting coverage areas of urban microcells," *IEEE Trans. Veh. Technol.*, vol. 117, pp. 1013–1023, Sep. 1993.

[22] G. L. Stüber, *Principles of Mobile Communications.* Kluwer Academic Publishers, 1996.

[23] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*. Academic Press, 4th ed., 1980.

[24] M. Nakagami, *Statistical Methods in Radio Wave Propagation*. Pergamon Press, Oxford, U.K., 1960.

[25] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 3rd ed., 1991.

[26] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 51, pp. 10–22, Feb. 1998.

[27] W. C. Jakes, *Microwave Mobile Communications*. IEEE Press, 1974.

[28] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Lett.*, vol. 27, pp. 2145–2146, Nov. 1991.

[29] R. Narasimhan, *Estimation of Mobile Speed and Average Received Power with Application to Corner Detection and Handoff*. PhD thesis, Stanford University, Stanford, California, U.S.A., Nov. 1999.

[30] R. Guérin, *Queueing and Traffic in Cellular Radio*. PhD thesis, California Institute of Technology, Pasadena, California, U.S.A., May 1986.

[31] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Technol.*, vol. 35, pp. 77–92, Aug. 1986.

[32] C. Jedrzycki and V. C. M. Leung, "Probability distribution of channel holding time in cellular telephony systems," in *Proc. VTC*, pp. 247–251, 1996.

[33] F. Barceló and J. Jordán, "Channel holding time distribution in cellular telephony," *Electron. Lett.*, vol. 34, pp. 146–147, 1998.

[34] P. V. Orlik and S. S. Rappaport, "A model for teletraffic performance and channel holding time characterization in wireless cellular communication with general session and dwell time distributions," *IEEE J. Selec. Areas Commun.*, vol. 16, pp. 788–803, Jun. 1998.

[35] M. Sidi and D. Starobinski, "New call blocking versus handoff blocking in cellular networks," *Wireless Networks*, pp. 29–41, 1997.

[36] S.-H. Oh and D.-W. Tcha, "Prioritized channel assignment in a cellular radio network," *IEEE Trans. Commun.*, vol. 40, pp. 1259–1269, Jul. 1992.

[37] D. L. Pallant and P. G. Taylor, "Modeling handovers in cellular mobile networks with dynamic channel allocation," *Operations Research*, vol. 43, pp. 33–42, 1995.

[38] G. J. Foschini, B. Gopinath, and Z. Miljanic, "Channel cost of mobility," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 414–424, Nov. 1993.

[39] M. W. Oliphant, "Radio interfaces make the difference in 3G cellular systems," *IEEE Spectrum*, pp. 53–58, Oct. 2000.

[40] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *IEEE Commun. Mag.*, vol. 38, pp. 54–64, Jan. 2000.

[41] S. Papavassiliou, L. Tassiulas, and P. Tandon, "Meeting QOS requirements in a cellular network with reuse partitioning," *IEEE J. Selec. Areas Commun.*, pp. 1389–1399, Oct. 1994.

[42] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2nd ed., 1989.

[43] A. H. Aghvami, "Digital modulation techniques for mobile and personal communication systems," *Electron. & Commun. Eng. J.*, pp. 125–132, Jun. 1993.

[44] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Prentice Hall, 1995.

[45] R. van Nobelen, N. Seshadri, J. Whitehead, and S. Timiri, "An adaptive radio link protocol with enhanced data rates for GSM evolution," *IEEE Commun. Mag.*, vol. 61, pp. 54–64, Feb. 1999.

[46] S. Kallel, "Analysis of a type II hybrid ARQ scheme with code combining," *IEEE Trans. Commun.*, vol. 38, pp. 1133–1137, Aug. 1990.

[47] J. Pons and J. Dunlop, "In-service link quality estimation for link adaptation algorithms, applied to GSM," in *Proc. ICUPC*, pp. 1169–1173, 1998.

[48] J. Dunlop, J. Irvine, and P. Cosimini, "Estimation of the performance of link adaptation in mobile radio," in *Proc. VTC*, pp. 326–330, 1995.

[49] X. Qiu and J. Chuang, "Link adaptation in wireless data networks for throughput maximization under retransmissions," in *Proc. ICC*, 1999.

[50] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Th.*, vol. 43, pp. 1986–1992, Nov. 1997.

[51] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power mqam for fading channels," *IEEE Trans. Commun.*, pp. 1218–1230, Oct. 1997.

[52] S. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *To appear in IEEE Trans. Commun.*, 2001.

[53] M.-S. Alouini, *Adaptive and Diversity Techniques for Wireless Digital Communications over Fading Channels*. PhD thesis, California Institute of Technology, Pasadena, California, U.S.A., 1998.

[54] J. C.-I. Chuang, "Improvement of data throughput in wireless packet systems with link adaptation and efficient frequency reuse," in *Proc. VTC*, pp. 821–825, 1999.

[55] L. Kleinrock, *Queueing Systems*, vol. I. Wiley, 1975.

[56] M. Frullone, G. Falciasecca, P. Grazioso, G. Riva, and A. M. Serra, "On the performance of packet reservation multiple access with fixed and dynamic channel allocation," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 78–85, Feb. 1993.

[57] X. Qiu and V. O. K. Li, "A unified performance model for reservation-type multiple-access schemes," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 173–189, 1998.

[58] G. Wu, K. Mukumoto, and A. Fukuda, "An integrated voice and data transmission system with idle signal multiple access – static analysis," *IEICE Trans. Commun.*, vol. E76-B, pp. 1186–1192, Sep. 1993.

[59] G. Wu, K. Mukumoto, and A. Fukuda, "An integrated voice and data transmission system with idle signal multiple access – dynamic analysis," *IEICE Trans. Commun.*, vol. E76-B, pp. 1398–1407, Nov. 1993.

[60] N. D. Wilson, R. Ganesh, K. Joseph, and D. Raychaudhuri, "Packet CDMA versus dynamic TDMA for multiple access in an integrated voice/data PCN," *IEEE J. Selec. Areas Commun.*, vol. 11, pp. 870–884, 1993.

[61] N. Amitay and L. J. Greenstein, "Resource auction multiple access (RAMA) in the cellular environment," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 1101–1111, 1994.

[62] G. Bianchi, F. Borgonovo, L. Fratta, L. Musumeci, and M. Zorzi, "C-PRMA: A centralized packet reservation multiple access for local wireless communications," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 422–435, May 1997.

[63] L. Wang, J. Wu, and A. H. Aghvami, "Performance of CDMA/PRMA with adaptive permission probability control in packet radio networks," in *Proc. VTC*, pp. 2044–2048, 1999.

[64] P. Taaghol, R. Tafazolli, and B. G. Evans, "Performance optimization of packet reservation multiple access through novel permission probability strategies," in *IEE Colloquium on Networking Aspects of Radio Communication Systems*, pp. 8/1–8/5, 1996.

[65] H. Qi and R. Wyrwas, "Capture effects on PRMA stability and performance," *Electron. Lett.*, vol. 30, pp. 539–540, 1994.

[66] X. Qiu and V. O. K. Li, "On the capacity of packet reservation multiple access with capture in personal communication systems," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 666–675, 1996.

[67] D. J. Goodman and S. X. Wei, "Efficiency of packet reservation multiple access," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 170–176, Feb. 1991.

[68] W.-C. Wong, "Packet reservation multiple access in a metropolitan microcellular radio environment," *IEEE J. Selec. Areas Commun.*, vol. 11, pp. 918–925, Aug. 1993.

[69] M. Frullone, G. Riva, P. Grazioso, and C. Carcioffi, "PRMA performance in cellular environments with self-adaptive channel allocation strategies," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 657–665, Nov. 1996.

[70] S. Nanda, D. J. Goodman, and U. Timor, "Performance of PRMA: A packet voice protocol for cellular systems," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 584–598, Aug. 1991.

[71] H. Qi and R. Wyrwas, "Performance analysis of joint voice-data PRMA over random packet error channels," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 332–345, May 1996.

[72] C. Wang, J. Wang, and B. Sukhbaatar, "Performance evaluation of PRMA system with channel fading," in *Proc. ICC*, pp. 1952–1954, 1993.

[73] P. T. Brady, "A model for generating on-off speech patterns in two-way conversations," *Bell Syst. Tech. J.*, pp. 2445–2472, Sep. 1969.

[74] R. W. Muise, T. J. Schonfeld, and G. H. Zimmerman, III, "Experiments in wideband packet technology," in *Proc. 1986 Zurich Sem.*, pp. 135–139, Mar. 1986.

[75] S. Mason and H. Zimmerman, *Electronic Circuits, Signals, and Systems.* Wiley, 1960.

[76] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing.* Prentice Hall, 1989.

[77] S. Tasaka, *Performance analysis of multiple access protocols.* MIT Press series in computer systems, 1986.

[78] A. Fukuda, "Equilibrium point analysis of ALOHA-type systems," *Trans. IECE Japan*, vol. J61-B, pp. 959–966, Nov. 1978.

[79] D. Grillo, M. Frullone, P. Grazioso, and P. Valigi, "A performance analysis of PRMA considering speech/data traffic, co-channel interference and ARQ error recovery," in *Seventh IEE European Conf. on Mobile and Personal Commun.*, Dec. 1993.

[80] H.-S. Wang and L.-S. Poon, "Rayleigh fading effects on packet reservation multiple access," in *Proc. VTC*, pp. 1360–1363, 1994.

[81] L. Hanzo, J. Cheung, R. Steele, and W. Webb, "Performance of prma schemes via fading channels," in *Proc. VTC*, pp. 913–916, 1993.

[82] M. K. Simon and D. Divsalar, "Some new twists to problems involving the gaussian probability integral," *IEEE Trans. Commun.*, vol. 46, pp. 200–210, Feb. 1998.

[83] S. K. Deora, *Channel Assignment Algorithms in Cellular Radio Networks.* PhD thesis, California Institute of Technology, Pasadena, California, U.S.A., May 1995.

[84] S. Tekinay and B. Jabbari, "Handover and channel assignment in mobile cellular networks," *IEEE Commun. Mag.*, pp. 42–46, Nov. 1991.

[85] R. Singh, S. M. Elnoubi, and C. Gupta, "A new frequency channel assignment algorithm in high capacity mobile communications systems," *IEEE Trans. Veh. Technol.*, vol. 31, pp. 125–131, 1982.

[86] M. Zhang and T.-S. P. Yum, "Comparisons of channel-assignment strategies in cellular mobile telephone systems," *IEEE Trans. Veh. Technol.*, vol. 38, pp. 211–215, 1989.

[87] D. Everitt and N. Macfayden, "Analysis of multicellular mobile radio telephone systems with loss," *British Telecom Technol. J.*, vol. 1, pp. 37–45, 1983.

[88] P.-A. Raymond, "Performance analysis of cellular networks," *IEEE Trans. Commun.*, vol. 39, pp. 1787–1793, Dec. 1991.

[89] Y.-B. Lin, A. R. Noerpel, and D. J. Harasty, "The sub-rating channel assignment strategy for PCS hand-offs," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 122–130, Feb. 1996.

[90] M. Naghshineh and M. Schwartz, "Distributed call admission control in mobile/wireless networks," *IEEE J. Selec. Areas Commun.*, vol. 14, pp. 711–717, May 1996.

[91] R. Ramjee, D. Towsley, and R. Nagarajan, "On optimal call admission control in cellular networks," *Wireless Networks*, vol. 3, pp. 29–41, 1997.

[92] S. Jordan and A. Khan, "A performance bound on dynamic channel allocation in cellular systems: Equal load," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 333–344, 1994.

[93] F. P. Kelly, "Blocking probabilities in large circuit-switched networks," *Adv. Appl. Prob.*, vol. 18, pp. 473–505, 1986.

[94] F. P. Kelly, *Reversibility and Stochastic Networks*. John Wiley, 1987.

[95] N. B. Mehta and A. Goldsmith, "Prediction-based techniques for hand-off prioritization in channel assignment schemes," in *Proc. Globecom*, pp. 2599–2604, 1998.

[96] N. B. Mehta and A. Goldsmith, "Effect of mobility on PRMA," in *Proc. ICC*, pp. 1110–1114, 1999.

[97] N. B. Mehta and A. Goldsmith, "Effect of fixed and interference-induced packet error probability on PRMA," in *Proc. ICC*, pp. 362–366, 2000.

[98] N. B. Mehta and A. J. Goldsmith, "Performance analysis of link adaptation in wireless data networks," in *Proc. Globecom*, 2000.

[99] N. B. Mehta and A. J. Goldsmith, "Throughput analysis of link adaptation in interference-limited cellular systems," To be presented at VTC (Spring), 2001.

[100] N. B. Mehta and A. J. Goldsmith, "Effect of mobility on PRMA," *Accepted for publication in IEEE Trans. Commun.*