ABSTRACT

Title of Dissertation: SCHEDULING IN PACKET SWITCHED CELLULAR WIRELESS SYSTEMS

Roshni Srinivasan, Doctor of Philosophy, 2004

Dissertation directed by: Professor John S. Baras
Department of Electrical and Computer Engineering

In cellular wireless networks where users have independent fading channels, throughput for delay tolerant applications has been greatly increased on the downlink by using opportunistic schedulers at the base station. These schedulers exploit the multiuser diversity inherent in cellular systems. An interesting question is how opportunistic schedulers will provide Quality of Service(QoS) guarantees for a mix of data traffic and traffic from delay-sensitive multimedia applications.

In the first part of this dissertation, we completely characterize the scheduled rate, delay and packet service times experienced by mobile users in a packet switched cellular wireless system in terms of a configurable base station scheduler metric. The metric used has a general form, combining an estimate of a mobile

user's channel quality with the scheduling delay experienced by the user. In addition to quantifying the scheduler performance, our analysis highlights the inherent trade-off between system throughput and the delay experienced by mobile users with opportunistic scheduling. We also use this analysis to study the effect of prioritized voice users on data users in a cellular wireless system with delay constrained opportunistic scheduling. Our statistical analysis of the forward link is validated by extensive simulations of a system architecture based on the CDMA 1xEV-DO system.

The increase in data traffic from mobiles to the base station has led to a growing interest in a *scheduled* reverse link in the 1xEV-DO system. We address the reverse link scheduling problem in a multi-cell scenario with interference constraints both within and outside the cell. This approach leads to a co-operative scheduling algorithm where each base station in a cellular network maximizes the sum of mobile data transmission rates subject to linear constraints on (1) the maximum received power for individual mobiles (2) the total interference caused by scheduled mobiles to (a) traffic and control channels of other mobiles within the cell and (b) mobiles in neighboring cells. Simulations of the reverse link structure based on the 1xEV-D0 system highlight the distinct advantages of this algorithm in ensuring predictable inter-cell interference and higher aggregate cell throughputs.

SCHEDULING IN PACKET SWITCHED CELLULAR WIRELESS SYSTEMS

by

Roshni Srinivasan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:

Professor John S. Baras, Chair
Professor Eric V. Slud
Professor Armand M. Makowski
Assistant Professor Richard J. La
Assistant Professor Sennur Ulukus

DEDICATION

In loving memory of my parents,

N.K. Prabhakara Rao and Savitha Rao

ACKNOWLEDGEMENTS

I am grateful to John Baras, my thesis advisor, for his support throughout my graduate studies. I have gained immensely from my discussions with him and my fellow graduate students in his group. I look forward to continued collaboration with him in the future.

I would like to thank the other members of my committee, Eric Slud, Armand Makowski, Richard La and Sennur Ulukus for their feedback and advice. Of the many outstanding teachers who have influenced and shaped my research at the University of Maryland, I wish to acknowledge Eric Slud, Adrian Papamarcou, Armand Makowski, Prakash Narayan and Abram Kagan.

My discussions on service differentiation with Hari Balakrishnan and Suchitra Raman exposed me to a number of interesting problems in the early part of my dissertation research. I thank Arnab Das, Sundeep Rangan, Tom Richardson and Shirish Nagaraj for reviewing my work and providing insightful comments. Their suggestions have greatly improved the quality of this dissertation.

Special thanks to Althia Kirlew for her patience and timely assistance with all my requests. I would also like to acknowledge Maria Hoo, Elisabeth

El-Khodary, Dorothy Chu, Diane Hicks and Tina Vigil for their advice and assistance with departmental procedures.

I have the deepest gratitude to my parents. My father's insatiable curiosity, incorrigible optimism and work ethic have inspired me. I could always count on my mother for her reassuring support and encouragement in difficult times. I cherish the memories of their love and affection for me.

My parents-in-law have stood by me through the ups and downs of this endeavor and never let me lose sight of my goals. My sisters, Anu and Suman, sister-in-law, Kamini and their families have been a great source of strength to me. I owe a special debt of gratitude to my family.

I would now like to thank the persons dearest to my heart. My husband, Murari has always believed in my abilities. We have had countless intellectually stimulating discussions on various aspects of my research that improved my understanding of the issues that were relevant to my thesis. The completion of this dissertation is a milestone that could not have been reached without his unconditional support and confidence in me. Finally, I would like to thank our son, Siddharth and our daughter, Leela for helping me discover the joys of learning and simple thinking once again. I will always treasure the wonderful times we share together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Cellular wireless technologies have traditionally been optimized for voice traffic. The first generation of cellular wireless systems like AMPS were built around analog transmission techniques such as frequency modulation (FM). In the early 1990's, second generation cellular systems such as CDMA IS-95 and GSM that employ digital communication technologies were designed and deployed. These systems are essentially circuit switched in nature, with airlink resources allocated to mobiles for the duration of a voice call. Circuit switched air-interfaces are typically inefficient for data delivery. Technologies that provide broadband data services in third generation (3G) wireless systems use a combination of circuit switching and packet switching for data transmission. The trend in evolving 3G systems is clearly towards a completely packet switched air-interface. A variety of mobile terminals such as cellphones, laptops, PDAs are beginning to use the data services available on these systems. The management of airlink resources such as power, spectrum among mobile terminals is critical to the performance of 3G systems. A packet switched air-interface facilitates efficient use of the airlink

resources by enabling scheduling mechanisms that arbitrate between the multiple mobiles served by a single base station. As in the case of schedulers for packet switched wireline systems, airlink schedulers are also required to provide high aggregate throughput and support for Quality of Service (QoS).

The quality of a wireless channel, typically measured in terms of the Signal-to-Interference+Noise-Ratio (SINR or SNR) fluctuates in time due to fading. Traditionally, in point-to-point wireless communications, fading has been viewed as a phenomenon that must be overcome in order to increase throughput. The large dynamic range of channel attenuations in a typical fading process requires complex resource allocation mechanisms that adapt transmission power and coding rate to the variations in channel quality. In recent years, the conventional view of fading has been turned around completely in the context of a multiuser point-to-multipoint system. It has been shown that the independent fading processes of multiple users can be exploited to achieve *multiuser diversity* [26, 51, 20] gain by scheduling users with relatively good channel quality at any given time instant. The concept of multiuser diversity has given rise a new class of schedulers in cellular wireless systems that are frequently referred to as *opportunistic* schedulers.

The Maximum SNR scheduler best illustrates the idea behind opportunistic schedulers. In a time-slotted system where mobiles constantly report channel quality to the transmitter, this scheduler maximizes system throughput on the

forward link by transmitting to the mobile with the best channel in every time slot. 3G systems such as the Enhanced Data rates for Global Evolution (EDGE) [18] extension of GSM/GPRS in Enhanced General Packet Radio Service (EGPRS) [39] or 1xEV(IS-856) [5, 49, 15, 61] in cdma2000 family of standards also give priority to mobiles with better channels by utilizing the forward link channel state information reported by mobiles. In the case of the aximum SNR scheduler, channel variations and low SNR may cause arbitrarily long scheduling delays for mobiles with poor channel conditions . It is easy to see that the gains in system throughput may come at the cost of unfair resource allocation and variability in the scheduled rate and delay. In order to support traffic with a wide range of QoS requirements, these schedulers must incorporate delay constraints.

On the forward link in 1xEV-DO [5, 49, 15, 61], the transmitter at the base station adapts the *rate* of the scheduled mobile to the fading channel environment by utilizing the channel state information provided by the mobile. In every time slot, the base station transmits at full power P to the mobile with the highest scheduling metric. Since only one mobile is scheduled, there is no interference to other mobiles within the cell. Furthermore, since the tranmission power is fixed, interference to other mobiles in neighboring cells on the forward link is predictable. Unlike the forward link, the reverse link traffic in current 1xEV-DO systems consists of circuit switched data channels and forward control channels. The transmission *power* of the data channels is adapted through a

distributed rate control algorithm. This algorithm probablistically varies the transmission power and therefore the rate in response to measured interference. This approach is well suited for low and medium rate data applications. However, file uploads and increasingly popular multimedia applications now require higher rates on the reverse link. This has led to a growing interest in a scheduled reverse link. A centralized scheduling algorithm that schedules mobiles with favorable channels to transmit at higher powers would undoubtedly increase data throughput and hence the aggregate system throughput. However, careful consideration has to be given to the resulting interference, which depends on the transmission power as well as the location of the scheduled mobiles. Within the same cell, high levels of interference from scheduled mobiles can degrade the performance for other cicuit switched voice users and reverse link control channels. Scheduled mobiles transmitting at very high power at the edge of a cell can cause interference in neighboring cells as well. It is therefore important to control interference while maximizing throughput through scheduling on the reverse link. We now motivate the problems studied in this thesis.

1.1 Motivation

1.1.1 Opportunistic Scheduling Delays on the Forward Link

In a typical cell, mobiles near the base station have better channel conditions than mobiles at the periphery of a cell. Since opportunistic schedulers typically favor mobiles with better channels, fading only exacerbates the unfairness in resource allocation to mobiles at the periphery. An opportunistic scheduler that is *rate fair* equalizes rates between mobiles by allocating a larger fraction of time slots to mobiles with weak channels. A *resource fair* scheduler on the other hand can ensure fairness by allocating the same number of time slots to all users at the expense of throughput. In both cases, opportunistic scheduling can large fluctuations in the scheduling delay. We now illustrate the trade-off between aggregate forward link throughput and scheduling delay through the following scheduler metric, $m(t)$ that combines multiuser diversity gain with delay constraints:

$$m(t) = R(t) + \alpha \frac{v(t)}{N} = R(t) + \alpha V(t), \tag{1.1}$$

For convenience, we use the word *user* interchangeably with mobile in the rest of this dissertation. $R(t)$ is the rate requested by the user at the beginning of time slot $[t, t+1)$, $v(t)$ is the delay since a waiting packet in the user's queue was previously served and α is a configurable control weight. The scheduling delay $v(t)$ at the beginning of time slot $[t, t+1)$, normalized by the number of users is

represented by $V(t)$. In the case of opportunistic schedulers, waiting packets could be delayed because the scheduler is serving other users. Such delays could be caused either because the scheduler is serving other users with better channel conditions or for reasons of fairness. We therefore refer to the normalized delay, $V(t)$ as scheduler *vacation time* in the rest of this dissertation. As the number of users increases, $v(t)$ increases proportionally. Using the normalized version of $v(t)$ ensures that the number of users does not affect the balance between multiuser diversity gain and delay in the metric.

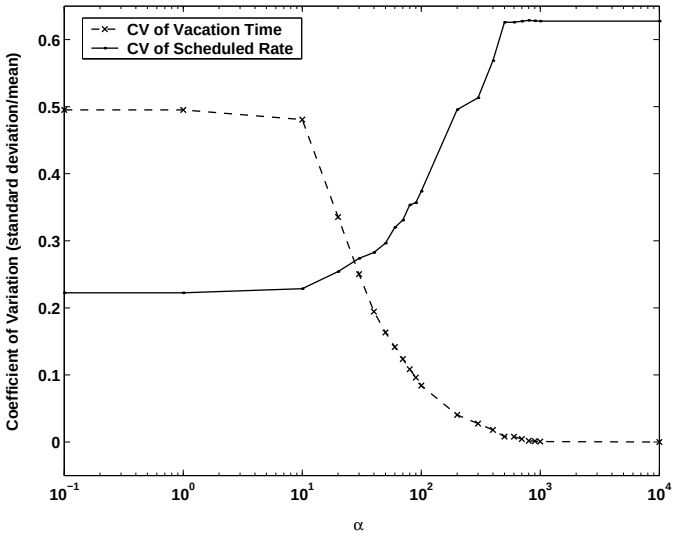

Figure 1.1: CV of scheduled rate and CV of vacation time vs α for 16 mobiles at nominal SNR of 2.5dB

In Figure 1.1, we plot the coefficient of variation (ratio of standard deviation to the mean) of the scheduled rate and the coefficient of variation (CV) of the delay between scheduling slots as a function of α. The scenario we consider is the

forward link of a cellular wireless system similar to the 1xEV-DO data system. In such a system, a base station serves N users in a cell using a time-slotted forward link combined with an asynchronous circuit switched reverse link. In this example, we consider 16 mobiles, each with a nominal SNR of 2.5dB. We assume a Rayleigh SNR distribution for the flat fading channel experienced by every mobile. The maximum sustainable rate, which is a function of the channel quality is constantly reported back to the base station by every mobile via a dedicated, circuit switched channel on the reverse link. These *requested rates* are independent of each other, independent across time slots and identically distributed. In every time slot, the scheduler at the base station computes a metric for each mobile as given by equation 1.1. The base station then transmits to the mobile with the highest metric.

We see from Figure 1.1 that for small values of α, the scheduler described above behaves like the Maximum SNR scheduler. Not only is the mean scheduled rate is large since the scheduler always serves the user with the best channel, but the standard deviation of the scheduled rate is also very low. The CV of the scheduled rate is therefore the lowest for the Maximum SNR scheduler. On the other hand, the scheduling delays are given no priority, causing large variations in the vacation time experienced by users and therefore a high coefficent of variation. As α is increased, the contribution of the delay towards the scheduling metric increases. For large α , the scheduler is channel agnostic. Users are served

cyclically, once in every N slots as in Longest Wait First (LWF) scheduling. The LWF scheduler, naturally, has the lowest CV of vacation time. The scheduled rate exhibits the opposite trend. We measure the vacation time in slots and the scheduled rate in Link Layer (LL) segments per slot. Table 3.1 lists the transmitted rate in link-layer segments as a function of the SNR. For 8 byte segments and a slot duration of 1.667ms, the maximum rate corresponds to 64 segments per slot. Although the dynamic ranges of the rate and normalized vacation time in the metric proposed in equation 1.1 are not of the same scale, normalizing the scheduled rate by the maximum rate of 64 segments resolves the issue of dimensionality. Figure 1.2 demonstrates that this normalization simply corresponds to a scaling in the range of α over which the scheduled rate and vacation time statistics vary.

1.1.2 Data Scheduling with Prioritized Voice

The support of real-time services and in particular, the integration of voice over IP (VoIP) and data traffic in packet switched 3G cellular wireless systems is currently a topic of active research. In 1xEV-DO, if 20 ms speech frames generated by VoIP codecs such as G.729 are to be delivered with minimal delay, 12 active users can be supported every 20ms. Assuming a voice activity factor of 0.5, the system can support roughly 24 voice users without jitter. A higher number of voice users can be served by either using higer compression rates, or

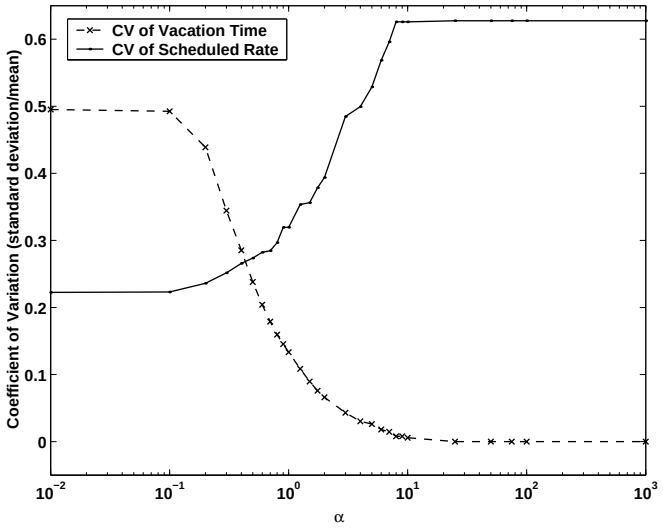

Figure 1.2: CV of scheduled rate and CV of vacation time vs α for 16 mobiles at nominal SNR of 2.5dB with scaled rate

by tolerating a larger amount of scheduling delay, both of which can adversely affect voice quality. The simplest technique to support delay-sensitive traffic such as packet voice in a packet switched cellular data system is to strictly prioritize it over data traffic. 1xEV-DO, for instance, supports QoS by prioritizing delay-sensitive data in the wireline backhaul network as well as over the airlink [63]. The residual bandwidth (time slots) available for data applications can be utilized most efficiently by exploiting multiuser diversity techniques. Naturally, this scheme could decrease throughput and increase delay for the data traffic in the cell. In order to quantify this impact, consider Figure 1.3, which illustrates the effect of admitting prioritized voice users on the packet call throughputs of data users. This figure plots the cumulative density function of

the effective throughput experienced by a "data call", which emulates the download of a standard 5KB web page. We model our evaluation of the performance of data users on standard evaluation techniques used in 3GPP [14]. The CDF is obtained by using Monte Carlo simulations to average the effective throughputs of the data call for a number of low-mobility users distributed under different channel conditions in an interference-constrained cell.

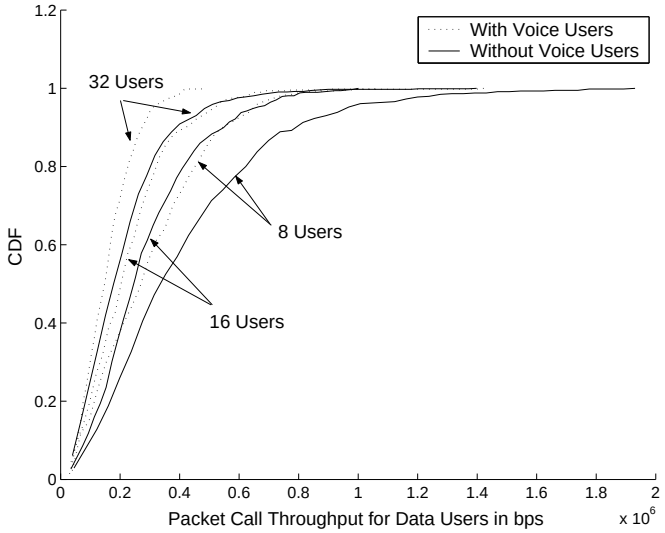

Figure 1.3: Impact of 25% voice calls on packet call throughputs of data users

We consider a system in which a base station serves N_v voice users and N_d data users using a combination of a time-slotted forward link and an asynchronous circuit switched reverse link. We assume a Rayleigh SNR distribution for the flat fading channel experienced by every data user. The Cumulative Distribution Function (CDF) of the correlated channel rates is

generated using the Jakes model [25]. In every time slot, the scheduler at the base station computes a metric for each user as given by equation 1.1. If there is no waiting voice packet, the base station transmits to the data user with the highest metric. Multiuser diversity gains are maximized for the data traffic by setting $\alpha = 0$ in the metric defined in Equation 1.1.

In our simulations, we allocate 25% of the total voice capacity of the system to data. Observe from Figure 1.3 that when this fraction of time slots becomes unavailable to the data users, the CDF's naturally shift to lower ranges of throughput. We plot the packet call throughputs in bits per second with and without voice traffic when 8, 16 and 32 data users share the available bandwidth in a time slotted system . We see that when voice users are multiplexed with data users, the performance that the system offers 8 data users is similar to that obtained by 16 users in the absence of voice. The level of degradation of the user experience depends not only on the fraction of time slots used by voice users but also on the number of data users sharing the unused voice bandwidth and the type of data traffic. Observe that the data throughput is maximized since multiuser diversity is exploted fully. If low latency data applications that require delay constraints in the metric have to be supported, throughput degradation will be even worse.

The integration of wireless telephony and data services in wireless systems that use packet switched air interfaces therefore poses new challenges in the

management of system resources. An analytical evaluation of the performance for data users in the presence of prioritized voice traffic could provide important tools for cellular network operators to quantify system performance and provision resources for traffic with varying QoS requirements.

1.1.3 Interference Constrained Reverse Link Scheduling

System architectures for data services in the cdma2000 1xEV (IS-856) [61, 62] standard were initially built on the assymetric link model. The forward link from the base station transceiver to the mobile was designed to be a fast forward path for data transfer at high rates. Traffic on the reverse link was expected to consist mainly of ACK packets and control information. 1xEV therefore provided lower data rates on the slower reverse path. The 1xEV standard uses a distributed rate control algorithm [10] to adjust transmission rates on the reverse link. If the total interference measured at the base station exceeds(drops below) a threshold, the base station sets(resets) the Reverse Activity Bit (RAB) to indicate that mobiles should probabilistically decrease(increase) their reverse link transmission rates. While such a distributed algorithm works well for low rate, circuit switched traffic, more precise control can be achieved for burstier, packet switched data traffic using centralized scheduling algorithm. The increase in reverse link traffic from mobiles that upload images, facilitate web browsing and large file transfers have led to provisions for a scheduled reverse link and QoS support in the

standards for the forthcoming Release A of 1xEV-DO.

Scheduling in the 1xEV reverse link requires power allocation algorithms that maximize the overall reverse link capacity while controling interference. In addition to managing the interference caused to other mobiles within a cell, interference caused by mobiles at the edge of the cell to neighboring cells is an important design consideration. This is an especially challenging issue when QoS considerations influence the choice of users to be scheduled. For example, a user with high transmission power at the edge of a cell could potentially transmit at a high rate if it is given priority over interior users. In order to control the interference caused to the reverse link channels of mobiles in neighboring sectors, the scheduler needs to intelligently constrain the transmit power of such users.

1.2 Contributions and Organization of Thesis

Scheduling algorithms for the forward link in cellular wireless systems have been well studied in the literature. Although scheduling on the reverse link has been given less attention, it is gaining importance with the increase in uplink traffic and the demand for service differentiation. We present related work in forward and reverse link scheduling in Chapter 2. In this chapter, we also describe the evolution of the packet switched air-interfaces for cellular wireless systems and describe system details that are relevant to the assumptions made in the analytical and simulation models used in this dissertation.

In addition to the design of an efficient scheduling algorithm, the evaluation of its performance in achieving the desired throughput or delay statistics is of great importance to a network operator. The results presented in Chapter 3 of this dissertation completely characterize the scheduled rate and delay experienced by mobile users in a typical cell in terms of a configurable scheduler metric. The metric used has a general form, combining the requested rate, which is a measure of the mobile's channel quality with the scheduling delay experienced by the mobile. Furthermore, the analysis highlights the inherent trade-off between system throughput and the delay experienced by mobile users with opportunistic scheduling. The results in this chapter address the important issue of providing QoS suport in cellular wireless systems. Quantifying the performance at the MAC layer also benefits higher layer protocols such as TCP which are impacted by rate and delay fluctuations [45, 11]. This cross-layer networking approach can further enhance system performance.

In Chapter 4, we focus on the impact of supporting VoIP services over a time-slotted packet switched air interface. The system is assumed to give strict priority to voice traffic, while data packets are opportunistically scheduled subject to delay constraints. We extend the analysis in Chapter 3 to study the effect of prioritized voice users on data users in a cellular wireless system with delay constrained opportunistic scheduling. We quantify the resulting delay and compute the packet service times for data users as a function of the number of

voice users in the system when all users have fully loaded queues and the resources of the system are completely utilized.

In Chapter 5, we propose and evaluate a co-operative reverse link scheduling algorithm in which every cell maximizes its total throughput while limiting the interference to neighboring cells to a globally known threshold. This strategy limits the probability of outages caused by mobiles transmitting at high power at the edge of a neighboring cell and reduces unpredictable out-of-cell interference. The proposed scheduling algorithm constructs a greedy solution to the reverse link throughput maximization problem with contraints on mobile power, total in-cell interference as well as out-of-cell interference. We study the performance of this algorithm and validate our analysis through simulations of a reverse link based on the IS-856 (cdma2000 1xEV) standard.

Finally, in Chapter 6, we present a summary of our work. We conclude this dissertation with a discussion of design considerations for QoS support in 3G cellular wireless systems and outline some directions for future work.

Chapter 2

Background and Related Work

2.1 3G : The Evolution of Packet Switched Systems

The new market dynamics created by mobility and the growth of the internet have led to the definition of technical requirements and standards for third generation (3G) cellular systems that provide higher voice capacity and data rates. The International Mobile Telecommunications IMT-2000 [23] standards approved by the International Telecommunication Union (ITU) in collaboration with industry bodies from around the world require that 3G networks deliver improved system capacity and spectral efficiency over second generation (2G) systems. The third generation partnership project, 3GPP2 [48] was formed to prepare, approve and maintain globally applicable technical sepcifications for 3G mobile systems. 3GPP2 also takes into account emerging ITU recommendations in the IMT-2000 initiative.

CDMA technology, which was first introduced in the IS-95 standard included several enhancements in the standards that followed. IS-95A/B and related

standards that formed the 2G cdmaOne suite evolved into cdma2000. The ITU approved cdma2000, also known as IMT-CDMA Multi Carrier as one of the five standard 3G radio interfaces.

The narrowband 1x standard in cdma2000 has significantly improved voice and data capacity using a single 1.25MHz carrier. 1x supports circuit switched voice and is also backward compatible with cdmaOne systems. The 1x mode supports voice and medium data speeds of up to 153kbps. The recently deployed 1xEV packet data standard is the *EV*olution of cdma2000 beyond the 1x standard. 1X Evolution Data Optimized (1xEV-DO) provides higher data rates on 1X systems. 1xEV-DO requires a separate carrier for data and is capable of delivering peak rates of 2MBps. This carrier will be able to hand-off to a 1X carrier if simultaneous voice and data services are needed. The technical specification for 1xEV-D0 is standardized by the Technical Specification Group - C (TSG-C) of 3GPP2 as IS-856 [9]. 1x Evolution Data and Voice (1xEV-DV) is the next phase of the 1x evolution. A single 1xEV-DV carrier will be used to carry high speed data and voice simultaneously through a packet switched network. In this dissertation, we focus on scheduling issues relating to QoS support in 1xEV-DO systems, both on the forward link as well as the reverse link.

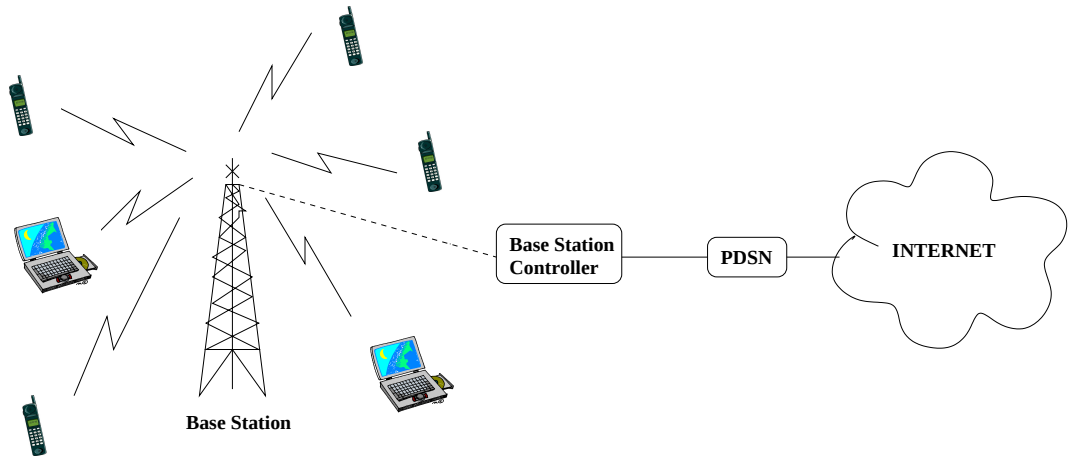

Figure 2.1: System architecture

2.2 System Architecture

In the 1xEV-DO system architecture [64] shown in Figure 2.1, each cell site comprises of a base station that communicates with mobile access terminals. The base station may be sectorized with a separate antenna for each sector in order to increase capacity. On the reverse path, data packets from the mobile are sent to the base station (BTS - Base Station Transceiver) over the airlink. Multiple base stations are connected to the Base Station Controller(BSC) through a packetized backhaul which is either wired or wireless. Packet data is handled by the Packet Control Function (PCF) which is integrated in the BSC. The PCF forwards data packets to the Packet Data Servicing Node (PDSN) which performs traditional Network Access Server (NAS) functionality. The PDSN then forwards data to the provider's IP network. On the forward path, data packets sent from the PDSN

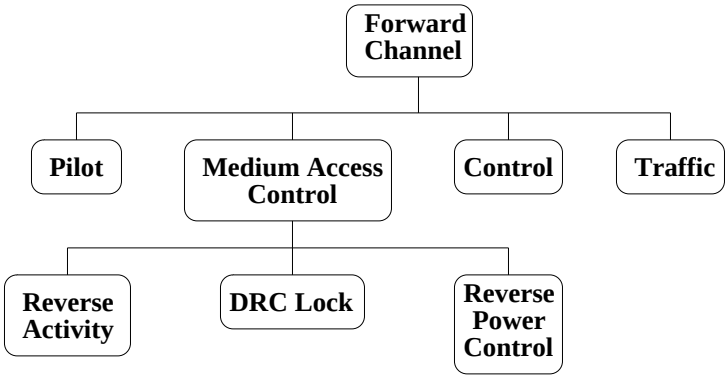

Figure 2.2: 1xEV-DO forward link structure

are forwarded to the base station by the BSC. Packets are then transmitted on the forward airlink from the base station to the destination mobile.

2.3 Channel Structure

The forward and reverse link models in the analysis and simulation in this dissertation are based on the 1xEV-DO standard. We now briefly describe relevant details of the 1xEV-DO link structure [49]. The 1xEV forward link consists of four time multiplexed channels as shown in the block diagram in Figure 2.2. User data is transmitted on the Traffic channel in a slotted manner, with each slot extending over 1.67 ms. Using the pilot signal that is transmitted periodically as reference, every mobile computes its SINR, the ratio of the signal strength to the sum of thermal noise and interference. The mobile then requests the maximum sustainable rate corresponding to the measured SINR using the

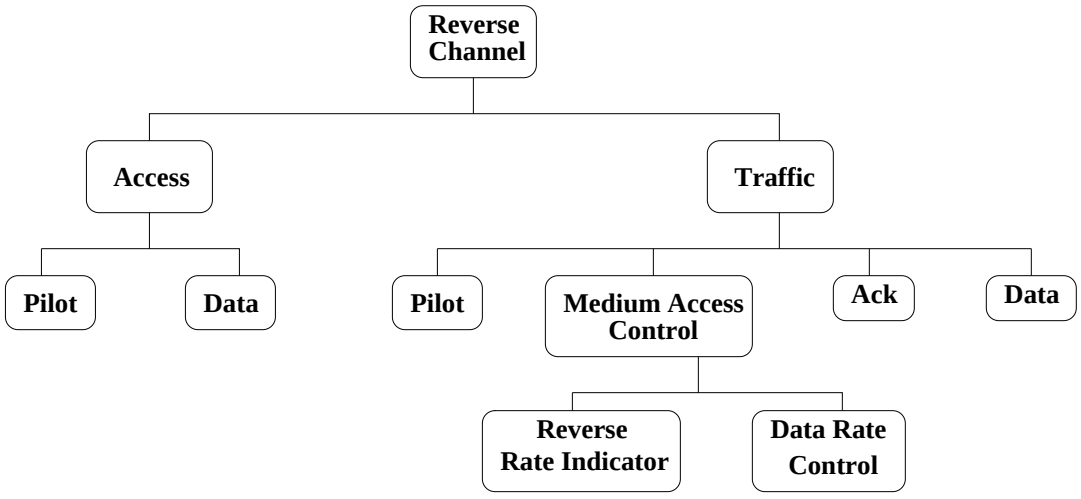

Figure 2.3: 1xEV-DO reverse link structure

reverse link Data Rate Control(DRC) channel. The airlink scheduler at the base station uses the requested rate to compute a scheduing metric for the mobile and selects a single user for transmission in very time slot. The base station transmits at full power to the scheduled mobile at the rate requested by the mobile.

The reverse channel structure is shown in Figure 2.3. The reverse link in 1xEV-DO is essentially identical to the 1xRTT reverse link. The mobile uses the Access channel on the reverse link [49] to communicate with the base station. Signaling information and user data is transmitted on the traffic channel which consists of the Pilot, Medium Access Control, Ack and Data channels. Unlike the forward link, the traffic channel on the reverse link is circuit switched. Hence, several users simultaneously transmit to the base station on their respective data channels. The Reverse Rate Indicator(RRI) is used to indicate the rate at which

reverse link data was sent to the base station. The base station controls the reverse link transmssion rate of the aggregate group of mobiles in a distributed manner using the Reverse Activity Bit (RAB). 1xEV-DO employs a fast ARQ mechanism on the forward link. Therefore, feedback corresponding to the forward link data transmission is transmitted to the base station on the ACK channel in the reverse link.

2.4 Related Work

2.4.1 The Forward Link

2.4.1.1 Fair Queueing

Schedulers for wireless networks were first designed along the lines of their fair queueing analogs in wireline systems. Packet Fair Queueing (PFQ) algorithms that approximate the idealized Generalized Processor Sharing (GPS) [13, 37] policy were extended to wireless networks. The Wireless Fluid Fair Queueing (WFFQ) and the Idealized Wireless Fair Queueing (IWFQ) scheduling algorithm proposed in [32] augment GPS by giving preference to sessions without location dependent wireless channel errors. In addition to the throughput and delay guarantees provided to error-free sessions, the Channel-condition Independent Packet Fair Queueing (CIF-Q) Algorithm proposed in [34] provides fairness guarantees to both error-free and error sessions. The authors in [38] address the

same issues in [34], but provide a different solution to provide long term fairness through the Server Based Fairness Approach (SBFA). Although these PFQ algorithms account for bursty channel errors and location dependent errors, the emphasis is on ensuring fair sharing of the channel resource. Multiuser diversity or link adaptation are not used to improve aggregate throughput.

2.4.1.2 Adapting to Channel Variations in Wireless LANs

Link layer protocols in wireless LANs used multiple retransmissions to recover from bursty packet errors in wirless LANs. However, retransmissions were mostly unsuccessful when the channel went into a burst error state, leading to poor utilization of the channel. The Head of Line blocking effect from FIFO scheduling in such a situation worsened the problem. The Channel State Dependent Packet (CSDP) scheduler in [8] addressed these issues and significantly improved channel utilization in typical wireless LAN configurations. The authors in [17] utilized channel state information in a class based queueing approach to ensure fairness among WLAN users.

2.4.1.3 Multiuser Diversity Gain : The Proportional Fair Scheduler

Traditional approaches to compensate for fading in wireless channels include power control, frequency hopping or interleaving. Knopp and Humblet [26] were among the first to recognize the multiuser diversity gains realizable in a cellular wireless system. They showed that in a single cell, where multiple mobile users

transmit to a base station, the total reverse link information-theoretic capacity is maximized by allowing the user with the best channel to utilize the common channel resource. On the forward link , a similar scheduling strategy at the base station has been shown to be maximize throughput [51]. Multiuser diversity gain grows with the number of users in the system since there is a greater likelihood of scheduling users at peak channel conditions.

The Proportional Fair scheduler used in 1xEV-DO [52] is the first application of an opportunistic scheduler in a commercial wireless system. The trade-off between system throughput and scheduling delay can also be seen in the metric used to select a user in the PF scheduler [52]:

$$m(t) = \frac{R(t)}{T(t)}, \tag{2.1}$$

where $T(t)$ is an estimate of the user's average MAC layer throughput in some window of time prior to the current instant. If i^* denotes the user with the highest metric, then $T(t)$ can be obtained by exponentially averaging the ith user's throughput over a scheduling time scale t_c i.e.,

$$\begin{aligned} T_i(t+1) = &\quad (1-\tfrac{1}{t_c})T_i(t) + (\tfrac{1}{t_c})R_i(t), \qquad i = i^* \\ = &\quad (1-\tfrac{1}{t_c})T_i(t) \qquad\qquad\quad i \neq i^* \end{aligned} \tag{2.2}$$

The parameter t_c is tied to the latency time scale and the QoS requirement of the application. If the latency time scale is large then the scheduler has more flexibility in scheduling the user, possibly waiting much longer until the user's

channel is at a very high peak. While the scheduled rate is bound to be higher, the variations in scheduling delay are larger. When $t_c \rightarrow \infty$, the long-term average throughput of each user is shown to exist and the algorithm is shown to maximize $\sum_{i=1}^{N} logT_i$ among the class of all schedulers. The PF scheduler therefore provides an implicit mechanism to increase aggregate cell throughput at the MAC layer at the expense of increased variability in scheduling delay jitter.

2.4.1.4 Providing QoS : Throughput and Delay Guarantees

In recent years, improving wireless technologies have increased the demand for wireless data services [40]. Real-time applications such as voice, video conferencing and games share network resources with non-real-time traffic such as file transfers and messaging. QoS support for wireless data is therefore a natural consequence of the integration of packet switched wireless networks with the internet, placing new demands on scheduling algorithms. The low delay scheduling algorithm proposed by Bettesh and Shamai [6] was one of the first to address the problem of arbitrarily long delays that could result from opportunistic scheduling. The theoretically complex evaluation of the performance of this algorithm limited its application. The Modified - Largest Weighted Delay First (M-LWDF) [2] rule attempts to optimally provide QoS guarantees in terms of predefined guarantees for the probability of loss and minimum long-term throughput for each user. The throughput optimal

algorithms studied in [2, 3, 4] combine delay and channel conditions in the scheduling metric. A scheduling algorithm is said to be throughput optimal if it is able to keep all queues stable if this is at all feasible to do with any scheduling algorithm. The exponential rule [43, 44] is optimized to share a time-varying channel among multiple real-time users with deadlines. In [1], the authors use a utility maximation formulation to propose a weighted proportional fair scheduler. In [28, 43], admission control and policing govern the queue dynamics and statistical delay guarantees are provided to users based on their current backlog. Unlike this Intserv approach, the authors in [1] use a DiffServ policy to provide service differentiation based on the user the service class and the network resource utilization, which is a function of the channel state.

A framework that provides QoS guarantees has two important elements. The first is an admission control policy that determines the extent to which a user can utilize system resources. The second is a mechanism to evaluate the QoS obtained by an appropriate scheduling algorithm in terms of its control parameters. In effect, this mechanism determines whether the QoS requirements of an admitted user can be supported. In Chapter 3, we address the latter issue by analytically computing the distributions for the scheduled rate and resultant delay as a function of the control parameters in the scheduler metric. The problem is complicated by the time-varying wireless channel capacity as a result of fading. In [30], a general framework for opportunistic scheduling is presented.

Optimal solutions for scheduling problems with temporal and utilitarian fairness requirements are presented. A scheduling algorithm for minimum performance guarantees is also presented. The authors in [60] introduce the notion of effective capacity to address this issue. Using effective capacity for for admission control, they design a scheduling algorithm that combines Round Robin Scheduling with Maximum SNR scheduling to provide QoS guarantees . In our work, we take a new approach and model the system formed by the scheduler at the base station and the users as a dynamical system. In the discrete case, the analysis models a time-slotted system with a finite number of users.

2.4.2 The Reverse Link

The reverse link in IS-95 systems was designed to support low rate voice applications. The analysis of reverse link capacity in [58, 36] is derived from the CDMA principles that form the basis of the IS-95 standard. In the 1xEV standard, higher data rates can be supported on the forward link by incorporating hybrid ARQ, adaptive modulation, channel feedback information in the base station scheduler etc. Many of these features require additional control channels on the reverse link such as a DRC channel, ACK feedback channel, etc. The interference overhead caused by the control channels on the reverse link limits the number of users that can be supported. The capacity of the reverse link in the 1xEV-DO is evaluated in [16]. The analysis in this paper also

evaluates the effect of antenna patterns in achievable gains in capacity.

Research on the 1xEV reverse link can be broadly classified into two areas. The first area of research comprises of distributed rate control arising from related work on power control algorithms for CDMA systems. Research on centralized scheduling algorithms with QoS capabilities is more recent.

2.4.2.1 Distributed Rate Control

In this section, we briefly describe the distributed rate control algorithm [10] that is used by the access network in 1xEV-DO to control the reverse link data traffic rates. We now introduce some notation to describe the transmit/receive power on the pilot and data channels for the ith mobile:

- g_i is the channel gain from the mobile to the base station
- $\overline{p}_i$ is the transmit power on the data channel
- $p_i = g_i\overline{p}_i$ is the corresponding receive power at the base station on the data channel
- $\overline{P}_i$ is the transmit power on the pilot channel
- $P_i = g_i\overline{P}_i$ is the corresponding receive power at the base station on the pilot channel
- The pilot signals of each of the mobiles are power controlled to be received at approximately the same power P_{tgt}. The corresponding transmit power

on the pilot channel alone is $\overline{P}_i = P_{tgt}/g_i$

Let $\mathcal{R} = \{0, R_{min}, 2R_{min}, ..., R_{max}/2, R_{max}\}$. denote the finite set of allowable transmission rates. Consider a sector that has N mobiles connected to it. Let the current data transmission rate for the ith mobile be $r_i \in \mathcal{R}$. The corresponding transmit power is naturally a function of the rate, and may be represented by $\overline{p}_i = \mathcal{F}(r_i)$. It is typical to express the total power transmitted relative to the pilot transmit power, a quantity referred to as Total-to-Pilot transmit power (TtP) ratio.

Every mobile monitors a forward link channel that conveys the *Reverse Activity Bit* (RAB), which is set by the base station. The Combined RAB is the logical OR operation on the most recent RABs from all sectors in the active set. The mobiles independently select their current transmit rate in a distributed manner.

- If the Combined RAB is 0, then the mobile is allowed to probabilistically double its transmission rate, i.e., the new transmission rate, $\tilde{r}_i = min(R_{max}, r_i * 2)$ with probability q_{inc} or $\tilde{r}_i = min(R_{max}, r_i)$ with probability $(1 - q_{inc})$.

- If the Combined RAB is set, then the mobile is required to probabilistically halve its transmission rate, i.e., $\tilde{r}_i = max(R_{min}, r_i/2)$ with probability q_{dec} or $\tilde{r}_i = max(R_{min}, r_i)$ with probability $(1 - q_{dec})$.

- The transition probabilities q_{inc} and q_{dec} are a function of the current rate, r_i. Table 2.1 shows typical set of transition probabilities given in [10].

- If the available power is not enough to support the new rate $\tilde{r}_i$, then the mobile decreases r_i to the maximum rate in $\mathcal{R}$ that can be accomodated by the available transmit power.

- If the amount of data to transmit is less than the payload size of $\tilde{r}_i$, then $\tilde{r}_i$ is decreased to the lowest data rate for which the payload size can be transmitted.

- The mobile transmits at $\tilde{r}_i$.

Table 2.1: Transition probabilities and total-to-pilot power ratios

	Data rate (kbps)					
	0	9.6	19.2	38.4	76.8	153.6
q_{inc}	1	3/16	1/16	1/32	1/32	0
q_{dec}	0	0	1/16	1/16	1/8	1
TtP	2	4.37	6.73	11.44	23.13	72.79

The base station constantly monitors the total energy it receives on the reverse link. The aggregate power of the receive signal is used to calculate a

quantity called the *Rise over Thermal* (RoT), which may be expressed in dB as

$$RoT = 10\ log_{10}(\frac{N_0W + I_c + I_o + \sum_{i=1}^{N} \overline{p}_i g_i}{N_0W}) \tag{2.3}$$

where I_c is the interference generated by circuit switched control channels within the cell and I_o is the out-of-cell interference received at the base station. When the RoT exceeds a certain threshold, the base station sets its RAB bit, which forces all the connected mobiles to start decreasing their transmission rates, and correspondingly the transmit powers. When the interference, and hence the RoT, have been sufficiently reduced, the base station clears its RAB bit. The RoT threshold is chosen to ensure that the reverse link control channels are decoded with high probability.

A subtle aspect of the distributed rate control algorithm is that every mobile is controlled by the RAB bits of all base stations in its active set. Mobiles in the interior of a sector are typically controlled by the RAB of that sector alone, while edge mobiles are controlled by more than one RAB. This ensures that the mobiles on the edge are probabilistically controlled to lower rates, and hence generate lower interference to neighboring cells.

The fairness of the Medium Access Control (MAC) algorithm in the cdma2000 1xEV-DO systems has been investigated through a utililty maximization formulation in [50]. The problem of allocating the total power to pilot (TtP) ratio for the mobiles is formulated as a competitive market problem. In this approach, each mobile consumes a portion of the available interference

power budget. The power budget, which is the margin of the RoT below the threshold is equivalent to a scarce resource in the pricing model. The base station (producer) utilizes the RAB to signal the price of the resource to the mobiles(consumers). In the competitive market, the price is adjusted until the supply equals the demand. The power allocation vector in the equilibrium state is fair and optimal. The authors in [50] identify the utility functions of the distributed closed loop rate control algorithms in [10, 15] as well as the Enhanced Reverse Link Medium Access Control (RL-MAC) algorithm proposed in [31]. The paper provides the sufficient condition for fair rate allocation in an isolated sector where mobiles have fully loaded queues and no power constraints.

2.4.2.2 Reverse Link Scheduling

One of the main problems with distributed rate control is that there is no ability to provide QoS guarantees of any sort on the reverse link. There is no mechanism to explicitly control rates or delays for particular users. Also, there is no way to award high uplink burst rates to specific users who may have short interactive sessions. These limitations have led to a growing interest in the standards bodies in the design of a scheduled reverse link. Algorithms that optimize overall system throughput in a *single cell* subject to constraints on the transmit powers of individual mobiles as well the total power contribution towards interference within the cell have been analyzed in [53, 27, 35, 22]. Consider a sector with N

mobiles, each with a maximum transmit power of $\overline{p}^{max}$. The maximum power that can be received from mobile i is $p_i^{max} = \overline{p}^{max} g_i$. The solution that maximizes the aggregate throughput is obtained by ordering the mobiles in the decreasing order of their maximum receive powers, and then allocating the maximum allowable power to each mobile until the in-cell interference budget is consumed. Without loss of generality, let the mobiles be indexed so that $p_1^{max} \geq p_2^{max} \geq \ldots p_N^{max}$ In the optimal solution, there exists an integer i_T such that the allocated *receive* power, p_i, is the maximum permissible receive power, p_i^{max}, for all mobiles with indices $i < i_T$. Also, mobiles with indices $i > i_T$ are allocated zero power. A solitary mobile with index $i = i_T$ may be allocated receive power that places it an an intermediate point i.e., $0 < p_K < p_K^{max}$. It is straightforward to observe that the mobiles with the best links end up being scheduled and consume the interference budget. The weaker mobiles do not get scheduled unless the interference budget is very large, or if the maximum transmit power on the mobiles is very small. This leads to a highly unfair allocation of resources in the cell, and makes it impossible to provide any QoS guarantees.

The issue of fairness between users at the interior of the cell and the edge of the cell is addressed in [53] by serving users with non-empty queues sequentially. Higher throughputs and significant power savings are obtained from statistical averaging of the optimal solution over a large sample space of experimental user distributions within the cell. In [27], Kumaran and Qian also propose throughput

optimal scheduling algorithms for mobiles in a single cell. The use of weighted throughputs in the objective function addresses (a) the inherent unfairness of algorithms that maximize throughput alone to mobiles with lower path gains and (b) the provision of different Qualities of Service (QoS) through appropriate choice of weights. The optimal schedule in this formulation orders the mobiles in decreasing order of the weighted rates, as against the ordering in terms of maximum receive powers in [35, 22].

A common thread that runs through the existing body of work on reverse-link scheduling is the explicit consideration of in-cell interference alone. The out-of-cell interference (I_o in equation 2.3) is typically assumed to be a constant and unrelated to the in-cell interference budget. In typical cellular deployment scenarios, this is an unrealistic assumption. The reverse link is symmetric in that the amount of interference generated by the mobiles in any one sector to the neighboring sectors is also experienced in turn by that sector itself. Hence, allowing different in-cell interference budgets while assuming a constant out-of-cell interference is unrealistic.

There appear to be two reasons for this assumption. Firstly, much of the work in reverse-link power allocation or scheduling has been motivated by information-theoretic considerations, in which it is not easy to incorporate the effects of a multi-cellular topology. The single sector reverse link has been well studied and characterized in the information theory literature. The second reason

for this assumption is that the reverse link in existing CDMA cellular systems is circuit switched. It is less critical to explicitly consider out-of-cell interference in the case of a circuit switched uplink. Because of a large number of mobiles transmitting simultaneously, the out-of-cell interference generated by mobiles in a sector has a strong statistical correlation to the in-cell interference that is generated. Also, approximately the same amount of interference is experienced by all neighboring sectors given that the population of mobiles is geographically distributed uniformly. The distributed rate-control mechanism driven off of the RAB bit works extremely well in this situation. The situation is very different in the case of a scheduled uplink. Often times, a single mobile or a small number of mobiles may be scheduled high burst rates. The amount of interference generated by the scheduled set can vary widely, ranging from a negligible amount for interior mobiles to a significant amount for an edge mobile. As may be expected, interference can be experienced in very different ways by neighboring sectors depending on the location of the scheduled set. All these factors make it very important to explicitly consider out-of-cell interference in the scheduling metric.

In our work, we study the problem in a co-operative framework where every cell not only maximizes the overall throughput within the cell but also limits the out-of-cell interference caused to neighboring cells by optimally scheduling edge and interior mobiles. In our formulation, the out-of cell interference generated by the set of scheduled mobiles is explicitly considered in the scheduling metric. We

provide the forms of the solution that satisfy the necessary conditions for optimality. We also provide a scheduling algorithm that constructs a greedy solution to the maximization problem. The results are supported by extensive multi-cell CDMA system simulations. Using these experiments, we evaluate and discuss the scenarios in which out-of-cell interference limits capacity gains and to what extent imposing a constraint affects the system performance.

Chapter 3

An Analysis of Delay-Constrained Opportunistic Scheduling

In this chapter, we analyze the system comprising of the delay-constrained opportunistic scheduler proposed in Section 1.1.1 and the time-varying channel conditions of multiple users as a dynamical system. The analysis outlined characterizes the probability distribution of the expected delay experienced between successive scheduling instants as well as the distribution of the scheduled rate. The approach used is to consider the scheduler as a dynamical system and then examine this system in its steady-state.

3.1 The Discrete Case : A Time-slotted System with a Finite Number of users

The analysis in this section rests on the following assumptions. The base station shares the forward link among a finite group of N users with identical channel

statistics. Let $\mathcal{N} = \{0, \ldots, N-1\}$ denote the set of users served by the base station. Each user $i \in \mathcal{N}$ indicates the maximum sustainable rate to the base station on a dedicated channel on the reverse link. Let $R_i(t)$ denote the rate requested by user i at the beginning of time slot $[t, t+1)$. In order to make the analysis more tractable, the channel rates for each user are assumed to be independent from one time slot to the other. Naturally, the requested rates of the users are independent of each other and identically distributed. Let $\mathcal{R} = \{r_0, r_1, \ldots, r_{max}\}$ denote the finite set of rates which can be requested by the users. This set is assumed to have a probability distribution $f_{R_i}(r) = f_R(r) = P(R = r), r \in \mathcal{R}, \forall i \in \mathcal{N}$. The delay experienced by each user $i \in \mathcal{N}$ since it was previously scheduled is $v_i(t)$, with $V_i(t) = v_i(t)/N$ representing the normalized vacation time at the beginning of time slot $[t, t+1)$.

In every time slot $[t, t+1)$, the base station transmits to the user with the highest metric computed from equation 1.1, applying a tie-breaking rule if necessary. The state space formed by considering all possible combinations of the time-varying channel conditions $R_i(t)$ and scheduling delays $V_i(t)$, $i \in \mathcal{N}$ for the N users in the system is very large. A Markov analysis on this state space is naturally very complex.

The objective of our analysis is to derive statistics for the scheduled rate and vacation time seen by the scheduled user in the steady state. Towards this end, we consider a permutation of the user space in which the users are rank-ordered

in every slot according to the delay they have experienced since they were last scheduled. While rank-ordering eliminates information on the sample paths of individual users, it enables us to capture the steady state statistics for the scheduled user in a greatly reduced state space.

Let the vector $\mathbf{U}(t)$ denote the rank-ordering of users at the beginning of time slot $[t, t+1)$.

$$\mathbf{U}(t) = \{u_0(t), u_1(t), ..., u_{N-1}(t)\}, \tag{3.1}$$

where $u_i(t) \in [0, 1, ..., N-1]$. In this space, $u_i(t)$ denotes the original index of the user who is ranked in the i^{th} position at at the beginning of time slot $[t, t+1)$. By definition, this permutation has the property,

$$V_{u_0}(t) \leq V_{u_1}(t) \leq ... \leq V_{u_{N-1}}(t), \tag{3.2}$$

where $V_{u_i}(t)$ is the vacation time seen by the user who is ranked in position i at the beginning of time slot $[t, t+1)$. Naturally, since $u_0(t)$ is the index of the user scheduled in the previous time slot, $V_{u_0}(t) = 1/N$. In the analysis that follows, we refer to the user with rank i in the ordered space $\mathbf{U}(t)$ as the ith user at time t.

At the beginning of time slot $[t, t+1)$, the scheduler selects a user whose rank $S^*(t)$ is given by

$$S^*(t) = \arg\max_i m_{u_i}(t), \quad i \in \mathcal{N} \tag{3.3}$$

In the event that more than one user has the highest metric, $S^*(t)$ is picked with uniform probability from among the users with the highest metric in order to

break the tie. The selection of a user in one slot causes the rank-ordering of the users to change at the beginning of the next slot. At the beginning of time slot $[t+1, t+2)$, the user, $S^*(t)$ that was selected in the previous time slot $[t, t+1)$ moves to position 0 in the rank-ordered space. Since users are arranged in ascending order of their vacation times, all users with rank greater than that of $S^*(t)$ do not change their order in any way, while all users below the rank of $S^*(t)$ increment their rank by one. Specifically, $\mathbf{U}$ evolves over time as:

$$u_i(t+1) = \begin{cases} u_{S^*}(t), & i = 0 \\ u_{i-1}(t), & 0 < i \leq S^*(t) \\ u_i(t), & S^*(t) < i \leq N-1 \end{cases} \tag{3.4}$$

where $u_{S^*}(t)$ is the index of the user who was scheduled in time slot $[t, t+1)$. Correspondingly, the vacation time seen by every user who was not scheduled increases, while the vacation time seen by the scheduled user is reset to the minimum possible value

$$V_{u_i}(t+1) = \begin{cases} \frac{1}{N}, & i = 0 \\ V_{u_{i-1}}(t) + \frac{1}{N}, & 0 < i \leq S^*(t) \\ V_{u_i}(t) + \frac{1}{N}, & S^*(t) < i \leq N-1 \end{cases} \tag{3.5}$$

This framework describes the evolution of a Markov process consisting of the rank-ordered user space, the corresponding channel conditions and scheduling delays.

We define a selection mass function, $\pi_{u_i}(t)$ which represents the probability of

scheduling the ith rank-ordered user u_i at the beginning of time slot $[t, t+1)$.

$$\pi_{u_i}(t) = Pr(S^*(t) = u_i(t)), \quad i \in \mathcal{N} \tag{3.6}$$

with the property, $\sum_{i=0}^{N-1} \pi_{u_i}(t) = 1$. The notion of a selection mass function is better understood through the following examples. First, consider a Longest Wait First (LWF) scheduler, which always serves the queue with the largest vacation time. This corresponds to a choice of $\alpha \to \infty$ in the composite metric in Equation 1.1. In the rank-ordered user space, the probability mass is concentrated entirely at rank $N-1$ in any slot, i.e.,the user with the highest vacation time.

$$\pi_{u_i}(t) = \begin{cases} 0, & i = 0, 1, \ldots, N-2 \\ 1, & i = N-1 \end{cases} \tag{3.7}$$

Contrast this with a Maximum SNR scheduler, which by definition is agnostic to the delay experienced by any user. This scheduler results from a choice of $\alpha = 0$ in Equation 1.1. If the N users have identical channel statistics, they are equally likely to be served, which results in a uniform selection mass function,

$$\pi_{u_i}(t) = \frac{1}{N}, \quad i \in \mathcal{N} \tag{3.8}$$

3.1.1 Computation of the Selection Mass Function

At the beginning of time slot $[t, t+1)$, the user with rank-ordered index i has a vacation time of $V_{u_i}(t)$. If $R_{u_i}(t)$ is the user's requested rate, the metric $m_{u_i}(t)$

for each user $i \in \mathcal{N}$ is given by

$$m_{u_i}(t) = R_{u_i}(t) + \alpha V_{u_i}(t)$$

The probability of selecting the i^{th} user is then given by

$$\begin{aligned}
\pi_{u_i}(t) &= P(m_{u_i}(t) > m_{u_j}(t), \quad \forall j \neq i) + P(u_i \text{ is selected in a tie}) \\
&= P(R_{u_i}(t) + \alpha V_{u_i}(t) > R_{u_j}(t) + \alpha V_{u_j}(t) \quad \forall j \neq i) \\
&\quad + P(u_i \text{ is selected in a tie}) \qquad (3.9)
\end{aligned}$$

We employ a simple tie-breaking rule in the event that more than one user has the highest metric. In such a case, a single user is picked with uniform probability from among the set of users with the highest metric. The computation of the probability of selecting the i^{th} user in the event of a tie is outlined in the Appendix. Since the channel rates are i.i.d. random variables with distribution $f_R(r)$ and the channel variations are independent of the vacation times, the first term in equation 3.9 can be computed as

$$P(m_{u_i}(t) > m_{u_j}(t), \quad \forall j \neq i) = \sum_{r=r_0}^{r_{max}} \prod_{j \neq i} F_R(r + \alpha(V_{u_i}(t) - V_{u_j}(t)))f_R(r)$$

where r_{max} is the maximum rate that can be supported by the user.

3.1.2 Computation of the Expected Vacation Time

The expected vacation time, $\overline{V}_u(t)$ characterizes the normalized delay experienced by the rank-ordered users in the system. In the analysis that follows,

we assume the existence of a selection mass function, $\pi_{u_j}(t)$ which represents the probability of scheduling the jth rank-ordered user, $u_j(t)$, at the beginning of time slot $[t, t+1)$.

Recall from equations 3.4 and 3.5 that in the set of rank-ordered users, $\mathbf{U}$, the selected user $S^*(t)$ at the beginning of time slot $[t, t+1)$ moves to position 0 at the beginning of time slot $[t+1, t+2)$. All users with rank greater than $S^*(t)$ experience an increase in delay, but do not change their rank in any way. Therefore, if $S^*(t) = \arg\max_i m_{u_i}(t)$, the vacation time for these users evolves as

$$V_{u_i}(t+1) = V_{u_i}(t) + \frac{1}{N}, \quad S^*(t) < i \leq N-1 \tag{3.10}$$

As a result of the scheduled user occupying the very first position in the next slot, all users with rank less than $S^*(t)$ experience an increase in delay and also an increase in rank by one.

$$V_{u_i}(t+1) = V_{u_{i-1}}(t) + \frac{1}{N}, \quad 1 \leq i \leq S^*(t) \tag{3.11}$$

Let $\overline{V}_{u_i}(t)$ denote the expected vacation time at the beginning of time slot $[t, t+1)$ for user i. Although the vacation function dynamically changes with time for any given rank-ordered position i, we have empirically observed the convergence of the *expected* vacation time to an equilibrium value. Furthermore, we have also observed the convergence of the the selection mass function in the steady state. In other words, we assume the existence of limiting values

$$\overline{V}_{u_i} \quad = \quad \lim_{t \to \infty} \overline{V}_{u_i}(t)$$

$$\pi_{u_i} = \lim_{t\to\infty} \pi_{u_i}(t) \tag{3.12}$$

Using equations 3.10 and 3.11 and conditioning on the selected user in time slot $[t, t+1)$, $S^*(t)$, we can compute the expected vacation time for the ith rank-ordered user at the beginning of time slot $[t+1, t+2)$ as

$$\begin{aligned}
\overline{V}_{u_i}(t+1) &= E[V_{u_i}(t+1)] \\
&= E[V_{u_i}(t+1)|S^*(t) > i]P(S^*(t) > i) \\
&\quad + E[V_{u_i}(t+1)|S^*(t) \leq i]P(S^*(t) \leq i) \\
&= E[V_{u_{i-1}}(t) + \frac{1}{N}](\sum_{j\geq i} \pi_{u_j}(t)) + E[V_{u_i}(t) + \frac{1}{N}](\sum_{j<i} \pi_{u_j}(t)) \\
&= (\overline{V}_{u_{i-1}}(t) + \frac{1}{N})(\sum_{j\geq i} \pi_{u_j}(t)) + (\overline{V}_{u_i}(t) + \frac{1}{N})(\sum_{j<i} \pi_{u_j}(t))
\end{aligned} \tag{3.13}$$

The expected vacation time in the steady state, $\overline{V}_u$ can be computed recursively from equation 3.13 using the steady state limiting values from equation 3.12 as

$$\overline{V}_{u_i} = \overline{V}_{u_{i-1}} + \frac{1}{N\left(1 - \sum_{j<i} \pi_{u_j}\right)}, \quad i = 1, \ldots N-1 \tag{3.14}$$

with the initial condition $\overline{V}_{u_0} = \frac{1}{N}$.

3.1.3 Worst-case Expected Delay

The upper bound on the normalized expected delay seen by any user can be computed from equation 3.14 as

$$\overline{V}_{max} = \overline{V}_{u_{N-1}} = \frac{1}{N} + \sum_{i=1}^{N-1} \frac{1}{N\left(1 - \sum_{j<i} \pi_{u_j}\right)} \tag{3.15}$$

Observe that the selection mass function completely determines this upper bound. In the case of the LWF scheduler, the selection density function is 1 for the highest rank-ordered user and 0 for all other users. No term in the summation in equation 3.15 is larger than $1/N$. On the other hand, schedulers that use finite α have selection mass functions with non-zero values for rank-ordered users with lower delays. The terms in the summation in equation 3.15 corresponding to every such user will be larger than $1/N$, thereby increasing the maximum expected normalized delay. The LWF scheduler, with $\alpha \to \infty$, therefore has the lowest worst-case normalized expected delay.

3.1.4 Distributions for Scheduled Rate and Vacation Time

The dynamics of the system can also be used to determine the distribution of vacation time at the scheduling instants as well as the scheduled rate. As can be seen from equation 3.5, the vacation time can only take values that are integral multiples of $1/N$. Let $\gamma = \{k/N,\ k = 1, 2, 3, ...\}$ represent the set of values that the vacation time can take.

The probability mass function of the vacation time experienced by the ith user in time slot $[t+1, t+2)$, can be derived by conditioning on the selected user in the previous time slot $[t, t+1)$ as follows:

$$\begin{aligned} P(V_{u_i}(t+1)=\gamma) &= P(V_{u_i}(t)=\gamma | S^*(t) \leq i) P(S^*(t) \leq i) \\ &\quad + P(V_{u_i}(t+1)=\gamma | S^*(t) > i) P(S^*(t) > i) \\ &= P(V_{u_i}(t)=\gamma - \frac{1}{N})(\sum_{j<i} \pi_{u_j}(t)) \\ &\quad + P(V_{u_{i-1}}(t)=\gamma - \frac{1}{N})(\sum_{j \geq i} \pi_{u_j}(t)) \end{aligned} \tag{3.16}$$

If we denote the probability mass function, $P(V_{u_i}(t)=\gamma)$ by $f_i(t,\gamma)$, the pmf of the vacation time can be obtained recursively as

$$f_i(t,\gamma) = f_i(t-1, \gamma - \frac{1}{N})(\sum_{j<i} \pi_{u_j}(t-1)) + f_{i-1}(t-1, \gamma - \frac{1}{N})(\sum_{j \geq i} \pi_{u_j}(t-1))$$

Assuming the existence of the limit

$$f_i(\gamma) = \lim_{t \to \infty} f_i(t,\gamma) \tag{3.17}$$

we obtain the steady state pmf recursion,

$$f_i(\gamma) = f_i(\gamma - \frac{1}{N})(\sum_{j<i} \pi_{u_j}) + f_{i-1}(\gamma - \frac{1}{N})(\sum_{j \geq i} \pi_{u_j}) \tag{3.18}$$

with $f_0(\frac{1}{N}) = 1$.

Let $\mathcal{V}_{S^*}$ denote the random variable representing the vacation time seen by the *scheduled* user. The distribution of $\mathcal{V}_{S^*}$ may be derived as

$$P(\mathcal{V}_{S^*}=\gamma) = \sum_{j=0}^{N-1} P(\mathcal{V}_{S^*}=\gamma | S^*=i) P(S^*=i)$$

$$= \sum_{j=0}^{N-1} P(V_{u_i} = \gamma)\pi_{u_i}$$
$$= \sum_{j=0}^{N-1} f_i(\gamma)\pi_{u_i} \tag{3.19}$$

The pdf of the scheduled rate (which is naturally different from that of the requested rate) may be derived as a function of α as

$$f_{R_{S^*}}(r) = \sum_{i=0}^{N-1} Pr(R_{u_i} = r, i\text{th rank-ordered user is selected}) \tag{3.20}$$
$$= \sum_{i=0}^{N-1} f_R(r)\prod_{j\neq i} F_R(r + \alpha(V_{u_i} - V_{u_j})) + Pr(\text{user } i \text{ is selected in a tie})$$

3.1.5 Packet Service Times

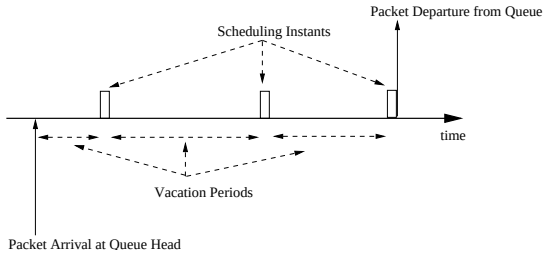

Figure 3.1: Illustration of packet service timeline

Observe from Figure 3.1 that the packet service time may be expressed as the sum of scheduling periods and vacation periods. The packet service time, X, is the sum of X_S, the total number of slots required to transmit all the LL segments corresponding to the packet at the head of the queue and X_V, the number of slots where the scheduler goes on vacation. We assume that packets of fixed

length, L are served by the base station in the order that they arrive. Furthermore, the packet service times are i.i.d. and independent of the interarrival times. Since the scheduling instants are independent of the vacation times, the distribution of service time may be calculated as

$$P(X = n) = \sum_{j} P(X_V = n - j)P(X_S = j) \tag{3.21}$$

If the packet completes service in j time slots, then the data user has experienced exactly j vacation periods between the scheduling instants. We represent this sum of j i.i.d. vacations as $V^{(j)}$. Let $R_{S,k}$ represent the scheduled rate corresponding to the k^{th} slot in the transmission of a given packet. The scheduled rates in different slots are assumed to be i.i.d. random variables. The distribution of the packet service times can therefore be computed from the distributions of the scheduled rate and vacation times.

$$P(X = n) = \sum_{j} P(V^{(j)} = n - j)P(R_S^{(j-1)} < L \leq R_S^{(j)}) \tag{3.22}$$

3.2 System Model and Implementation Issues

In this section we describe the system model and the wireless channel model. We also discuss various aspects of the implementation of the scheduler that uses the metric introduced in Equation 1.1.

3.2.1 System Model

In our simulation experiments, we use a system architecture that is similar to the 3G CDMA wireless data systems such as 1xEV-DO and 1xEV-DV. We combine a time-slotted forward link with an asynchronous cicrcuit switched reverse link. We now outline some important features in the system model.

Table 3.1: Transmission rate per slot as a function of SNR

SNR (in dB)	Rate (Kb/s)
-12.5	38.4
-9.5	76.8
-6.5	153.6
-5.7	204.8
-4	307.2
-1.0	614.4
1.3	921.6
3.0	1228.8
7.2	1843.2
9.5	2457.6

Packet streams for individual users are assigned separate queues by the base

station. Fixed length packets of 512 bytes are segmented into link-layer (LL) segments of 8 bytes for transmission over the air link. At the beginning of each time slot, the scheduler at the base station computes the metric as in equation 1.1 and selects the data user with the highest metric. The number of segments transmitted in a slot depends on the current SNR of the selected data user; this correspondence is enumerated in Table 3.1. The slot duration of 1.667ms and peak rate of 2.45Mbps achievable in this model are similar to the 1xEV-DO system. When all the LL segments corresponding to the packet at the head of the queue for a particular user have been transmitted over the airlink, the packet is dequeued. Transmission errors can be simulated by probabilistically delaying packet transmission. Since we assume that the channel state is known to a high degree of accuracy, we assume a negligible loss probability. Every user is always assumed to have data in the queue. This ensures that the scheduling metric is the sole criterion for selecting a user.

3.2.2 Wireless Channel Model

Under a flat fading assumption, the channel response for every user is assumed to be constant over the duration of the slot. If $x_i(t)$ and $y_i(t)$ denote the vectors of transmitted and received symbols for user i at the beginning of time slot $[t, t+1)$, then

$$y_i(t) = h_i(t)x_i(t) + z_i(t), \quad i = 1, 2, \ldots, N \tag{3.23}$$

where $h_i(t)$ is the time-varying channel response from the base station to the user and $z_i(t)$ is a noise vector with i.i.d., zero mean Gaussian noise components with variance σ_i^2. Assuming unit-energy signals, the nominal SNR for user i is $C_{NOM,i} = \frac{1}{\sigma_i^2}$ with the instantaneous SNR for this user, $C_i(t)$ given by $C_i(t) = \frac{h_i(t)}{\sigma_i^2}$. The probability distribution for the rates requested by the users is generated by using the Jakes [25] model to simulate time-varying channels. The Jakes model uses a sum of K complex exponentials to approximate a single-path Rayleigh fading channel. The complex channel gain at time t is given by

$$h_i(t) = \sum_{j=0}^{K-1} h_{i,j} exp(j2\pi f_d^i t cos(2\pi\phi_j)) \tag{3.24}$$

where $h_{i,j}, j = 0, .., K-1$ are complex, unit variance gaussian random variables with zero mean representing the magnitudes of the subpaths. Each subpath has a phase delay, ϕ_j, which is uniformly distributed in $[0, 2\pi]$. The doppler frequency of the user is given by f_d^i. The Jakes model produces a sequence of attenuation coefficients that is very close to a Rayleigh fading process, and in particular has the same correlation properties.

We study the performance of 16 users ($N = 16$), all with $C_{NOM,i} = 2.5dB$ and a doppler frequency of 10Hz. A scenario with identical channel statistics for all users was selected to enable comparison between analysis and simulation. For the case of i.i.d. channel fades, the distribution of the channel rates is chosen to be identical to the marginal distribution obtained with correlated channel fades at the same nominal SNR.

3.3 Simulation Results

In the analysis in Section 3.1, we assume the existence of steady state limits for the expected vacation time in equation 3.12 and the probability mass function of the vacation time for rank-ordered users in equation 3.17. In this section, we first provide computational results that show the existence of these limits. We then highlight some properties of the expected vacation time that can be verified through simulations. Finally, we present a comparison of analytical and simulation results for the distribution of the scheduled rates and vacation times parametrized by α.

3.3.1 Numerical Convergence

We now present computational results that support the claim in Section 3.1 that the pmf of the vacation time, $f_i(t, \gamma)$, converges to $f_i(\gamma)$ and the expected vacation time $\overline{V}_{u_i}(t)$ converges to $\overline{V}_{u_i}$ as $t \to \infty$. In our simulation experiments, we observe the evolution of the vacation time for 16 users with i.i.d. channel rates and various values of α. We present the results for an arbitrarily chosen value of $\alpha = 10$. In each of the 10,000 iterations of the Monte Carlo simulation, we record the vacation time of all users in every slot. Users are also ranked in increasing order of vacation time in every slot. We compute the probability mass function, $f_i(t, \gamma) = P(V_{u_i}(t) = \gamma)$, for the ith rank-ordered user at the beginning of time slot $[t, t + 1)$ for $\{\gamma = k/N, k = 1, 2, 3, \ldots\}$. We obtain this pmf from the

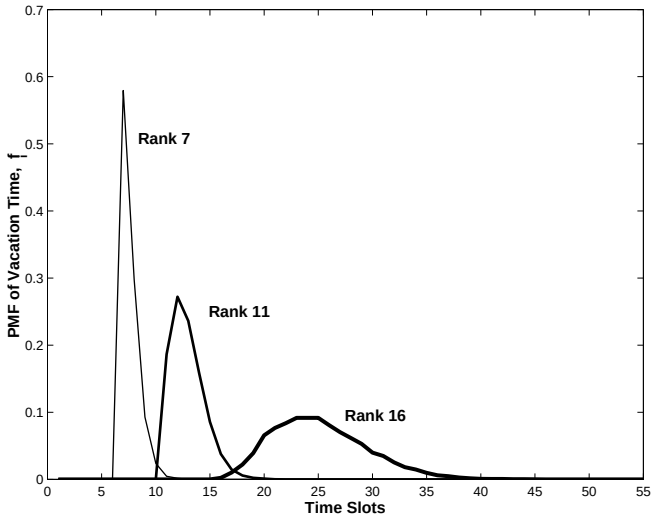

Figure 3.2: Steady state PMF of vacation time for rank-ordered users

vacation times of the *ith* rank-ordered user in the tth slot of each of the 10,000 iterations by computing a normalized histogram across the iterations. Each iteration is run for a duration of 1000 slots.

Figure 3.2 shows the pmf for three arbitrarily chosen rank-ordered users for clarity of exposition. As expected, the pmf corresponding to a higher ranked user in the rank-ordered space is spread over a larger range of values when compared with a lower ranked user. For example, the pmf for the user with rank 16 is spread over a larger range of values as compared to the user with rank 11. This is because the number of users with rank below 16 is larger than that of the number of users below rank 11 and it is possible to select lower ranked users with a non-zero probability if their channel rates dominate the metric. The same relationship can be observed with the pmf's for users with rank 11 and rank 7,

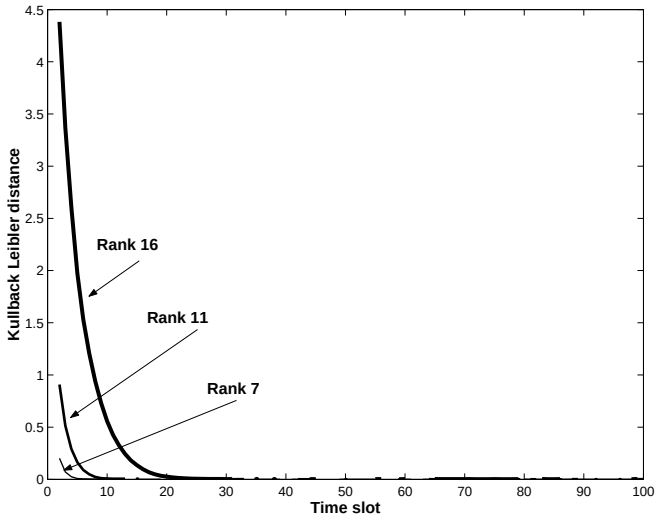

Figure 3.3: Kullback-Leibler distance between steady state PMF and PMF in each iteration for vacation time of rank-ordered users

with the spread of the distribution decreasing with rank. The pmf of the user with the lowest rank has a point mass at $1/N$.

In order to study the convergence of the pmf of the vacation time, $f_i(t, \gamma)$ to $f_i(\gamma)$ as $t \to \infty$, we consider the Kullback Leibler distance [12]. The Kullback Leibler distance, or relative entropy, between two distributions $p(x), x \in \mathcal{X}$ and $q(x), x \in \mathcal{X}$ is defined as

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) log \frac{p(x)}{q(x)} \tag{3.25}$$

Based on continuity arguments, we define $0 log \frac{0}{q} = 0$ and $p log \frac{p}{0} = \infty$. The relative entropy is always non-negative and $D(p||q) = 0$ if and only if $p = q$. Therefore, although it does not satisfy the triangle inequality, it is often used to characterize the *distance* between two probability distributions.

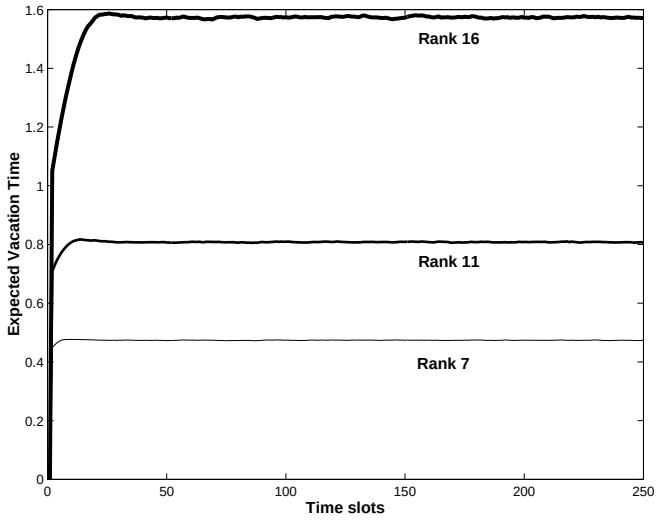

Figure 3.4: Convergence of expected vacation time for rank-ordered users

We compute the Kullback Leibler distance, $d_i(t, t_{max})$ for the ith rank-ordered user as the divergence between $f_i(t, \gamma)$, the pmf of the vacation time at time slot t, and the steady state pmf at the last time slot of the iteration, $f_i(t_{max}, \gamma)$. We define $d_i(t, t_{max})$ as follows:

$$d_i(t, t_{max}) = D(f_i(t, \gamma)||f_i(t_{max}, \gamma)) \tag{3.26}$$

$$= \sum_{\gamma} f_i(t, \gamma) log \frac{f_i(t, \gamma)}{f_i(t_{max}, \gamma)}, \quad \gamma = k/N, k = 1, 2, 3, \ldots \tag{3.27}$$

with the convention that $0 log \frac{0}{f_i(t_{max}, \gamma)} = 0$ and $f_i(t, \gamma) log \frac{f_i(t,\gamma)}{0} = \infty$.

Figure 3.3 shows the Kullback-Leibler distance for the first 100 time slots for users with rank 16, 11 and 7. We see that $d_i(t, t_{max}) \rightarrow 0$ as $t \rightarrow t_{max}$ for all three users, thereby illustrating the convergence of the pmf $f_i(t, \gamma)$ to the steady state value, $f_i(t_{max}, \gamma)$.

We use the pmf of the vacation time for rank-ordered users to compute the expected vacation time for ith rank-ordered user. Observe from figure 3.4, where we plot the expected vacation time for 3 randomly picked rank-ordered users that the expected vacation time also converges to a steady state value. Naturally, the users with lower rank converge to their steady state values faster than users with higher rank.

3.3.2 Worst-Case Expected Delay

The local slope of the expected vacation time at any position i in the rank-ordered user space is given by

$$\Delta\overline{V}_{u_i} = \overline{V}_{u_i} - \overline{V}_{u_{i-1}} \tag{3.28}$$

Observe from equation 3.15 that $\Delta\overline{V}_{u_i}$ is given by $1/\left(N\left(1-\sum_{j<i}\pi_{u_j}\right)\right)$, and is therefore a montonically increasing function of the rank index. The slope at a rank-ordered index i, $\Delta\overline{V}_{u_i}$ is constant $(1/N)$ as long as the probability of scheduling users with lower vacation times is zero i.e., $\pi_{u_j} = 0, \forall j < i$. With increasing i, as the cumulative probability of scheduling, $\pi_{u_j}, \forall j < i$, increases, $\Delta\overline{V}_{u_i}$ increases as well. For every non-zero probability mass in the selection mass function, there is a corresponding increase in the local slope of the expected vacation time. This increase in slope is captured by the summation in the second term of equation 3.15. The interdependence of the expected vacation time and the selection mass function in the proposed scheduling metric gives rise to

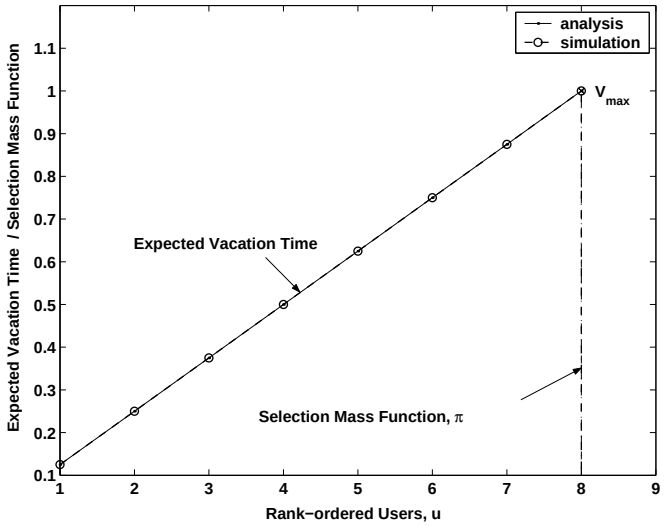

Figure 3.5: LWF scheduler : $\overline{V}$ and π for $\alpha = 1000$

implicit upper and lower bounds for the expected worst-case scheduling delay.

At one extreme, we have the channel agnostic LWF scheduler in which the delay constraint dominates. Fig. 3.5 shows the expected vacation time and the selection mass function for a choice of $\alpha = 1000$ and 8 users. This approximates the behaviour for the LWF scheduler with $\alpha \to \infty$. Using the selection mass function π_{u_i} from the simulations, we compute the expected vacation time from equation 3.14. As expected, we see that the expected vacation time, $\overline{V}_u$ is linear, with slope $= 0.125$ and the maximum normalized delay, $\overline{V}_{max} = 1$ both analytically and from the simulations. The selection mass function has all the probability mass concentrated at 1, since the scheduler always schedules the user with the highest delay.

At the other extreme, we have the Maximum SNR scheduler in which the

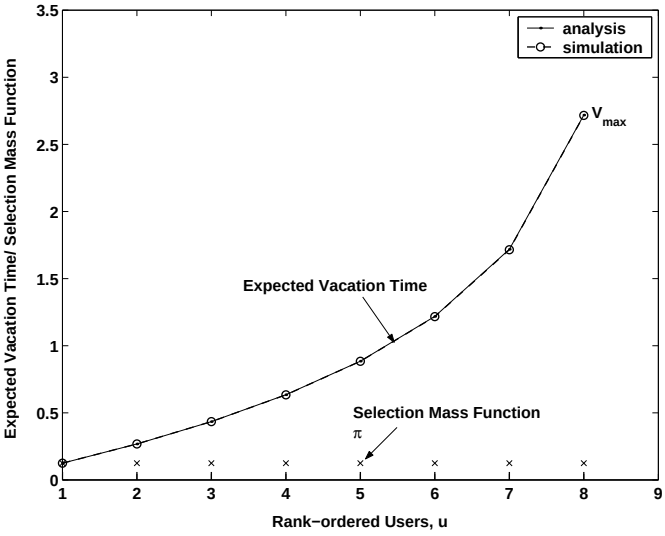

Figure 3.6: Maximum SNR scheduler : $\overline{V}$ and π for $\alpha = 0$

channel conditions dominate the metric with no constraint on the delay. As in Fig. 3.6, the selection mass function is shown to be uniform (0.125). The slope of the expected vacation time increases at every point in the user space. From equation 3.15, the worst case normalized delay is $\overline{V}_{max} = 2.7179$. The simulation results for the mean normalized delay experienced by the rank-ordered users closely match the analysis with $\overline{V}_{max} = 2.7157$.

3.3.3 Distributions for Scheduled Rates and Vacation Time

For the purpose of clarity, we plot the CDFs of the vacation time for the scheduled user and the scheduled rates in Figures 3.7 & 3.8 resp. for only three values of α. The graphs illustrate the close correspondence of the simulation results with the analysis in Section 3.1. In Figure 3.7, we see both from

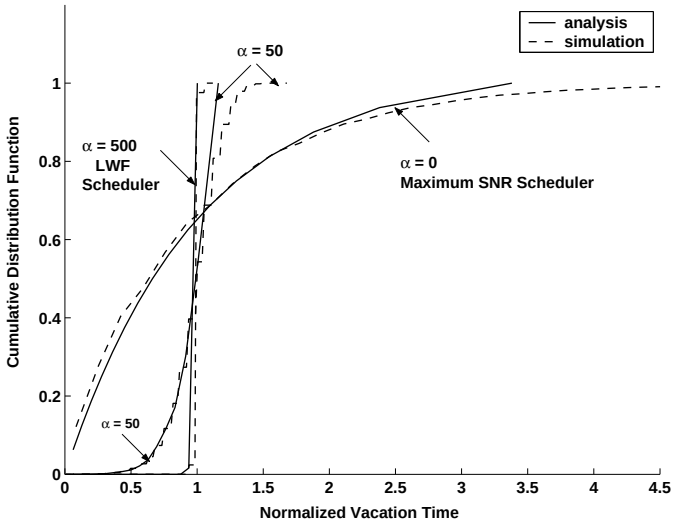

Figure 3.7: CDF of vacation time of scheduled user

simulation and analysis that the LWF scheduler ($\alpha = 500$), concentrates the mass of the CDF at the normalized vacation time of 1, scheduling users in a Round Robin manner. Since the scheduler is channel agnostic, the probability of a user being selected is zero for all but the user with the highest delay. The Maximum SNR scheduler ($\alpha = 0$), on the other hand, ignores delay, always favoring users with higher SNR. The highest normalized vacation time for the scheduled user is almost 5 times higher than that of the LWF scheduler.

Figure 3.8 illustrates the increasing throughput obtained by relaxing the delay constraint (smaller α) which manifests as the density of the scheduled rates concentrating in the higher end of the range. Contrast the distribution of scheduled rates in the Maximum SNR scheduler and the LWF Scheduler. The former maximizes the gain from multiuser diversity and schedules users at higher

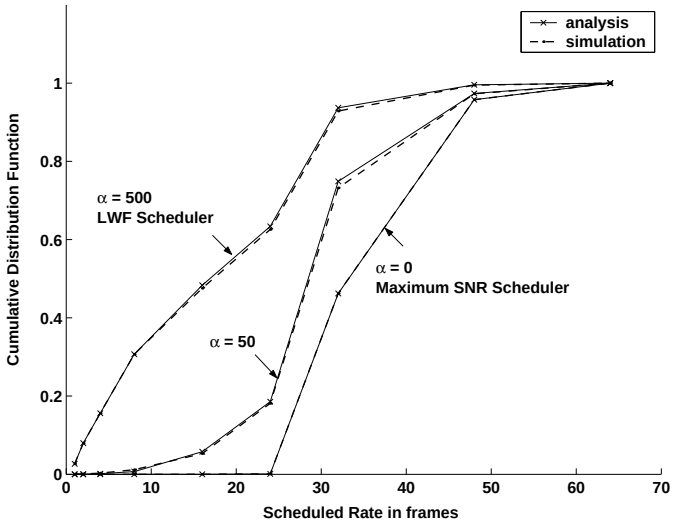

Figure 3.8: CDF of scheduled rate

rates. The CDF of the scheduled rates for the Maximum SNR scheduler is the product of the CDFs of the 16 users. Since there is a uniform probability of picking an user in any given slot, the distribution of the scheduled rate for the LWF scheduler is simply the distribution of the requested rates, $f_r(r)$ for any user as obtained from the Jakes model. In both figures, the distributions of the scheduled rate and vacation time for a value of $\alpha = 50$ lie between the two extremes.

3.3.4 Correlated Rates

In our analysis, we assume that the channel rates and therefore the requested rates for each user are independent of each other and identically distributed. Furthermore, the analysis also assumes that the requested rates are independent

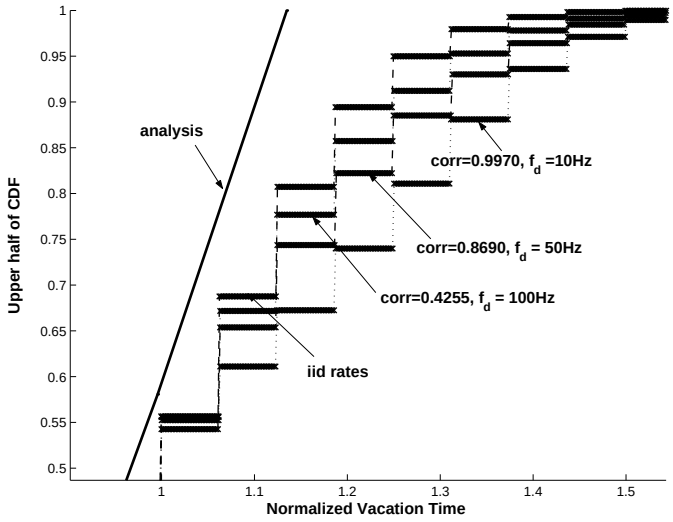

Figure 3.9: Comparison of vacation time CDFs (above median) for $\alpha = 50$ when channel rates are i.i.d./ correlated across time slots

from one time slot to another. In our simulations, we also study the performance of the system when channel rates are correlated across time slots. With correlated fades, the channel remains in a *good* state or *bad* state across several consecutive slots, depending on the coherence time. We now compare the CDFs of vacation time obtained from analysis with the CDFs obtained through simulation in four cases. In the first case, channel rates are assumed to be i.i.d across slots. By changing the doppler frequency, which directly changes the coherence time, we can simulate channels with varying correlation between channel SNR in consecutive slots. We then compare the CDFs from simulations when the SNR correlation is $0.9970(f_d = 10Hz)$, $0.8690(f_d = 50Hz)$ and $0.4255(f_d = 100Hz)$.

For clarity, we show the CDF of the vacation time above and below the

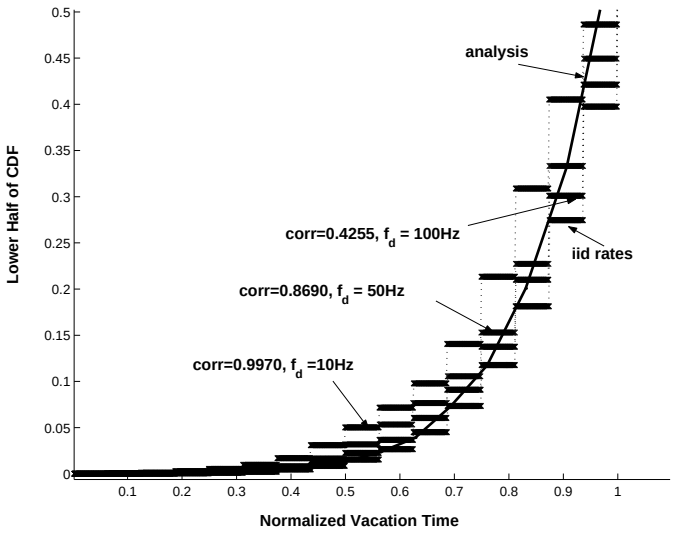

Figure 3.10: Comparison of vacation time CDFs (below median) for $\alpha = 50$ when channel rates are i.i.d./ correlated across time slots

median in two separate figures. As can be seen from figure 3.9, which shows the CDF of the vacation time for $\alpha = 50$ above the median (0.5 to 1), the divergence in the tail of the vacation time distribution from the CDF obtained from i.i.d channel SNRs increases with more highly correleated channels. When the channel corelation is large, there is a greater probability of scheduling users with higher vacation times. This can happen when (a) the user remains in a bad fade for a long duration i.e. in consecutive slots or (b) the user is pre-empted by other users with better channel conditions for a sustained duration. We see the opposite effect in the CDF of vacation time below the median in figure 3.10. In this case, the probability of being scheduled at at lower vacation times increases

as correlation between channel SNR in consecutive slots decreases. As expected, CDF from the simulation with i.i.d. rates in consecutive slots has the closest correspondence to the CDF obtained from analysis.

3.4 Summary

Innovative scheduling algorithms for packet switched air-interfaces have been shown to support higher data rates by exploiting the *multiuser diversity* inherent to cellular wireless systems. While such *opportunistic* schedulers significantly improve the system throughput, they could degrade the user experience through unfair resource allocation and increased variability in the scheduled rate and delay. The growing demand for service differentiation between real-time multimedia traffic and data traffic underscores the need for these schedulers to incorporate delay constraints.

The analytical results in this chapter not only highlight the inherent trade-off between system throughput and the delay experienced by mobile users with opportunistic scheduling, but they also provide a novel way of quantifying the QoS offered by a general opportunistic scheduler. Our analysis is strongly supported by system simulations of a time-slotted cellular downlink shared by multiple mobile users with independent, time-varying channels.

Chapter 4

Analyzing the Performance of Data Users in Packet Switched Wireless Systems with Prioritized Voice Traffic

4.1 Introduction

Packet switched cellular wireless systems have benefited greatly from multiuser diversity, which enables capacity gains by opportunistically scheduling users at favorable channel instants. While this is suitable for data applications, opportunistic scheduling does not fit the requirements of real-time or delay-sensitive applications such as Voice over IP (VoIP). The integration of voice and data services in cellular wireless networks that use packet switched air interfaces poses new challenges to network operators.

There has recently been significant interest in supporting real-time services in 3G wireless systems such as 1xEV-DO. This system, in particular, supports QoS

in a time-slotted packet switched air-interface by prioritizing delay-sensitive data in the wireline backhaul network as well as over the airlink [63]. The number of voice users supported by the system is an important factor determining the performance for data users that share system resources with voice users. Highly compressed voice traffic with stringent delay requirements constrains the sharing of system resources with data applications. We now highlight the important factors that affect the multiplexing of voice and data traffic over packet switched air interfaces.

4.2 Packet Voice in Cellular Wireless Systems

Circuit switched voice codecs used for wireline telephony in the PSTN encode telephone audio at 64kbps using pulse code modulation (PCM) schemes. Circuit switched cellular wireless systems such as GSM and IS-95 have conventionally used low bit-rate vocoders which operate at rates between 8kbps and 13.2 kbps in order to support a reasonable number of voice users with time-varying channel conditions in the available spectrum. These codecs use voice activity detection at the sender and comfort noise generation at the receiever to reduce the transmission rate during inactive speech periods. Speech frames are sampled at periodic intervals intervals typically between 10ms and 30ms. Voice over IP (VoIP) enables voice transmission over packet switched cellular air-interfaces such as 1xEV-DO. VoIP codecs which implement ITU standards such as G.729

and G.723.1 compress speech up to 8kbps and 5.3/6.3kbps respectively without significantly compromising speech quality. IP header compression is a typcial option for reducing the bandwidth of VoIP packets before transmission over the air-interface. If we assume 20ms speech frames with a voice activity factor of 0.5, then the base station in a time-slotted, packet switched air interface such as 1xEV-DO [5], [49] data system with a 1.667ms time slot duration can serve up to 24 voice users. A lower voice activity factor and longer interval between speech frames can increase the number of voice users that can be supported by the system, e.g., a maximum of 45 voice users can be supported in a system where voice is sampled by a codec with 30ms frames and a voice activity factor of 0.4. While the higher voice compression rates support a more efficient use of bandwidth, any further scheduling delay or delay jitter to voice packets could adversely affect the voice quality. In addition to providing end-to-end network transport functions for packetized voice by using protocols such as RTP [42], schedulers at base stations can employ suitable QoS support to ensure that the low delay requirements for voice traffic are met.

4.3 Expected Vacation Time with Prioritized Voice

We build on the analytical framework developed earlier in Chapter 3 to understand the effect of prioritized voice on the computation of the expected vacation time for the data users. In every time slot, the base station scheduler

serves voice users with the highest priority on a first-come first-served (FCFS) basis. If there is no waiting voice packet, the base station transmits to the data user with the highest metric as computed from equation 1.1. The tie-breaking rule described Section 3.1 is applied in the event that more than one user has the highest metric.

Observe that equations 3.1, 3.2, 3.4 and 3.5 remain unchanged when there is no voice packet scheduled, except that the number of users, N is now replaced by N_d. If a voice packet is scheduled, the expected vacation time of all users increases

$$\overline{V}_{u_i}(t+1) = \overline{V}_{u_i}(t) + \frac{1}{N_d}, \quad 0 \leq i < N_d \tag{4.1}$$

Assuming the existence of limiting values $\overline{V}_{u_i} = \lim_{t\to\infty} \overline{V}_{u_i}(t)$ and $\pi_{u_i} = \lim_{t\to\infty} \pi_{u_i}(t)$ the expected vacation time for the data user with rank i may be computed recursively as

$$\overline{V}_{u_i} = \overline{V}_{u_{i-1}} + \frac{1}{N_d\,(1-p_v)\left(1-\sum_{j<i}\pi_{u_j}\right)} \tag{4.2}$$

where p_v is the probability of scheduling a voice user in any slot. Observe from equation 4.2 that the increase in delay experienced by rank-ordered users is inversely proportional to the number of voice users in the system (p_v). The distribution for the vacation time for the ith rank-ordered user can can be computed in the same manner as in equation 3.18 by including the probability of

scheduling a voice packet as

$$f_i(\gamma) = p_v \left(f_i(\gamma - \frac{1}{N_d})(\sum_{j<i} \pi_{u_j}) + f_{i-1}(\gamma - \frac{1}{N_d})(\sum_{j \geq i} \pi_{u_j}) \right) + (1 - p_v) \left(f_i(\gamma - \frac{1}{N_d}) \right) \tag{4.3}$$

The distributions for the scheduled rate and vacation time for the scheduled user, as well as the packet service times can then be computed as given by equations 3.19, 3.20 and 3.22 respectively.

4.4 Simulation Results

4.4.1 Effect of Voice Calls on Vacation Time

The gains from multiuser diversity are maximized by setting $\alpha = 0$ in equation 1.1. In this case, the selection density function, π_u is uniformly distributed among the data users since the scheduler picks the data user with the best channel without any constraint on delay. Observe from equation 3.14 that the local slope of the expected vacation time is given by $\frac{1}{N_d(1-p_v)\left(1-\sum_{j<i}\pi_{u_j}\right)}$. For a fixed number of data users and $\pi_u = 1/N_d$, as the number of voice users increases, p_v increases, thereby causing the local slope to increase at every point. As expected, we see in Figure 4.1 that the increase in slope for the higher rank-ordered data users increases with p_v. We see from Figure 4.2 that normalized vacation time for data users also increases as they contend for system resources with more voice users.

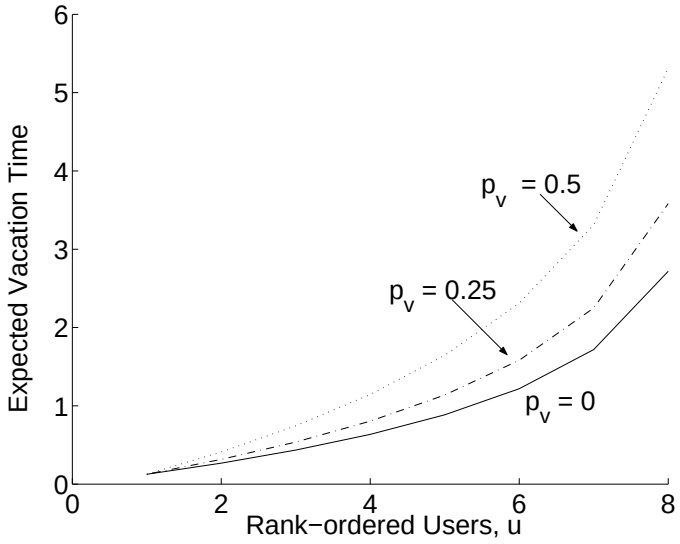

Figure 4.1: Impact of voice on the expected vacation time of data users

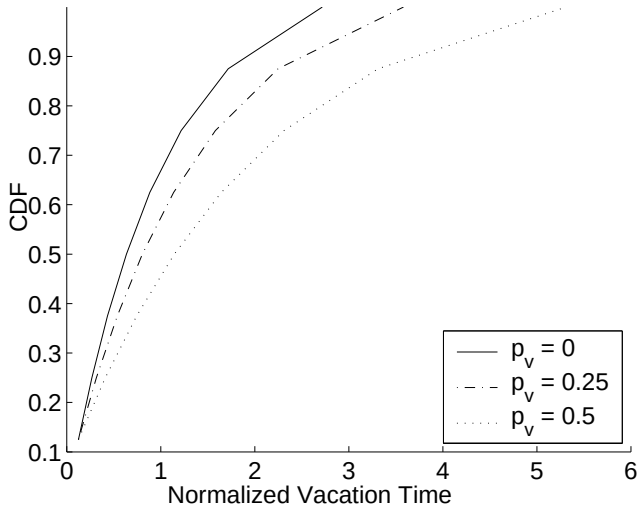

Figure 4.2: Impact of voice on CDF of vacation time for data users

Multiuser diversity gains are maximized for the data traffic by setting $\alpha = 0$ in the metric defined in Equation 1.1 for our simulations. It is important to note that a metric with α set to some non-zero positive value will only bias the scheduler to favor data users with higher values of vacation time, i.e., lower scheduling delays at the expense of multiuser diversity gain. These results are not included in this chapter for reasons of compactness. Furthermore, increasing the number of voice users will only cause scheduling delay to further dominate the scheduling metric at the expense of overall system throughput and degrade the performance for best-effort data applications even more.

4.4.2 Packet Service Time Statistics

The admittance of voice users naturally causes packet service times and delays for data users to increase. Figure 4.3 illustrates the CDF of packet service times experienced by data users in the absence of any voice calls. The CDF obtained by numerically evaluating the analytical expression from Section 3.1.5 is compared with simulated results when the channel rates are i.i.d., and shows very close correspondence. In the case of i.i.d. channel fades, the median packet service time is about 20 slots, while the packet service time at the 90th percentile is about 30 slots. In comparison, Figure 4.4 illustrates the effect on the packet service times when 50% of the time slots are occupied by voice calls. For i.i.d. fades, the median packet service time is about 35 slots, while the packet service

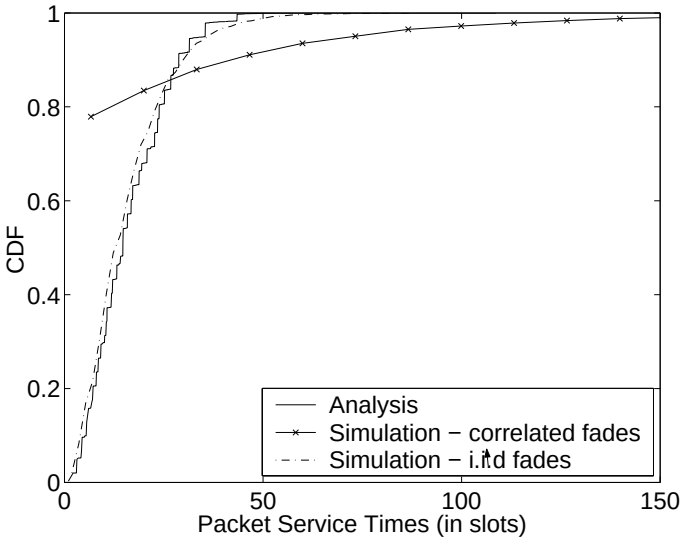

Figure 4.3: Packet service time CDF for 8 data users in the absence of voice

time at the 90th percentile increases to 60 slots.

The distribution of packet service times in the presence of correlated channel fading, which is obtained through simulation, is also included in these figures. In both cases, the distribution of packet service times for correlated channel fading occupies a much wider dynamic range. This is because a data user may be scheduled repeatedly with lower delays when the channel remains in a good state and higher delays when it remains in a bad state.

4.5 Summary

Cellular wireless systems that traditionally supported voice traffic now see an increasing demand for data services. Packet switched time-slotted air interfaces

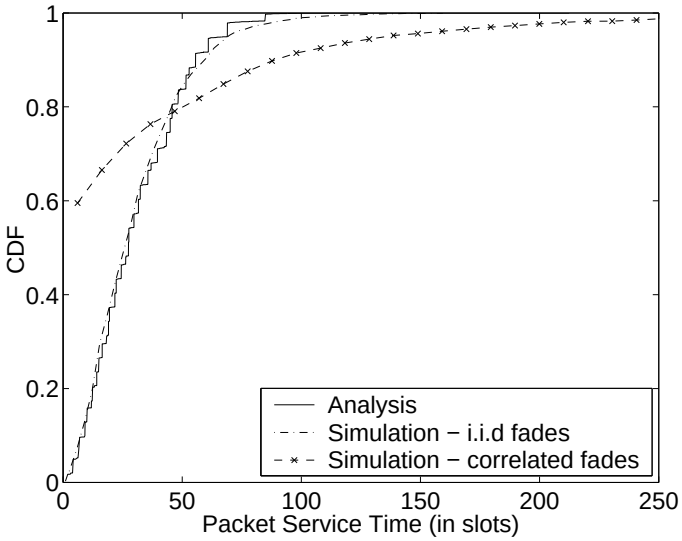

Figure 4.4: Packet service time CDF for 8 data users, with 50% probability of voice calls

such as 1xEV-DO that facilitate wireless data need to incorporate support for the varying QoS requirements of voice and data applications. The stringent delay requirements of compressed voice used in wireless telephony can be met by simply giving voice calls priority over data. While network operators would benefit from allocating system resources unused by voice users to data traffic, prioritized voice naturally limits the bandwidth and time slots available to data users. Data throughput, however, can be significantly improved by using scheduling algorithms that exploit multiuser diversity gain.

In this chapter, we present analysis and simulation of the scheduled rates, delays and packet service times experienced by data users multiplexed with prioritized voice users in a packet switched airlink. The analytical results in this

chapter provide useful tools for a network operator to evaluate whether a given mix of high priority voice users and lower priority data users achieves the desired performance objectives for the data users in terms of throughput and delay.

Chapter 5

A Co-operative Framework for Interference Controlled Reverse Link Scheduling in Multi-Cell CDMA Systems

In this chapter, we investigate reverse link scheduling as a capacity maximization problem with constraints on the maximum transmit power and interference generated both within and outside the cell. We first introduce the problem and then derive the necessary conditions for the optimal solution. The nature of the optimal solution suggests a greedy algorithm to allocate power to individual mobiles. We study the performance of this scheduling algorithm in maximizing aggregate throughput through simulations of a multi-cell reverse link architecture based on the 1xEV (IS-856) system.

5.1 The CDMA reverse link

The CDMA reverse link is an example of a multiple-access channel, where every mobile transmits simultaneously over a common uplink channel. Each mobile uses a distinct signature sequence or code to modulate its information sequence [58]. Practical CDMA systems use Walsh codes as signature sequences. In principle, the base station receiver on the uplink can use a *multiuser detector* [54] to jointly demodulate and decode the signals of all the simultaneously transmitting mobiles. The capacity of the Gaussian CDMA channel was first studied by Verdú in [55, 56]. Optimal signature sequences and the sum capacity of the CDMA reverse link in symbol synchronous CDMA systems have been studied in [41] when users have equal power and in [57] for asymmetric user powers.

Symbol synchronization and design complexity are important considerations that limit the use of multiuser detectors in base station receievers for cellular wireless systems. A simpler, but sub-optimal design with a bank of matched filters can be used in the receiver, where each matched filter corresponds to the signature sequence of a distinct mobile user. In this scenario, the output of any of the matched filters is the sum of two components. The first is a signal component that is proportional to a particular mobile's signal. The second component, whose energy is the sum of the receive powers corresponding to the signals of the other users, contributes to interference. The interference

component results from non-orthogonality of the signature sequences caused by timing offsets between the signals corresponding to different mobiles. In addition, the receiver also experiences thermal noise which is mostly related to the bandwidth of operation and the receiver noise figure. In essence, a matched filter treats the signals of other mobile users as noise in the process of extracting a particular mobile user's signal. Therefore, assuming the use of a conventional receiver with matched filters, every mobile interferes with every other mobile to the full extent of its receive power. While multiuser detectors have been actively researched for well over a decade, practical CDMA systems typically employ matched filter banks in the base station receiver.

5.2 The Reverse Link Scheduling Problem

Assuming a matched filter receiever, we now outline the reverse link scheduling problem by considering a network of cells in a cellular wireless system. To begin with, we focus on a single cell and its immediate neighbors in this network. Without loss of generality, we assign an index c_0 to the cell under consideration. Let $\mathcal{N}$ denote the set of mobiles served by the base station in the cell c_0. For each mobile $i \in \mathcal{N}$, let $\overline{p}_i^{max}$ denote the maximum transmit power of the mobile and g_{i0} denote the reverse link path gain. If the actual transmit power of the mobile is $\overline{p}_i$, the corresponding receive power p_i may be expressed, as in 2.4.2.1, as

$$p_i = \overline{p}_i g_{i0} \tag{5.1}$$

We assume that the reverse link signals transmitted by the mobiles in cell c_0 interfere with the reverse link transmissions in M neighboring base stations, which are denoted $\{c_j, j = 1, 2, \ldots, M\}$. Let $g_{ij}, i \in \mathcal{N}, j \in \{1, ...M\}$ denote the path gain from mobile i to the base station in cell j. The corresponding interference power is $\overline{p}_i g_{ij}$. The aggregate interference caused by the scheduled mobiles in c_0 to any neighboring cell is a function of the transmit powers and locations of the mobiles. Naturally, the aggregate out-of-cell interference experienced on the reverse link by the base station in cell c_0, denoted by I_o, is a function of the mobiles scheduled in neighboring cells.

Interference caused to any mobile $i \in \mathcal{N}$ has three components. The first component, I_c is the interference generated by circuit switched voice channels and control channels within the same cell. I_c scales linearly with the number of mobiles in the cell. For simplicity, we assume a fixed number of users, N in the system and therefore assume I_c to be constant. The second component comes from other data users within the same cell and is given by $(I_{in} - p_i)$, where $I_{in} = \sum_i p_i$. The third component I_o is the out-of-cell interference generated by scheduled mobiles in neighboring cells. The power of additive thermal noise is given by $N_0 W$, where W is the reverse link bandwidth and N_0 is the noise spectral density. Therefore, the ratio of the signal to interference plus noise

(SINR) for this mobile is given by

$$SINR_i = \frac{p_i}{N_0 W + I_c + I_o + \sum_{j \neq i} p_j} \tag{5.2}$$

Then, assuming capacity achieving rates, the total reverse link data capacity of the cell c_0 is given by

$$\begin{aligned} C &= \sum_{i=1}^{N} W log(1 + (SINR)_i) \\ &= \sum_{i=1}^{N} W log(1 + \frac{p_i}{N_0 W + I_c + I_o + \sum_{j \neq i} p_j}), \end{aligned} \tag{5.3}$$

In the literature, typical scheduling problems consider the allocation of transmit (or, equivalently, *receive*) power among the mobiles in order to maximize the sum capacity in the cell expressed in Equation 5.3, assuming that I_o is a constant and known value. In this work, instead of assuming complete independence between cells, we explictly consider the effect of reverse link scheduling on neighboring cells.

In CDMA systems, mobiles maintain *active sets* in order to aid soft handoff. For any particular mobile, an active set is a list of base stations whose pilot signal strengths are received above a pre-defined threshold. The active set membership evolves dynamically. If the quality of the received downlink pilot signal from a particular base station exceeds a threshold, then that base station is added to the mobile's *active set*. Similar criteria are used to remove a base station from the active set. Naturally, the received signal from the *connected* base station is the strongest. The active set consists of those base stations that the

mobile can potentially handoff to. The reverse link transmissions of the mobile are most likely to impact the reverse links of the base stations in its active set. This property is central to the scheduling paradigm developed in this work.

We classify mobiles into two types, *interior* mobiles and *edge* mobiles, depending on the quality of the signals received from base stations in the vicinity of the mobile. We define interior mobiles as those mobiles that have only the connected base station in their active sets. Edge mobiles have at least one neighboring base station in addition to the connected base station in their active sets. The motivation behind this classification is that edge mobiles are more likely to cause interference in neighboring base stations, and must therefore be scheduled more carefully. The main constraint while scheduling interior mobiles is the resulting in-cell interference.

In a symmetric cell geometry, it is reasonable to expect the aggregate out-of-cell interference experienced by c_0 to be approximately equal to the aggregate interference generated by the mobiles scheduled in c_0 to the neighboring cells, $\{c_j, j \in 1, 2, ..., K\}$. Therefore, if every cell controls the aggregate out-of-cell interference generated, the interference in the network is controlled and predictable and this results in a co-operative scheduling strategy across cells. As a practical matter, every cell can limit the aggregate out-of-cell interference generated to $I_o = \beta I_{in}$, and assume that it will experience the same level of out-of-cell interference for scheduling purposes.

Earlier research on reverse link scheduling algorithms involves the determination of optimal power allocation vectors that maximize total capacity in a single cell with in-cell-interference constraints. Out-of-cell interference has either been assumed constant or absent in these studies. We argue that these approaches do not model a practical system well since they consider a single cell in isolation. In a real system, optimal power allocation in one cell could significantly increase the interference in a neighboring cell depending on the location and transmission power of scheduled edge mobiles. If one cell increases power to compensate for increased interference from a neighbor, both cells would subsequently experience a decrease in throughput. In order to better approximate practical systems, we incorporate an additional *out-of-cell* interference constraint in a throughput maximimation problem. To the best of our knowledge this work is the first to address the reverse link scheduling problem with interference constraints both within and outside the cell. This approach leads to a co-operative scheduling algorithm where each base station in a cellular network maximizes the sum of mobile data transmission rates subject to linear constraints on (1) the maximum received power for individual mobiles (2) the total interference caused by scheduled mobiles to (a) other mobiles within the cell and (b) mobiles in neighboring cells.

5.3 Interference Constrained Optimization

Let $\mathbf{p} = \{p_1, p_2, \ldots, p_N\}$ represent the vector of receive powers at the base station where $p_i, i \in \mathcal{N}$ is the receive power corresponding to the reverse link signal transmitted by the ith mobile connected to that base station. If interference were to remain constant, the capacity achieving rate of mobile i on the reverse link is a function of the total receive power p_i. We first outline the typical constraints that determine the performance of the 1xEV reverse link.

1. *Power Constraint:* Every mobile typically has a maximum transmit power, $\overline{p}_i{}^{max}$. If the channel gain from this mobile to the base station is represented by g_{i0}, the maximum power that can be received at the base station is $p_i^{max} = \overline{p}_i{}^{max} g_{i0}$. The receive power constraint can be expressed as

$$0 \leq p_i \leq p_i^{max}, \quad \forall i \in \mathcal{N} \tag{5.4}$$

2. *Rise over Thermal (RoT) Constraint:* Circuit switched CDMA systems require the total interference to remain below a threshold [14] in order that the outage probability for circuit voice users and control channels be kept below acceptable levels. If we assume that the out-of-cell interference is controlled to be no more than $I_o = \beta I_{in}$, this constraint translates to an *rise over thermal* (RoT) constraint within the cell that can be expressed as

$$\frac{N_0 W + I_c + (1+\beta) I_{in}}{N_0 W} < \gamma_{RoT}$$

$$\text{which implies that} \quad I_{in} = \textstyle\sum_i p_i < \gamma \tag{5.5}$$

where $\gamma = \frac{1}{1+\beta}(N_0 W(\gamma_{RoT} - 1) - I_c)$. In the event that the reverse link primarily consists of scheduled data users, the RoT is less critical to system performance. However, restrictions on the RoT may still be required in order to reliably decode the reverse link control channels. Furthermore, when legacy circuit switched voice users are mixed with scheduled data users, as proposed in the 1xEV-DV system, the RoT continues to be an important constraint.

3. *Out-of-cell Interference Constraint:* Every mobile that transmits to its connected base station also generates interference in the reverse link receivers of neighboring base stations. This is termed out-of-cell interference, and is a particularly challenging aspect in the design of a scheduled reverse link. In legacy circuit switched systems, all mobiles in the cell transmit simultaneously, typically at low rates with correspondingly low power. The average interference is therefore a good measure of the aggregate interference seen at a neighboring base station. However, in the case of a scheduled reverse link, control of out-of-cell interference is more challenging for two reasons. Firstly, users at the edge of the cell need to be scheduled carefully in the event that QoS or other requirements require higher rates and therefore higher transmit power. Secondly, the nature of the interference changes from a heavily averaged Gaussian like signal to a bursty and unpredictable signal since mobiles located in different parts of

the cell are scheduled at different rates at different times.

In order to express this constraint mathematically, we introduce some notation. As defined earlier, mobiles connected to any base station are classified either as interior mobiles or edge mobiles. Let the *edge set*, $\mathcal{S}_k$ denote the set of mobiles in a particular cell which have the kth neighboring base station in their active sets. Specifically, a mobile i is assigned to $\mathcal{S}_k$ if the ratio g_{ik}/g_{i0} exceeds a pre-determined threshold. Observe that the edge sets are not mutually exclusive and that a mobile may be a member of more than one edge set. We denote the set of interior mobiles by $\mathcal{N}_{int}$, where

$$\mathcal{N}_{int} = \mathcal{N} - \bigcup_k \mathcal{S}_k \tag{5.6}$$

If I_o is the expected level of aggregate out-of-cell interference in any cell, and a cell is assumed to have K neighbors, then the out-of-cell interference contribution to a single cell from any one of its neighbors in a symmetric cell structure is no greater than I_o/K. In order to ensure this, every base station must control the aggregate interference generated to any particular neighboring base station to a level no greater than I_o/K. For instance, assuming hexagonal geometry and making the assumption that mobiles in a cell only affect the base stations that are immediate neighbors, K is 6. For a particular mobile i that is *received* at its own base station c_0 at a power p_i, the interference caused to a neighboring base station c_k is $(g_{ik}/g_{i0})p_i$. However, this is not a quantity that can be determined at the base station

since the ratio of path gains is unknown. The *estimated* aggregate interference caused by all mobiles in any edge set to the corresponding base station may be expressed by the following constraint

$$f(\sum_{j \in \mathcal{S}_k} p_j) \leq I_o/K \quad \forall k \in \{1, 2, \ldots K\} \tag{5.7}$$

where f is the attentuation factor. This factor represents the ratio of path gains g_{ik}/g_{i0} averaged over an area that would typically be occupied by mobiles belonging to a particular edge set. This is completely analogous to the factor f that is calculated in the uplink capacity analysis in [16, 59], and may be determined experimentally. In a symmetric network geometry, this factor is independent of the actual edge set under consideration.

We now introduce some notation that will be used in the rest of this section. In a cellular environment, where the path gain can vary by three orders of magnitude, it is possible for a single mobile to violate the interference constraints by transmitting at its maximum power. Under the constraints posed, the maximum permissible receive power for any mobile is given by

$$\begin{aligned} \widetilde{p_i}^{max} &= min(p_i^{max}, I_{in}),\ i \in \mathcal{N}_{int} \\ &= min(p_i^{max}, I_{in}, \frac{I_o}{fK}),\ i \in \bigcup_k \mathcal{S}_k \end{aligned} \tag{5.8}$$

We define a mobile to operate at *full* power if $p_i = \widetilde{p_i}^{max}$. A mobile is defined to operate at an *extreme* point if p_i is either equal to $\widetilde{p_i}^{max}$ or if $p_i = 0$. Therefore, for any mobile operating at an *intermediate* point, $0 < p_i < \widetilde{p_i}^{max}$.

We also define S^F, S^I and S^0 to be the sets of interior mobiles operating at full power, intermediate power and zero power respectively. Similarly, S_k^F, S_k^I and S_k^0 are the sets of edge mobiles with the kth neighboring base station in the active set that operate at full power, intermediate power and zero power respectively. We assume that interior mobiles do not cause interference to neighboring cells, i.e.,

$$(g_{ik}/g_{i0})p_i \approx 0, \quad \forall i \in \mathcal{N}_{int}, \forall k \in \{1, 2, \ldots K\} \tag{5.9}$$

We consider the following optimization problem

$$\max_{\mathbf{p}} \quad C = \sum_{i=1}^{N} Wlog(1 + \frac{p_i}{N_0 W + I_c + I_o + \sum_{j \neq i} p_j}) \tag{5.10}$$

subject to the three linear constraints given by the equations 5.4, 5.5 and 5.7.

Lemma 1 *The optimal solution* $\mathbf{p}^*$ *has no more than one interior mobile at an intermediate point.*

Proof: The proof by contradiction for this Lemma is along the lines of the proof for Theorem 1 in [22]. Assume that Lemma 1 is not true and that the optimal solution $\mathbf{p}^*$ has more than one interior mobile at an intermediate point. Suppose this is true for two mobiles, i.e.,

$$0 < p_i^* < \widetilde{p}_i^{max}, \quad i \in S^I$$

$$0 < p_j^* < \widetilde{p}_j^{max}, \quad j \in S^I$$

Let $C(\epsilon)$ denote the objective function obtained by increasing p_i^* by ϵ and decreasing p_j^* by ϵ. Observe that this change does not alter either the aggregate interference generated out-of-cell or the total power received within the cell. Hence, neither the out-of-cell nor the in-cell constraint are violated by this change in power. If $I = N_0W + I_c + I_o + I_{in}$, we have

$$\begin{aligned} C'(\epsilon) &= \frac{1}{I - p_i^* - \epsilon} + \frac{1}{I - p_j^* + \epsilon} \\ C''(\epsilon) &= \frac{1}{(I - p_i^* - \epsilon)^2} + \frac{1}{(I - p_j^* + \epsilon)^2} > 0 \end{aligned} \tag{5.11}$$

We see from the above equation that $C(\epsilon)$ is a strictly convex function. The objective function can therefore be increased by choosing a non-zero value for ϵ. However this contradicts our assumption that the solution with two interior mobiles at intermediate points is optimal. Therefore the optimal solution has no more than one interior mobile at an intermediate point.

Lemma 2 *The optimal solution* $\mathbf{p}^*$ *has no more than one edge mobile at an intermediate point in any edge set.*

Proof: As in the proof of Lemma 1, assume that Lemma 2 is not true and that the optimal solution $\mathbf{p}^*$ has more than one edge mobile at an intermediate point in the kth edge set, $k = 1, \ldots, K$. Suppose this is true for two mobiles, i.e.,

$$0 < p_i^* < \widetilde{p}_i^{max}, \quad i \in S_k^I$$

$$0 < p_j^* < \widetilde{p}_j^{max}, \quad j \in S_k^I$$

The out-of-cell interference generated by these mobiles at the base station corresponding to the edge set they belong to is $(g_{ik}/g_{i0})p_i^*$ and $(g_{jk}/g_{j0})p_j^*$ respectively. Without loss of generality, we can assume that

$$\frac{g_{ik}}{g_{i0}} < \frac{g_{jk}}{g_{j0}}, \quad i \in S_k^I, \quad j \in S_k^I \tag{5.12}$$

Let $C(\epsilon)$ denote the objective function obtained by increasing p_i^* by ϵ and decreasing p_j^* by ϵ. Observe that this change does not alter the total power received within the cell. If we denote the difference in out-of-cell interference at base station k due to this new power assignment by $\Delta I_o^k(\epsilon)$, then

$$\Delta I_o^k(\epsilon) = \left(\frac{g_{ik}}{g_{i0}} - \frac{g_{jk}}{g_{j0}}\right)\epsilon \tag{5.13}$$

By equation 5.12, $\Delta I_o^k(\epsilon) < 0$ for $\epsilon > 0$. This implies that the out-of-cell constraint is not violated by the transfer of power. However, the optimality of the new solution can be contradicted as in the proof of Lemma 1. By the strict convexity of $C(\epsilon)$ as derived in equation 5.11, the objective function can be increased by transferring power from mobile j to mobile i without violating either the in-cell constraint or the out-of-cell constraint until Lemma 2 is satisfied. Therefore, The optimal solution $\mathbf{p}^*$ has no more than one edge mobile at an intermediate point in any edge set.

Lemma 3 *If there exists an interior mobile at an intermediate point in the optimal solution* $\mathbf{p}^*$*, there cannot exist an edge mobile at an intermediate point in any edge set.*

Proof: Assume that Lemma 3 is not true and that the optimal solution $\mathbf{p}^*$ has one interior mobile at an intermediate point and one edge mobile at an intermediate point in the kth edge set, $k = 1, \ldots, K$. Suppose this is true for two mobiles, i.e.,

$$0 < p_i^* < \widetilde{p}_i^{max}, \quad i \in S^I$$

$$0 < p_j^* < \widetilde{p}_j^{max}, \quad j \in S_k^I$$

From Equation 5.9 and the proofs for Lemma 1 and Lemma 2, it may easily be seen that for some $\epsilon > 0$, power can be transferred from the edge mobile j, whose receive power decreases to $p_j^* - \epsilon$, to the interior mobile i, whose receive power increases to $p_i^* + \epsilon$, without violating either the in-cell constraint or the out-of-cell constraint. Furthermore, from equation 5.11, the objective function $C(\epsilon)$, obtained by increasing p_i^* by ϵ and decreasing p_j^* by ϵ is a strictly convex function of ϵ. The objective function can therefore be increased by transferring power from the edge mobile j to the interior mobile i without violating the interference constraints until Lemma 3 is satisfied.

Converse: *If there exists an edge mobile at an intermediate point in any edge set in the optimal solution* $\mathbf{p}^*$*, there cannot exist an interior mobile at an intermediate point.*

We now provide the necessary conditions for the optimal solution.

Theorem 1 *The optimal solution* $\mathbf{p}^*$ *is of the form* $\mathbf{p}_a^*$ *or* $\mathbf{p}_b^*$*:*

1. *In $\mathbf{p}_a^*$, only one interior mobile operates at an intermediate point while all other interior and edge mobiles operate at extreme points.*

2. *In $\mathbf{p}_b^*$, all interior mobiles operate at full power while at most one mobile in any edge set operates at intermediate power.*

Proof: It follows from Lemmas 1 to 3.

5.4 The Scheduling Algorithm

Motivated by the form of the optimal solution outlined in Theorem 1, we now provide a scheduling algorithm that constructs a solution through a greedy allocation of power in each step. For simplicity, we assume that interior mobiles do not cause out-of-cell interference. The algorithm can easily be generalized to accomodate changes in this assumption.

The algorithm is initialized with the constraints and proceeds through a sequence of steps, scheduling and allocating receive power to a single mobile in each step. The algorithm terminates when the constraints are satisfied, or when no more mobiles can be scheduled. Let the vector $\mathbf{p} = \{p_1, p_2, \ldots p_N\}$ denote the allocated receive power for the mobiles. We define $I_{in}^{(l)} = I_{in}^{(0)} - \sum_i p_i$ as the in-cell interference budget that can be used in the lth step. Similarly, $I_{o,k}^{(l)}$ denotes the out-of-cell interference budget for each edge set S_k in the lth step. The in-cell interference constraint (out-of-cell constraint for edge set S_k) is satisfied in step l

if the corresponding interference budget $I_{in}^{(l)}$ ($I_{o,k}^{(l)}$) is zero.

Algorithm

Initialize

- Set $l = 0$
- Initialize interference budgets

  $I_{in}^{(0)} = \gamma$, the maximum permissible in-cell interference

  $I_{o,k}^{(0)} = I_o/K, \quad k = \{1, 2, \ldots K\}$, where I_o is the out-of-cell interference threshold.

- Initialize adjusted maximum recieve power

$$\widetilde{p_i}^{(0)} = \widetilde{p_i}^{max} = \begin{cases} \min(p_i^{max}, I_{in}^{(0)}), & i \in \mathcal{N}_{int} \\ \min_{k:i\in\mathcal{S}_k} \min(p_i^{max}, I_{in}^{(0)}, \frac{I_{o,k}^{(0)}}{f}), & i \in \bigcup_k \mathcal{S}_k \end{cases}$$

- Initialize allocated recieve power

  $p_i = 0, \quad \forall i \in \mathcal{N}$

While($\exists$ at least one $i \in \mathcal{N}$such that $\widetilde{p_i}^{(l)} > 0$)

1. In this step, select the mobile with the highest adjusted maximum receive power to be scheduled,

   $i^* = \underset{i}{\text{argmax}} \ \{\widetilde{p_i}^{(l)}\}$

In the event of a tie, pick one user at random from the set of users with the highest adjusted receieve power

2. Allocate receieve power to the selected mobile and mark it as scheduled

   $p_{i^*} = \widetilde{p}_{i^*}^{(l)}$

   $p_{i^*}^{max} = 0$

3. Updates

   - $l \rightarrow (l+1)$
   - In-cell interference budget

     $I_{in}^{(l+1)} = max(0, (I_{in}^{(l)} - p_{i^*}))$.

   - Out-of-cell interference budget

     If $(i^* \in \bigcup_k \mathcal{S}_k)$,

     $I_{o,k}^{(l+1)} = max(0, (I_{o,k}^{(l)} - f p_{i^*}))\ \forall k : i^* \in \mathcal{S}_k$

4. Set the adjusted receive power for each mobile

$$\widetilde{p}_i^{(l+1)} = \begin{cases} \min(p_i^{max}, I_{in}^{(l)}), & i \in \mathcal{N}_{int} \\ \min_{k:i\in\mathcal{S}_k} \min(p_i^{max}, I_{in}^{(l)}, \frac{I_{o,k}^{(l)}}{f}), & i \in \bigcup_k \mathcal{S}_k \end{cases}$$

Continue

5.4.1 Algorithmic Complexity and Suboptimality

The optimal power allocation has to be obtained from the vector of permissible receive powers given by equation 5.8. In the simple case with no constraints on

the out-of-cell interference, there is a single constraint on the total in-cell interference in addition to the usual power constraints on the mobiles. In this case, finding the optimal solution is a simple procedure with linear complexity since there is at most one mobile at an intermediate point, as described in Section 2.4.2.2.

In the presence of out-of-cell constraints, the solution procedure is not as straightforward because of the partial dependencies that exist among subsets of mobiles that are members of the same edge set. A simple ordering procedure does not solve the problem since the maximum adjusted receive powers of the mobiles, $\widetilde{p}_i, i \in \mathcal{N}$, changes differently with power allocation for different groups of mobiles. The greedy algorithm proposed in Section 5.4 provides a low-complexity, suboptimal solution to the constrained throughput maximization problem.

In each step, one mobile is selected to be added to the set of scheduled mobiles. In the lth step, the constraints are first applied to the $N - l$ unscheduled mobiles and the mobile with the maximum adjusted power $\widetilde{p}_i$ is scheduled. The number of operations in each step is therefore $2(N - l)$. In the worst case, the number of operations required is $\sum_{l=0}^{l=N-1} 2(N - l)$ i.e. the worst-case complexity is $\mathcal{O}(N^2)$. Observe that in the case of high-powered mobiles, only a few mobiles will consume the in-cell interference budget. In this case, the algorithm could terminate in one or two steps, making the complexity linear.

A simple counter-example can be provided to show that the algorithm

provides a suboptimal solution. Suppose the ith mobile belonging to some edge set is scheduled in the lth step at an intermediate point. This implies that $0 < \widetilde{p}_i < \widetilde{p}_i^{max}$ and $\widetilde{p}_i \geq \widetilde{p}_j$ for all mobiles j in the set of remaining mobiles. Now consider the case when this set of unscheduled mobiles contains an interior mobile. It is feasible that this mobile is scheduled in a subsequent step at an intermediate point, thus consuming the remainder of the in-cell interference budget. Lemma 3 will then be violated, thereby making the solution suboptimal. We will show in our simulations that although this algorithm provides a suboptimal solution, it captures most of the benefits of interference constrained reverse link scheduling with highly reduced complexity.

5.4.2 Practical Considerations and Fairness

We will see in the simulation section that typical models for propagation loss assume that loss varies inversely as the third to fourth power of distance. As a consequence, unless edge mobiles in a typical cell of radius 1 km transmit at powers that are three to four orders of magnitude larger than interior mobiles, the base station scheduler will allocate all of the available interference budget to interior mobiles. Throughput optimal scheduling strategies [53, 35, 22, 27] are inherently unfair and typically schedule mobiles in the interior of a cell. Hence, alternatives have been proposed in the form of schedulers that use the achievable rate of an mobile augmented by a weight that reflects QoS or fairness constraints.

Out-of-cell interference control becomes all the more necessary with fair schedulers since edge mobiles are just as likely to be selected as interior mobiles. To illustrate this, consider a scheduling metric for the i^{th} mobile defined as

$$w_i.log(1 + \frac{\widetilde{p}_i}{N_0W + I_c + I_o + \sum_{j \neq i} \widetilde{p}_i}) \tag{5.14}$$

where $w_i > 0$ can be configured to implement varying degrees of fairness in the scheduling algorithm. One method, which we use in all our simulation experiments, is to set $w_i = q_i^{\eta}$, where q_i is the mobile's queue length and η is a scaling constant. In order to determine the power/rate allocation schedule with a fair scheduler that uses this metric, the algorithm in Section 5.4 is modified to rank mobiles in decreasing order of their maximum weighted rate.

5.5 Simulation Results

In this section we present simulation results that complement the analysis in Section 5.3. We first describe the system simulation model and the choice of parameters.

5.5.1 System Simulation Model

We assume a hexagonal cell geometry, where each cell consists of a base station with an omnidirectional antenna located at the center of a hexagonal area of

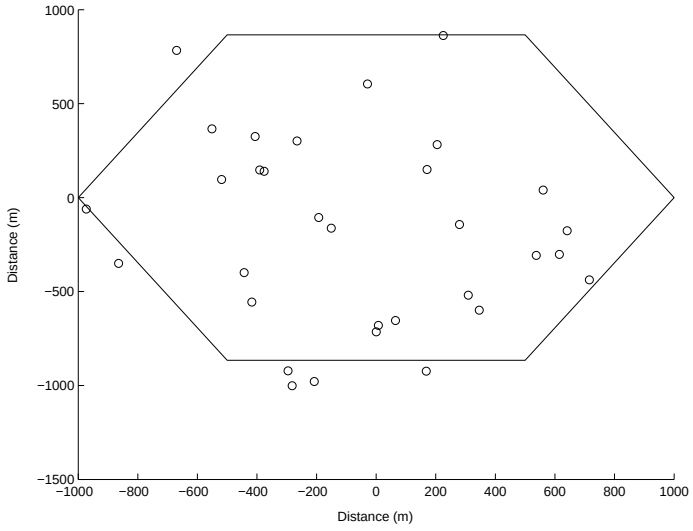

Figure 5.1: Location of 32 mobiles in a cell

radius 1km. 32 mobiles are randomly placed in the cell. In order to be consistent with the analysis, we neglect the effects of fast fading in our channel model.

The empirical Hata [21] urban propagation model was modified for the outdoor PCS environment by EURO-COST (European Co-operative for Scientific and Technolgical Research) as the COST-231 propagation model. In this model, which is well suited for cellular mobile communications, the path loss is given by

$$L = 46.3 + 33.9log(f_c) + 13.82log(h_t) - \alpha(h_r) + (44.9 - 6.55log(h_t))log(d) + C_M \tag{5.15}$$

where f_c is the frequency of operation in MHz, h_t is the effective transmitter(base station) height in meters, h_r is the effective receiver (mobile) antenna height in meters, d is the transmitter-receiever separation measured in meters and $\alpha(h_r)$ is the correction factor for effective mobile height. $C_M = 0dB$ for medium sized city

and suburban areas and $C_M = 3dB$ for metropolitan centers.

We assume the above COST 231 Modified Hata Urban propagation loss model for a carrier frequency of 1.9GHz and a minimum separation of 35m between the base station and mobile. The propagation loss in dB according to this model for a typical base station height, h_t of 34m and a mobile antenna height h_r of 1.5m is given by $(28.6 + 35log_{10}(d))$. Furthermore, we assume that the standard deviation of the lognormal shadowing component is 8.9dB. We assume an operational bandwidth of 1.25MHz, a noise spectral density of -174dBm/Hz and a Receiver Noise Figure of 5dB.

The structure of the reverse link data and control channels is described in Section 2.2 (Figure 2.3). In our simulation model, we assume that the Pilot, ACK and MAC channels for each mobile contribute to I_c. Since the ACK channels are transmitted only by the mobile that has received a downlink transmission in the corresponding slot, the interference caused by this channel may be ignored for analysis. The SINR required for pilots depends on the mobility level of the users as well as the number of receive antennas used on the base station. Assuming two receive antennas, the required pilot SINR is about -23dB [16] and the gain required for the DRC channel is about -1.5dB relative to the pilots. Therefore, I_c may be calculated as

$$I_c = N\ 10^{0.1*(-23-1.5)} P_{total} \tag{5.16}$$

where P_{total} is the total power received by the base station as expressed in

Equation 5.5.

As described earlier in this chapter, the out-of-cell interference is controlled by taking into account the cells that are affected by transmissions of different mobiles. The mobiles in the cell are grouped into edge sets. An edge set S_k corresponding to a neighboring base station c_k consists of those mobiles for which c_k lies in the active set. This implies that the ratio of the pilot strengths from c_0 and c_k exceeds a threshold. In our simulations. this threshold is chosen to be 6 dB. This is a simplification of the hysteresis loops that govern soft handoff behavior (and therefore candidate and active set memberships) in CDMA systems.

Finally, the traffic model used to generate the queue dynamics that are a part of the scheduling metric in Equation 5.14 is a simple Bernoulli model. In every slot, a mobile sees a probabilistic packet arrival. Those mobiles that end up being scheduled see their queues being depleted by their respective transmission rates in the slot.

5.5.2 Simulation Methodology

In our simulations, we schedule mobiles in a single cell using the scheduling metric outlined in Equation 5.14. The scheduler assumes the same level of incoming aggregate out-of-cell interference, I_o, as it is imposing as a constraint. In order to determine the actual rates that would result when multiple

neighboring cells schedule their mobiles independently, we adopt the following procedure. For any parameter set, the distribution of I_o generated to neighboring cells is first calculated. The simulation is repeated for the same parameters, with the throughput calculation using a value of I_o that is statistically sampled from this distrbution (uniform sampling of the CDF). This is completely equivalent to generating mobiles in every cell in a grid and using the interference produced to calculate throughputs.

For the same set of parameters, mobiles are randomly generated. For each such scenario, the simulation is run for a large number of slots (10,000) in order to fully sample the queue dynamics that result from application of the scheduler. The aggregate sector throughput, aggregate out-of-cell interference as well as in-cell interference distributions are collected across an ensemble of scenarios through Monte Carlo iterations.

We use the following convention in all the experimental results presented. The solid lines in the graphs represent simulation results when the scheduling algorithm constrains the out-of-cell interference caused by edge mobiles to the desired threshold. The dashed lines represent results when the out-of-cell interference is assumed to be a constant value, which is the same as the one used in the constrained case. In this case, the scheduling algorithm simply picks the user with the best weighted rate and no constraints are applied when scheduling edge mobiles.

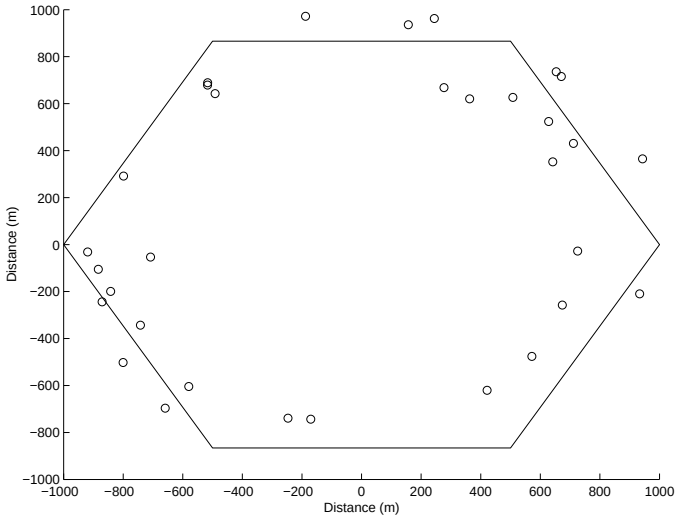

Figure 5.2: A pathological distribution of mobiles at the edge of the cell.

5.5.3 Interference Control in the Low-Power Scenario

In a circuit switched reverse-link, the nature of reverse link interference is predictable because of the geographical averaging over a number of mobiles. The power of the out-of-cell interference has been characterized as a fraction of the in-cell interference power in numerous publications [16, 59]. In the case of a scheduled reverse link, a similar pattern is to be expected when the transmit power of the mobiles is very low, which causes them to be scheduled together constantly in a circuit-like manner. In fact, explicit control of the out-of-cell interference has limited value even in a pathological case where the proportion of edge mobiles is very high as illustrated in Figure 5.2, with no mobiles within a circle of radius of 800m from the base station.

The distribution of the aggregate out-of-cell interference generated is shown

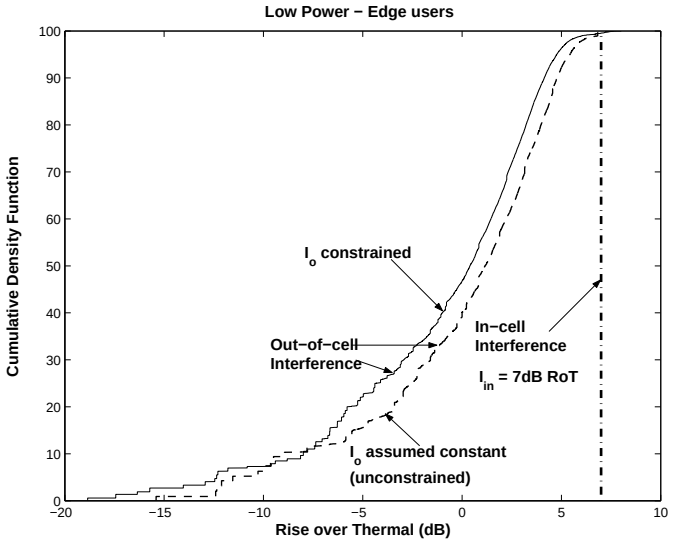

Figure 5.3: CDF of aggregate interference with low-power (50mW) edge mobiles, $I_{in} = 7dB$ RoT, $I_o = 4dB$ RoT.

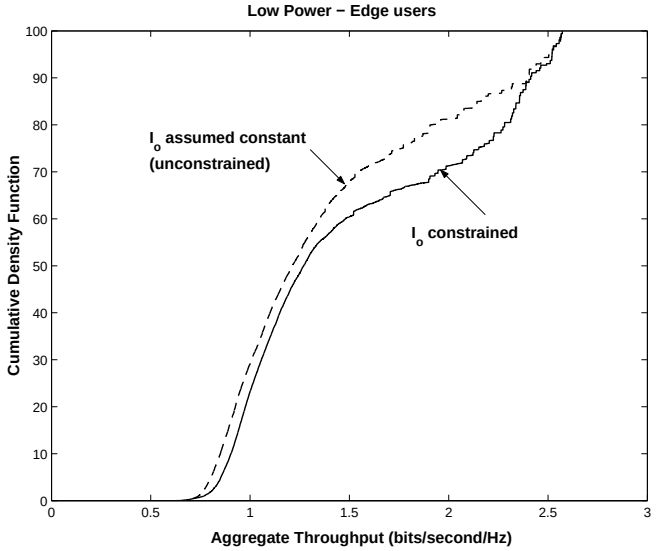

Figure 5.4: CDF of aggregate cell throughput with low power(50mW) edge mobiles, $I_{in} = 7dB$ RoT, $I_o = 4dB$ RoT.

in Figure 5.3. Corresponding throughput curves, shown in Figure 5.4 are generated using mobiles that are power limited to 50 mW. As is evident, interference control has no utility in this situation.

5.5.4 Interference Control in the High-Power Scenario

The situation is quite different when high power mobiles that are uniformly distributed in the cell are scheduled. It has been shown in this situation that the optimal scheduling strategy is to allow the best positioned mobile (capable of highest rate) to consume the entire available interference budget. This scheme by itself naturally leads to a high degree of unfairness, which is addressed using the weighted fair scheduler described above. We assume 32 mobiles per cell, each with a maximum transmit power of 1W. While this is high in comparison with cellular phones, it is a reasonable value devices powered by vehicle batteries or residential/desktop wireless modems.

In order to underscore the importance of enforcing out-of-cell interference constraints, consider Figure 5.5. This figure plots the distribution of the actual out-of-cell interference power that results from enforcing and ignoring the out-of-cell constraint. The in-cell interference maximum of 7dB RoT is typical in practical CDMA systems in order to allow successful decoding of control channels. We see from the curves that the out-of-cell interference is limited as required to 4dB RoT. This threshold, which is 3dB below the in-cell budget is

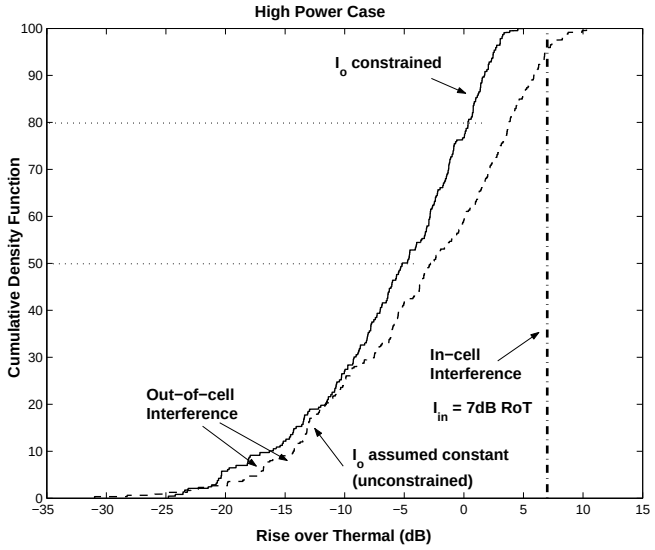

Figure 5.5: CDF of interference with uniformly distributed high power (1W) mobiles $I_{in} = 7dB$ RoT, $I_o = 4dB$ RoT.

enforced by the scheduler. Observe that the median interference power with out-of-cell constraints is 3dB RoT below or half the interference without constraints. At the 80th percentile this difference is 5dB. The smaller difference in interference between the two curves in the lower range of interference values can be attributed to the slots where interior mobiles are scheduled. As more edge mobiles are scheduled, the difference between the curves increases as can be seen from the divergence between the two curves.

Figure 5.6 illustrates the CDF of the aggregate cell throughputs that are obtained in the two cases. Not only is the throughput significantly higher in the case where out-of-cell constraints are enforced, but it is also much less *variable*. The throughput varies between 1.6bits/second/Hz and 2.6bits/second/Hz with

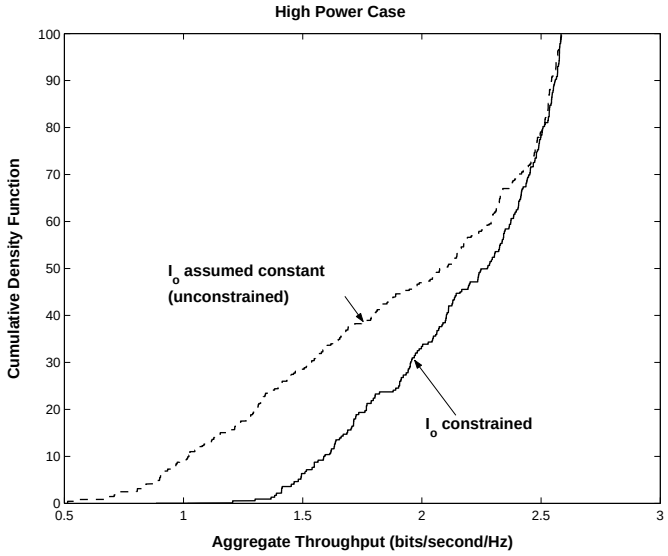

Figure 5.6: CDF of aggregate cell throughput with uniformly distributed high power (1W) mobiles $I_{in} = 7dB$ RoT, $I_o = 4dB$ RoT.

constraints enforced as against 0.4bits/second/Hz and 2.6bits/second/Hz without constraints

The effects of the interference control mechanism can be seen more clearly from Figures 5.7 and 5.8 when the maximum in-cell interference is 10dB RoT. We see the same trends in variations in throughput and interference as the 7dB RoT case. In this case, however, the higher in-cell interference budget allows more edge mobiles to be scheduled, and at higher powers when out-of-cell interference constraints are not imposed, leading to a drop in throughput. It must be noted that, in practical systems, interference of 10dB RoT would result in high outage for control channels and is therefore not feasible.

The effects of throughput variability are understated in this chapter because

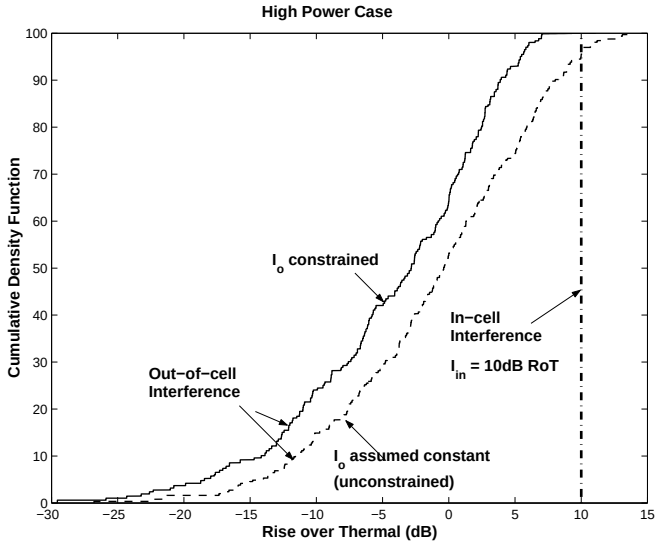

Figure 5.7: CDF of interference with uniformly distributed high power (1W) mobiles, $I_{in} = 10dB$ RoT, $I_o = 7dB$ RoT.

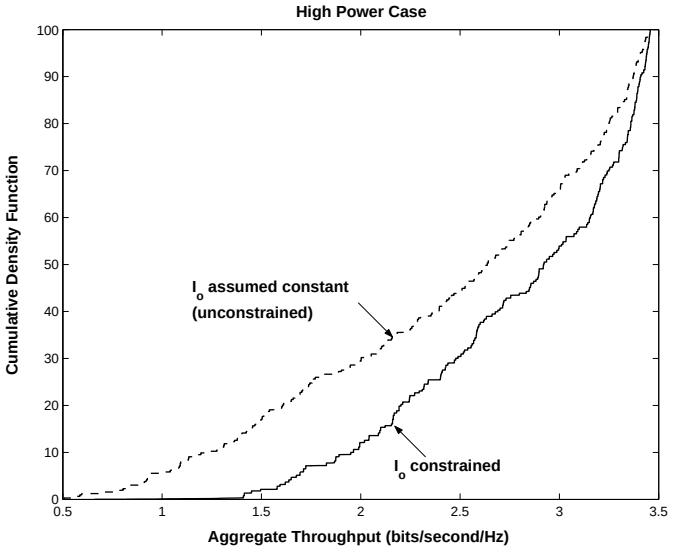

Figure 5.8: CDF of aggregate cell throughput with uniformly distributed high power (1W) mobiles, $I_{in} = 10dB$ RoT, $I_o = 7dB$ RoT.

we assume capacity-achieving rates. In a practical system, the wide dynamic range of out-of-cell interference caused by ignoring out-of-cell constraints will result in significantly higher segment error rates since scheduling decisions will have to be made assuming some knowledge of I_o. Alternatively, lower segment error rates can be obtained by building in sufficient margin to account for worst-case I_o, but this affects throughput to a large extent when the dynamic range of I_o is large. Note that in Figures 5.5 and 5.7, the probability with which the aggregate out-of-cell interference exceeds the expected value (4dB and 7dB respectively) is about 20% when the out-of-cell interference control mechanism is disabled.

5.6 Summary

This work in this chapter highlights and addresses the issues that come to the fore in scheduled reverse-links in practical multi-cell CDMA systems. An analytical characterization of the nature of the optimal solution leads to a greedy scheduling algoritm that computes the solution to the throughput maximization problem under power and out-of-cell interference constraints. Simulation results highlight the distinct advantages of this cooperative technique in ensuring predictable interference and much higher throughputs. This paradigm has direct application in evolving versions of the cdma2000 1xEV standard family.

Chapter 6

Conclusions

6.1 Forward Link Scheduling

The time-varying wireless channel capacity adds a new dimension to the problem of supporting broadband data services in cellular wireless networks. Implicit in the use of channel-state dependent scheduling algorithms are the questions of how these algorithms will provide QoS guarantees for a mix of traffic from delay-sensitive multimedia applications and non-real-time data traffic. The analytical results in Chapter 3 address the important issue of quantifying the QoS provided on the forward link of a time-slotted packet switched cellular wireless system. We completely characterize the distributions of the scheduled rates and delay in a general scheduler which realizes multiuser diversity gain with constraints on scheduling delay. In order to study this trade-off, we employ a general scheduling metric, $m = R + \alpha V$. The scheduler can be tuned to achieve the desired performance by varying the control parameter, α to balance the role of the channel rate, R or the delay, V in scheduling a user.

6.1.1 Fairness and Optimality of the Scheduling Metric

Fairness is an important performance criterion for a scheduler, given the large dynamic range of channel conditions in a cellular wireless system. The Proportional Fair (PF) [52] scheduler equitably shares forward link resources in a time-slotted system by ensuring all users get equal fractions of time-slots irrespective of their channel conditions. The scheduler metric proposed in equation 1.1 can also be shown to be *resource fair* with a simple change in the metric. The modified metric for user i is

$$m_i(t) = (R_i(t) + \beta_i) + \alpha V_i(t) \tag{6.1}$$

where β_i can be chosen optimally to maximize the total scheduled rate while ensuring resource fairness without delay constraints ($\alpha = 0$). This result is proved in [30].

Consider the typical case of users at different nominal SNRs distributed throughout a cell. It would be reasonable to assume that statistical variations about the nominal SNR are identical for all users in the same cell. In such a scenario, any scheduler that exploits multiuser diversity gain alone will favor the users at higher nominal SNRs. Let $\gamma_i(t)$ and $\gamma_{NOM,i}$ represent the instantaneous SNR reported by the user and the nominal SNR of the user respectively. Since the rate requested by user i is proportional to $log(\gamma_i)$, setting $\beta_i = -log(\gamma_{NOM,i})$ in equation 6.1 is equivalent to a *Normalized* Maximum SNR Scheduler. As can be seen from Figure 6.1, this scheduler results in an equitable distribution of time

slots among users with different nominal SNRs. For the case of 16 users with identical channel statistics about their nominal SNRs obtained from the Jakes channel model, we randomly distribute the nominal SNRs of the users in the range of 0dB to 8dB. While the Normalized Maximum SNR scheduler ensures a resource fraction that is roughly equal for all users, the Maximum SNR scheduler unfairly allocates a large fraction of time slots to users at the higher end of the range. In order to focus on the trade-off between throughput and delay alone, we assume that all users experience identical channel statistics in our analysis and therefore drop the fairness term, β_i, in the scheduling metric.

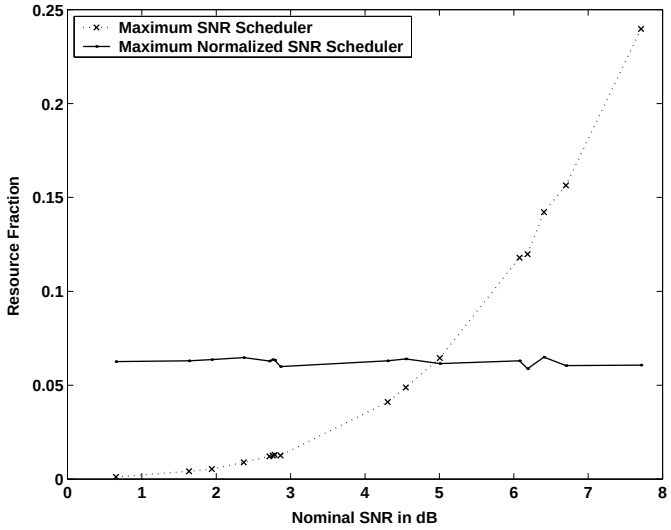

Figure 6.1: Resource fraction for 16 mobiles at different nominal SNRs but identical channel statistics

A natural question that would arise in the choice of a scheduler metric is whether the metric optimizes the scheduler performance with respect to some

criterion of interest. The objective of this work is not so much the design of an optimal scheduler as it is an analytical characterization of the trade-off between aggregate throughput and delay in a general opportunistic scheduler. The scheduler metric given by equation 1.1 can be configured to achieve a balance between multiuser diversity gain and delay constraints. The proposed scheduler lends itself well to statistical analysis while being sufficiently simple and versatile to be implemented in a real system. Our statistical analysis of forward link throughput and delay is validated by extensive simulations of a system architecture similar to a 1xEV-DO base station serving users.

6.1.2 Provision of Statistical QoS Guarantees

Consider the forward link of a CDMA cell that supports N mobile data users. Statistical QoS requirements for each user can be specified as

$$Pr(D_i > d_i) \leq \epsilon_i \tag{6.2}$$

where D_i is the steady-state packet delay for mobile i, d_i is the delay threshold for that user and ϵ_i is the probability of violating the delay threshold. The scheduling poilcy and traffic arrival pattern completely determine the probability of exceeding a given delay threshold. In Chapter 3, we combine the knowledge of the statistics of the time-varying channels of mobile users with a delay-constrained opportunistic scheduling policy to arrive at the packet service time statistics. In our model, we assume fully-loaded queues for all mobile users

in the system. While this simplifying assumption is very useful in evaluating the performance of the scheduler, a natural extension to the problem would incorporate an admission control scheme that would better describe the dynamics of individual user queues. The delay and throughput statistics in Chapter 3 are functions of the control parameter α and the channel statistics. We are currently investigating the choice of α that will provide feasible QoS guarantees when traffic characteristics and channel statistics are known.

6.1.3 Data User Experience with Prioritized Voice

The integration of wireless telephony and data services in 3G and 4G wireless systems that use packet switched air interfaces poses new challenges in the management of network resources. Although highly compressed voice traffic is given priority over data traffic, scheduling algorithms which exploit multiuser diversity have been shown to significantly improve the data throughput. In Chapter 4, we quantify the effect of prioritized voice traffic on the performance of data users in the system using a mix of analysis and simulation. We build on the analysis in Chapter 3 to characterize the scheduled rate, delay and packet service times for data in the presence of prioritized voice traffic by using a general scheduling metric that incorporates a measure of the user's channel quality in addition to a delay constraint. The results provide important tools for cellular network operators to evaluate system performance and provision resources for

traffic with varying QoS requirements.

6.2 Reverse Link Scheduling with Interference Constraints

The work in Chapter 5 highlights and addresses important aspects of scheduled reverse-links in practical multi-cell CDMA systems. We formulate a reverse link throughput maximization problem with three constraints. The transmission power of individual mobiles forms the first constraint. The second constraint requires a minimum SNR for decoding circuit switched channels that include voice users and control channels. This constraint is also called the Rise over Thermal (RoT) constraint since the threshold SNR is measured as a multiple of the thermal noise. We introduce a third constraint that limits the interference caused by scheduled mobiles in one cell to mobiles in neighboring cells. This constraint is plays an important role in limiting the interference caused by scheduling high powered mobiles at the edge of a cell.

Earlier approaches to the reverse link scheduling problem in CDMA systems considered a single cell in isolation by assuming the out-of cell interference to be constant. Scheduling algorithms in such a framework tend to decouple interference produced by scheduled mobiles within the cell and the interference from scheduled mobiles in neighboring cells. In our framework, we consider mobiles in single cell and the effect of mobiles in its six neighbors in a hexagonal cell geometry. We show that a co-operative scheduling algorithm that limits the

potential out-of-cell interference benefits all cells in the system. Predictable interference patterns are a natural result of the additional out-of-cell interference constraint. Higher throughputs with less variation also make the scheme more advantageous.

An analytical characterization of the nature of the optimal solution leads to a greedy scheduling algorithm that maximizes reverse link throughput under power and interference constraints in every step. Simulation results highlight the distinct advantages of this co-operative technique in ensuring predictable interference and higher throughputs. We now highlight some design considerations that are relevant in reverse link architectures for practical CDMA systems.

6.2.1 Maximum Recieve Power, p_i^{max}

As in earlier work [53, 27, 22, 35], we assume that the peak transmit power, $\overline{p}_i^{max}$ is known at the base station. While this is a reasonable assumption, current implementations of the 1xEV-D0 systems do not explicitly require mobiles to communicate this information with the base station. The base station *power controls* only the pilot/Reverse Rate Indicator(RRI) channels of all mobiles to be receieved at a fixed power. The power allocated to the DRC, ACK and Data channels are adjusted by a fixed gain relative to the Pilot/RRI channel in order to guarantee the desired performance of these channels. In particular, the relative

gain of the data channel increases with the data rate so that the received E_b/N_t is adjusted to achieve the required packet error rate (PER). As in the case of the RRI channel, the maximum receieved power, p_i^{max} can potentially be signaled periodically to the base station. Furthermore, in order to compute the maximum receieve power, which is used in scheduling, the channel gain also needs to be known since $p_i^{max} = \overline{p}_i{}^{max} g_i$. If the current transmit power, $\overline{p}_i$ is signaled periodically, then the channel gain g_i can be computed from the measured received power, p_i as $p_i/\overline{p}_i$.

6.2.2 Attenuation factor, f

The attenuation factor, f depends on the cell geometry, antenna pattern and the path loss model. For 3 sector hexagonal cells, the attenuation factor for different distributions of the lognormal shadowing component are computed in [16]. The value of f used in our algorithm can cause the scheduler to be either aggressive or conservative in allocating power to edge mobiles depending on how well the factor captures the typical interference pattern in the cell. In the field, the value of f used in the algorithm has to be tuned to maximize throughput as well as control interference. One possible approach is to use the measured out-of-cell interference to adjust the value of the attenuation factor f.

6.2.3 Active Set Knowledge

In the 1xEV-DO model shown in Figure 2.1, the Base Station Controller(BSC) has explicit knowledge of the active sets of all mobiles connected to base stations that report to the BSC. This information is necessary to facilitate soft hand-offs as the mobile moves from one cell to the next. In our scheduling algorithm, the base station needs to know the candidate sets of all connected mobiles in order to apply the out-of-cell interference constraints. While active set memberships are not known by the base station in current implementations, this information can be updated by the BSC whenever mobiles make changes to their active sets.

6.2.4 Implications on Soft-Handoff

In circuit switched links, soft-handoff enables an edge mobile to transmit at a lower power than it otherwise would, thus reducing the overall level of interference. In the case of the scheduled reverse link, soft handoff cannot be used to decode the data channels since sectors schedule mobiles independent of each other. Reverse link scheduling at the BSC, which would allow soft handoff, is an unattractive and perhaps infeasible option. The lack of soft handoff capability may potentially reduce some of the benefits of scheduled reverse links over circuit switched links in terms of sector throughputs.

6.3 Final Remarks

Research interest in providing QoS in broadband wireless data has grown with the increase in demand for high speed data services. In this dissertation, we have analyzed scheduling algorithms for the forward and reverse links in packet switched cellular data systems that support service differentiation. Through analysis and simulation models based on the 1xEV-DO link structure, we evaluate delay constrained algorithms that maximize scheduling gains by exploiting the channel variations of mobile users.

Appendix A

Computation of the Probability of Selecting a User in the Event of a Tie

This section details the computation of the probability of selecting a user in the event of a tie. Let $\mathcal{M}$ denote the set of unique values that the user metric, m_u, can take. Now consider a metric value, $m \in \mathcal{M}$. Suppose there are N_m users out of a total of N users who can take on this metric value. We define the event $\mathcal{E}_m$ as

$$\mathcal{E}_m = \text{two or more users with metric } m\text{, all other users with metric less than } m \tag{A-1}$$

This event may be represented as the union of mutually exclusive events as follows

$$\mathcal{E}_m = \bigcup_{k \geq 2} \mathcal{E}_{m,k} \tag{A-2}$$

where $\mathcal{E}_{m,k}$ denotes the event where exactly k users take on the metric m, with all other users having a lower metric. Since there are $\binom{N_m}{k}$ combinations of exactly k users from among N_m users, there are $\binom{N_m}{k}$ events that constitute the

event $\mathcal{E}_{m,k}$. We denote this partition as

$$\mathcal{E}_{m,k} = \bigcup_i \mathcal{E}_{m,k}^{(i)} \tag{A-3}$$

Let $\mathcal{S}_{m,k}^{(i)}$ denote the particular set of k users which take on the metric value m to form the event $\mathcal{E}_{m,k}^{(i)}$. The probability of the event may be computed as

$$P(\mathcal{E}_{m,k}^{(i)}) = \prod_{u \in \mathcal{S}_{m,k}^{(i)}} P(R_u = m - \alpha V_u) \prod_{u' \in (\mathcal{S}_{m,k}^{(i)})^C} P(R_{u'} < m - \alpha V_{u'}) \tag{A-4}$$

A user $u \in \mathcal{S}_{m,k}^{(i)}$ is picked with uniform probability, $\frac{P(\mathcal{E}_{m,k}^{(i)})}{k}$. Finally, the probability of picking a user, u in the event that at least one other user has the same metric is

$$\begin{aligned} P_{tie-break}(u) &= \sum_m P_{tie-break}(u,m), \quad \text{where} \\ P_{tie-break}(u,m) &= \sum_{i,k} \frac{P(\mathcal{E}_{m,k}^{(i)})}{k}, \quad \forall i,k \text{ such that } u \in \mathcal{S}_{m,k}^{(i)} \end{aligned} \tag{A-5}$$

BIBLIOGRAPHY

[1] R. Agrawal, A. Bedekar, R. J. La, and V. Subramanian, *C3WPF Scheduler for GPRS/EDGE*, ITCom '01, 2001.

[2] M. Andrews, K. Kumaran, K. Ramanan, A.L. Stolyar, R. Vijayakumar and P. Whiting, *CDMA Data QoS Scheduling on the Forward Link with Variable Channel Conditions*, Bell Labs Technical Report, April 2000.

[3] M. Andrews, S. Borst, F. Dominique, P. Jelenkovic, K. Kumaran, K. Ramakrishnan and P. Whiting, *Dynamic Bandwidth Allocation Algorithms for High-Speed Data Wireless Networks*, Bell Labs Technical Report, 2000.

[4] M. Andrews, K. Kumaran, K. Ramanan, A.L. Stolyar, P. Whiting and R. Vijayakumar, *Providing Quality of Service Over a Shared Wireless Link*, IEEE Communications Magazine, Vol.39 No.2, pp. 150-153, February 2001.

[5] 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 Communications Magazine, pp. 70-77,July 2000.

[6] I. Bettesh and S. Shamai, *A Low Delay Algorithm for the Multiple Access Channel with Rayleigh Fading*, IEEE Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC '98, September 1998.

[7] I. Bettesh and S. Shamai, *Fading Channel : Information Theoretic and Communication Aspects*, IEEE Transactions on Information Theory, Vol 44, pp. 2619-2692, October 1998.

[8] P. Bhagwat, P. Bhattacharya, A. Krishna and S. Tripathi, *Enhancing Throughput Over Wireless LANs Using Channel State Dependent Packet Scheduling*, Proceedings, IEEE Infocom April 1996.

[9] *cdma2000 High Rate Packet Data Interface Specification*, 3GPP2 C.S0024-A Version 1.0, March 2004, available at http://3gpp2.org/Public_html/specs/C.S0001-D_v1.0_031504.pdf

[10] S. Chakravarty, R. Pankaj and E. Esteves, *An Algorithm for Reverse Traffic Channel Rate Control for cdma2000 High Rate Packet Data Systems*, Proceedings, Globecom'01, November 2001.

[11] M.C.Chan. and R. Ramjee, *TCP/IP Performance over 3G Wireless Links with Rate and Delay Variation*, Mobicom, September 2002.

[12] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[13] A. Demers, S. Keshav and S. Shenker, *Analysis and Simulation of a Fair Queueing Algorithm*, Journal of Internetworking Research and Experience, pages 3-36, October 1990.

[14] R.T. Derryberry, *1xEV-DV Evaluation Methodology* 3GPP2/TSG-C.R1002.

[15] E. Esteves, *The High Data Rate Evolution of the cdma2000 Cellular System*, Multiaccess, Moblity and Teletraffic for Wireless Communications, Vol. 5, pp 61-72, Kluwer Academic Publishers, 2000.

[16] E. Esteves, *On the Reverse Link Capacity of CDMA2000 High Rate Packet Data Systems*, Proceedings, IEEE ICC'02, Vol. 3, May 2002.

[17] C. Fragouli, V. Sivaraman and M. Srivastava, *Controlled Multimedia Wireless Link Sharing via Enhanced Class Based Queueing with Channel-state-dependent Packet Scheduling*, Proceedings of IEEE Infocom, pp.572-580, April 1998.

[18] A. Furuskar, s. Mazur, F. Muller, and H. Olofsson, *EDGE: Enhanced Data Rates for GSM and TDMA/136 Evolution*, pp. 56-66, IEEE Personal Communications Magazine, June 1999.

[19] K.S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver Jr. and C.E. Wheatley III, *On the Capacity of a Cellular CDMA System*, IEEE Transactions on Vehicular Technology, vol. 40, pp. 303-312, May 1991.

[20] M. Grossglauser and D. Tse, *Mobility Increases the Capacity of Wireless Adhoc Networks*, IEEE Infocom '01, April 2001.

[21] M.H. Hata, *Empirical Formula for Propagation Loss in Land Mobile Radio Services*, IEEE Transactions on Vehicular Technology, VT-29(3):317-323, 1980.

[22] P. Hosein, *On the Optimality Conditions for 1xEV-DV Reverse Link Scheduling*, 3GPP2-C30-20030616-019.

[23] International Mobile Telecommunications-2000 (IMT-2000), Publications available at http://www.itu.int/home/imt.html

[24] *Introduction to cdma2000 Spread Spectrum Systems - Revision D* 3GPP2 C.S0001-D Version 1.0, February 2004, available at http://www.3gpp2.org/Public_html/specs/C.S0001-D_v1.0_031504.pdf

[25] W. C. Jakes, *Microwave Mobile Communications*, Piscataway, NJ : IEEE Press, 1993.

[26] R. Knopp and P. Humblet, *Information Capacity and Power Control in Single Cell Multiuser Communications*, IEEE ICC '95, Seattle WA, June 1995.

[27] K. Kumaran and L. Qian, *Uplink Scheduling in CDMA Systems*, Proceedings IEEE Infocom'03, April 2003.

[28] R. Leelahakriengkrai, R. Agrawal, *Scheduling in Multimedia DS-CDMA Wireless Networks*, Technical Report ECE-99-3, ECE Department, University of Wisconsin, Madison, July 1999.

[29] X. Liu, K.P. Chong and N.B. Shroff, *Opportunistic Transmission Scheduling with Resource-sharing Constraints in Wireless Networks*, IEEE Journal on Selected Areas in Communications, Vol 19, No. 10, October 2000.

[30] X. Liu, K.P. Chong and N.B. Shroff, *A Framework for Opportunistic Scheduling in Wireless Networks*, Computer Networks, vol. 41, no. 4, p. 451-474, 2003.

[31] Christopher Lott, *Enhanced RL-MAC for 1xEV-DO*, 3GPP2 Standard Meeting, April 2003.

[32] S. Lu, V. Bhargavan and R. Srikant, *Fair Scheduling in Wireless Packet Networks* IEEE ACM Transactions on Networking, vol 7, pp. 473-489, August 1999.

[33] X. Meng, Z. Fu and S. Lu, *Robust Packet Scheduling in Wireless Cellular Networks* Mobile Networks and Applications, Vol. 9, Issue 2, pp113-123, April 2004.

[34] T.S.E. Ng, I. Stoica and H. Zhang, *Packet Fair Queueing Algorithms for Wireless Networks with Location-dependent Errors*, Proceedings, IEEE Infocom '98, pp. 1103-1111, March 1998.

[35] S.-J. Oh, A.D. Damnjanovic, A.C.K. Soong, *Information-Theoretic Sum Capacity of Reverse Link CDMA Systems* Proceedings IEEE VTC, Korea, April 2003.

[36] R. Padovani, *Reverse Link Performance of IS-95 Based Cellular Systems*, IEEE Personal Communications Magazine, pp.28-34, Third Quarter, 1994.

[37] A. Parekh and R. Gallager, *A Generalized Processor Sharing Approach to Flow Control- the Single Node Case*, ACM/IEEE Transactions on Networking, 1(3): 344-357, June 1993.

[38] P. Ramanathan and P. Agarwal, *Adapting Packet Fair Queueing Algorithms to Wireless Networks*, Proceedings, ACM Mobicom '98, October 1998.

[39] T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, Upper Saddle River, NJ, 2002.

[40] T.S. Rappaport, A. Annamalai, R.M.Buehrer and W.H. Tranter, *Wireless Communications : Past Events and a Future Perspective*, IEEE Communications Magazine, vol 40, pp.148-161, May 2002.

[41] M. Rupf and J.L. Massey, *Optimum Sequence Multisets for Synchronous Code-Division Multiple-Access Channels*, IEEE Transactions on Information Theory, Vol. 40, No.4, pp 1261-1266, July 1994.

[42] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, *RTP: A Transport Protocol for Real-Time Applications*, IETF RFC 3550, July 2003.

[43] S. Shakkottai and A.L. Stolyar, *A Study of Scheduling Algorithms for a Mixture of Real and Non-Real Time Data in HDR*, Proc. 17th International Teletraffic Congress, September 2001.

[44] S. Shakkottai and A.L. Stolyar, *Scheduling for Multiple Flows Sharing Time-varying Channel : The Exponential Rule*, American Mathematical Society Translations, Series 2, A Volume in Memory of F. Karpelevich, Vol. 207, 2002.

[45] R. Srinivasan and J. Baras, *Opportunistic Scheduling in Wireless Systems and TCP Performance*, Technical Report, TR 2002-48, I.S.R. University of Maryland, October 2002.

[46] R. Srinivasan and J.Baras, *Understanding the Trade-off between Multiuser Diversity Gain and Delay - an Analytical Approach*, To appear in Proceedings IEEE Vehicular Technology Conference, Spring 2004, May 2004.

[47] R. Srinivasan and J.Baras, *Analyzing the Performance of Data Users in Packet Switched Wireless Systems with Prioritized Voice Traffic*, P. Langendoerfer et al. (Eds.): Proceedings WWIC 2004, LNCS 2957, pp. 236-247, February 2004.

[48] Third Generation Partnership Project 2, Contributions available at http://www.3gpp2.org/

[49] TIA/EIA IS-856 , *CDMA 2000 : High‘ Rate Packet Data Air Interface Specification*, Standard, November 2000.

[50] P. Tinnakornsrisuphap and C. Lott, *On the Fairness of the Reverse-Link MAC Layer in cdma2000 1xEV-DO*, To appear in Proceedings, IEEE ICC'04, June 2004.

[51] D. N. C. Tse, *Optimal Power Allocation over Parallel Gaussian Channels*, in Proceedings, International Symposium on Information Theory, Ulm, Germany, 1997.

[52] D. N. C. Tse, *Opportunistic Communications: Smart Scheduling and Dumb Antennas*, presentation available at http://www.eecs.berkeley.edu/ dtse/intel128.pdf, January 2002.

[53] S. Ulukus and L.J. Greenstein, *Throughput Maximization in CDMA Uplinks Using Adaptive Spreading and Power Control*, IEEE Sixth International Symposium on Spread Spectrum Techniques and Applications, Parsippany, NJ, September 2000.

[54] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1988.

[55] S. Verdú, *Capacity Region of Gaussian CDMA channels: the Symbol-synchronous Case*, Proc. 24th Allerton Conf., pp.1025-1034, October 1986.

[56] S. Verdú, *The Capacity Region of the Symbol-Asynchronous Gaussian Multiple-access Channel*, IEEE Transactions on Information Theory, Vol. IT-35, pp.735-751, July 1989.

[57] P. Viswanath and V. Anantharam, *Optimal Sequences and Sum Capacity of Synchronous CDMA Systems*, IEEE Transactions on Information Theory,Vol. 48, No. 6, pp. 1295-1318, June 2002.

[58] A.J. Viterbi, *CDMA Principles of Spread Spectrum Communications* Addison-Wesley, Boston 1995.

[59] A.J. Viterbi, A.M.Viterbi, K. S. Gilhousen and E. Zehavi, *Soft Handoff Extends CDMA Cell Coverage and Reverse Link Capacity*, IEEE Journal of Selected Areas Communication, vol. 12, pp. 1281-1287, Oct. 1994.

[60] D. Wu and R. Negi, *Effective Capacity - A Wireless Channel Model for Support of Quality of Service*, To appear in IEEE Transactions on Wireless Networking 2003.

[61] Q. Wu and E. Esteves, *The cdma2000 High Rate Packet Data System*, in Chapter 4 of Advances in 3G Enhanced Technologies for Wireless Communications, March 2002.

[62] *1xEV: 1xEVolution IS-856 TIA/EIA Standard, Airlink Overview*, Qualcomm White Paper, November 2001. http://www.qualcomm.com/technology/1xev-do/whitepapers.html

[63] *1xEV-D0 Quality of Service* Qualcomm White Paper, 2003.

[64] *1xEV-DO System Architecture* Qualcomm White Paper, 2003.