McGill University
Information Networks and Systems Laboratory

Flow Control for Packet Data in an Integrated Wireless Access Network

Jean-Pierre Rabbath

Department of Electrical & Computer Engineering,
McGill University, Montréal.

May, 1998

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of
Master of Engineering

0-612-44037-0

Table of Contents

List of Figures

List of Tables

Abstract

We consider flow control in the downlink direction for a microcellular CDMA system carrying integrated packet level-data and voice services with the objective of providing maximum traffic capacity without service degradation due to overload. The goal of this research is to develop flow control techniques that regulates the admitted calls based on the prediction of the total transmitted power requirements over the next packet transmission interval. If the probability of exceeding the maximum power value is sufficiently low, the extra power required to accommodate the data transmission is unlikely to result in outage and, therefore, data flow is permitted. The autocovariance function of the transmitted power under maximally loaded traffic conditions allows such predictions.

With packet-level flow control, one can potentially transmit useful volumes of packet data even when no fixed mean data transmission rate can be guaranteed. The traffic admitted to any cell is increased if the traffic in the neighboring cells happens to be low. Using constant bit rate (CBR) speech transmission with 100% activity, the system can accommodate about 67 mobiles. With available bit rate (ABR), we can accommodate up to 82 data mobiles receiving data at the same average rate of 8 Kb/sec.

Sommaire

On considère le contrôle du flux dans la direction base-à-mobile pour un système micro-cellulaire CDMA comprenant des services intégrés comme les paquets de données et les services vocaux, avec l'objectif de maximiser la capacité du trafic sans la dégradation de la qualité due à une surcapacité. Le but de cette recherche est de développer des techniques de contrôle du flux, qui régulariseront les appels admis et ce, en se basant sur la prédiction des exigences de la puissance totale transmise durant le prochain interval de transmission. Si la probabilité d'excéder la valeur de puissance maximale est suffisamment faible, il est improbable que la puissance excedentaire requise pour accommoder la transmission de données, résulte en un excès, permettant ainsi la transmission de données. La fonction d'auto-covariance de la puissance transmise sous des conditions de trafic maximal permet de telles prédictions.

Avec le control de flux, on peut potentiellement transmettre des volumes de données utiles même quand aucun taux de transmission de données est garanti. Le trafic admis dans une cellule est augmenté si le trafic de la cellule voisine est bas. En utilisant un taux de transmission constant, avec 100% d'activité, le système peut accommoder environ 67 mobiles. L'utilisation d'un taux de transmission à disponibilité permet quant-à-lui, d'accommoder jusqu'à 82 mobiles recevant des données à un taux moyen de 8 Kb/sec.

Acknowledgments

This work was supported in part by the Canadian Institute for Telecommunications Research under the Canadian Government's Network of Centers of Excellence program as well as Nortel and FCAR Industrial Research Chair in Personal Communications.

Thank you to all the people who read and commented this document, especially to Salvatore D. Morgera and Paul Mermelstein whose critics and advises always proved to be useful and relevant. Many thanks to Elias Nemer who proposed many pertinent editorial changes in this report.

I would like to thank too Rafi Rabipour for making me part of the Wireless Technologies team in Nortel while I was still working on my Masters degree. Also, I'd like to thank Solange Rivest and Robert Dijkerman for encouraging and pushing me into finishing my Masters degree.

I finally want to express my gratitude and devotion to my parents and my friends who supported me when writing this report, especially Nadim Batri, Joe Nader, Marianne Zakhour, Akram Rahal and Aya BouChedid.

1. Introduction

1.1 Evolution of Personal Communications Services (PCS)

The objective of personal communication systems is to provide ubiquitous wireless communications coverage, enabling users to access the telephone network for different types of communication needs, with no regard to the location of the user or the location of the information.

Clearly, today's cellular communication is only a subset of true PCS where voice service is only one of the many features provided to the user. It is expected that services traditionally associated with other media, such as television and computers, will slowly become part of PCS phone capabilities.

Users will be expected to view still and moving images, listen to high fidelity music, order groceries, and connect to the Internet via cell phones in the near future. To provide these services, however, portable phones will have to be capable of receiving and sending much more information than is possible at present - in technical terms, they will have to be capable of handling high-bandwidth transmissions.

Over the last decade, the telecommunication industry has witnessed a rapid rate of evolution from the use of wireline to pagers to faxes to cell phones and electronic mail. Market researchers forecast that the scope and penetration levels of wireless services will continue to grow and reach mass-market levels of over 25% by the end of the millennium. As PCS services become more versatile and more affordable, they are expected to eventually replace most of today's wireline services, thus reaching the real objective of ubiquitous communication coverage.

1.1.1 Current Cellular Standards

Different types of cellular systems employ various methods of multiple access.

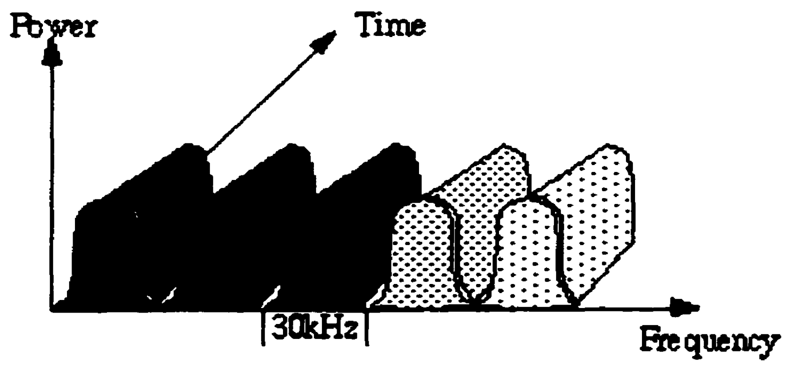

FIGURE 1. Frequency and time domain representation of FDMA

FDMA Standard. The traditional analog cellular systems, such as those based on the Advanced Mobile Phone Service (AMPS) and Total Communications System (TACS) standards, use Frequency Division Multiple Access (FDMA). With FDMA, only one subscriber at a time is assigned to a channel, as illustrated in Figure 1. No other conversations can access this channel until the subscriber's call is finished, or until that originated call is handed off to a different channel by the system. Traditional FDMA systems divide bandwidth into smaller channels of 30 kHz each, with one channel used for transmission and one for reception.FDMA systems are not bandwidth efficient since a frequency channel is permanently allocated to a user regardless of speech activity. Statistically, a user talks only about 40% of the time.

TDMA Standard. A common multiple access method employed in new digital cellular systems is Time Division Multiple Access (TDMA). TDMA digital standards include North American Digital Cellular (known by its standard number IS-54), Global System for Mobile Communications (GSM), and Personal Digital Cellular (PDC).

TDMA systems commonly start with a slice of spectrum, referred to as one *carrier*. Each carrier is then divided into time slots as shown in Figure 2. Only one subscriber at a time is assigned to each time slot, or channel. No other conversations can access this channel until the subscriber's call is finished, or until that original call is handed off to a different channel by the system.This method offers a threshold increase in capacity over an FDMA system, since a frequency channel is used for multiple users. For example, in IS-54, three users share a channel otherwise used by only one user in the AMPS system.

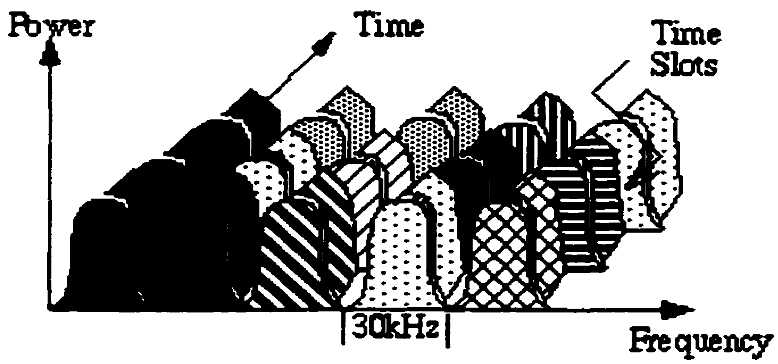

FIGURE 2. Frequency and time domain representation of TDMA

CDMA Standard. While Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA) both use barriers such as time and frequency to separate users, CDMA combines several users in the same channel at once as shown in Figure 3. Each user is assigned a unique pseudonoise, or spreading, (PN) code to allow the decoding of individual transmissions. The number and length of these PN codes depends on the system implementation. For example, in IS-95, 64-bit Walsh codes are used, thus allowing 64 perfectly orthogonal transmissions and resulting in a spreading over a 1.25 MHz band.

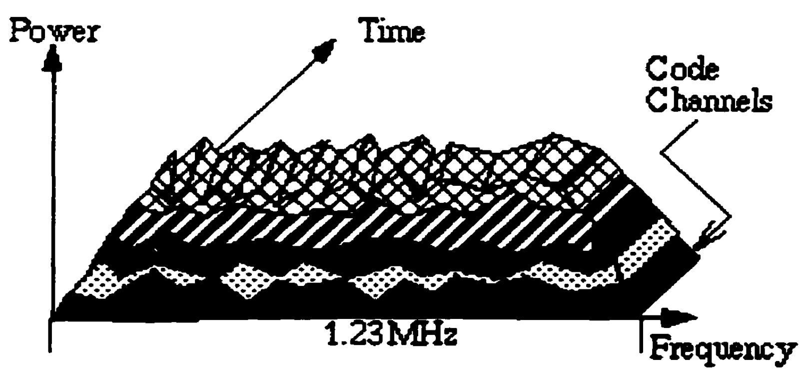

FIGURE 3. Frequency and time domain representation of CDMA

For cellular telephony, CDMA is a digital multiple access technique specified by the Telecommunications Industry Association (TIA) as IS-95. IS-95 systems divide the radio spectrum into carriers which are 1.25 MHz *CDMA channels* or *carriers*, each supporting these multiple code channels which share the 1.25 MHz bandwidth simultaneously. In each cell, each RF carrier can be further split into sectors that concentrate capacity in a particular direction, thus permitting additional reutilization of the spectrum.

One of the unique aspects of CDMA is its "soft capacity limit", i.e., the number of phone calls that can be handled by a carrier is not a fixed number. Rather, the capacity of the system is dependent on a number of different factors.

A detailed description of current CDMA technology is presented in the next chapter.

1.1.2 The Data Business

Service integration is considered particularly important to PCS providers whose traffic mix can be expected to significantly shift from an early predominance of voice with some low-speed data to an increased penetration of high-speed data services and service packages that combine voice with a number of features, including messaging, still image and video transmission, local area network connections, and Internet information services. in addition, it is desirable to avoid modifying the network as new services are introduced and the traffic mix changes.

The demand for data services has been growing exponentially, as is apparent by the growth of Internet users and service providers, modems and content providers in recent years. According to the latest statistics, more than 50% of transatlantic traffic is data [Wireless97]. Worldwide pager and cellular phone subscribers have also been steadily rising in the past few years. One out of every three computers sold in 1996 was a laptop. This phenomenon is mostly due to the fact that businesses are becoming increasingly global, which has resulted in an increased number of mobile professionals who need timely access to information. In addition, wireline technologies are adding mobility management functions, such as Mobile IP, to their architecture. Wireline and wireless backbone technologies are also starting to converge. All of these trends lead to the prediction that by the year 2000, 30% to 50% of all wireless traffic will be data. High speed data services will become common in home and office environments, and lower bandwidth solutions will remain the norm in mobile environments.

1.1.3 Data Services in PCS systems

It can be argued that personal computing was the most significant development in information technology in the 1980s. By 1997, millions of users had grown accustomed to having a computer on their desks, with each computer organized to meet the needs and preferences of its owner. Since the early 1990s, two powerful trends have propelled advances in personal computing. One is the popularity of portable computers: laptops, notebooks, and personal digital assistants.

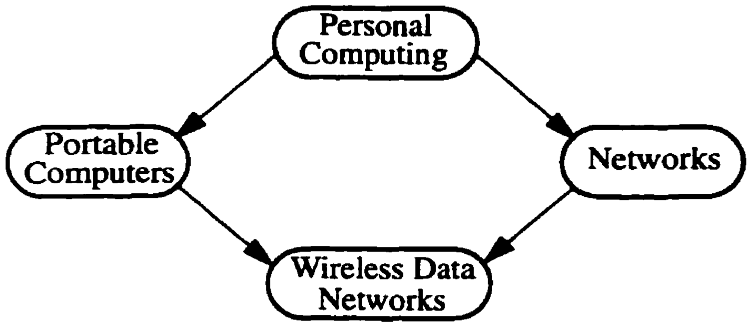

FIGURE 4. Wireless data networks' advantages

The other major trend in personal computing is networking which is triggered by the need to connect to other computers or databases and to have a variety of information sessions. The simultaneous popularity of portable computing and networking poses a paradox, because portable computers are "disconnected" from the wires of conventional networks and from public supplies. As illustrated in Figure 4, wireless data networks resolve this paradox. By making use of wireless data networks, the owners of portable computers retain the advantages of mobility while remaining connected to their important information services [1].

1.2 Integration of Stream and Packet Based Services

The research reported here builds on recent innovative solutions for the integration and management of stream- and packet-based services within the context of a CDMA wireless network. Specific objectives include:

- the efficient provisioning of packet data services,
- the integration of these services with voice in a bandwidth-on-demand fair-sharing manner, and
- the determination of available capacity for packet data transmission when stream calls have been admitted to capacity.

A solution will have two components, one at the call level, in the form of a call-admission controller, and the other, a time manager or flow control algorithm applied to packet-based services and operating at the burst or packet level.

Recent results include an approach to call-admission control, applicable to stream sources at different rates, which is fair, efficient, and novel in its use of both measured and declared traffic parameters for the estimation of the volume of resources available for new connections [Mermelstein95]. A new algorithm, using short-term predictive estimates of downlink power, is proposed here for intra-call management of packet flows, an activity which is central to the implementation of Available Bit Rate (ABR)-like services in the wireless environment.

1.2.1 Flow Control

Admission control prevents congestion by regulating the number of active stream calls on a network. At the time of call establishment, a stream call may be allowed or denied access to a network depending on the degree of network congestion. Flow control avoids network congestion by modulating the packet input traffic and reducing the burstiness of the packet input traffic as is shown in Figure 5. Flow control could be implemented at a traffic source by buffering incoming calls and injecting them into a network when resources are available. This requires the knowledge of both the current traffic conditions and the impact of adding a new connection.

Our main objective is to build the traffic manager based on short term resources. This information is provided by identifying the available capacity for packet data transmission by predicting the response of the total transmitted power to the transmission of a packet of data and by finding new methods of studying mixed data and voice traffic. The existence of fast Rayleigh fading and slow shadow fading suggests some correlation over time. The proposed flow control algorithm presents applicable traffic management procedures which are sufficiently simple to implement in practical applications, assuming that we know, or can estimate, the probability mass (discrete random variable) function of the incoming traffic source.

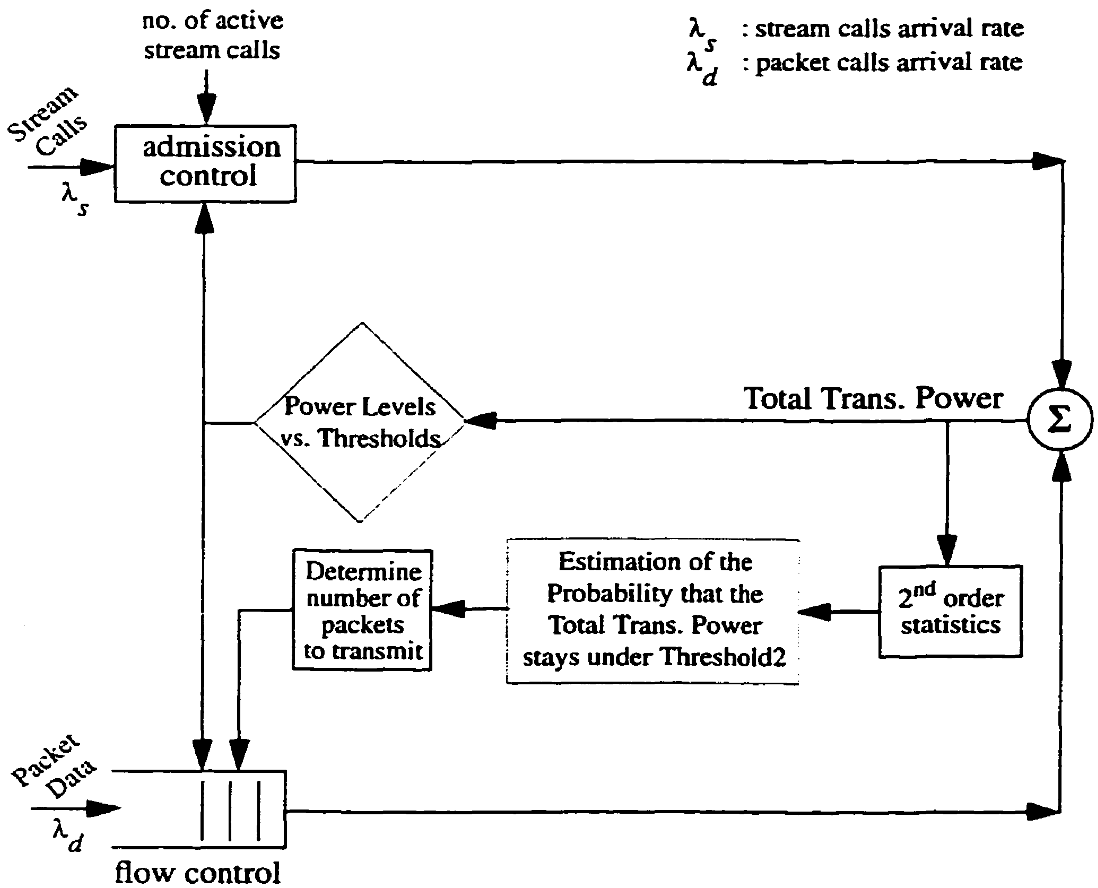

FIGURE 5. Approach to the integration of packet level data and voice

1.3 Thesis flow

This thesis is organized as follows: first, the wireless network transport level is described in Chapter 2. It is a model for a third-generation wireless access system that can support a wide variety of wireless services at different and variable transmission rates in both indoor and outdoor environments. Then, in Chapter 3, the objectives are clearly stated with an overview of previously suggested algorithms. The proposed flow control algorithm is then outlined in detail. It is based on the prediction of the power requirements over the next packet transmission interval using 2nd order statistics. Chapter 4 presents numerical results and discusses the simulations for various traffic conditions. Finally, conclusions are presented in Chapter 5.

2. Wireless Access Systems

Code-Division Multiple-Access (CDMA) is an attractive technology for personal communication systems. One of its main advantages is its ability to provide high utilization of the bandwidth, which is a scarce resource in most wireless environments. The high bandwidth utilization is achieved as a result of the statistical multiplexing of many independently fading signals with independently variable transmission requirements in the allocated bandwidth.

In this chapter, the transport level of the wireless network is described as mentioned in [IWAN]. In this network, Direct-Sequence Code-Division Multiple Access (DS-CDMA) is employed to accommodate as many high-demand services as possible within the bandwidth limitations of the environment. As only the downlink is considered in this project, the details of the downlink process are emphasized vis-á-vis uplink details.

This chapter is divided as follows: First, the CDMA system is described. Then, the transport level for an Integrated Wireless Access Network (IWAN) is outlined in details.

2.1 The CDMA System

CDMA is a multiple-access technique for application in cellular networks. Theoretically, CDMA offers significant advantages over existing analog and digital cellular technologies, since it appears to better exploit the statistical idle times in conversations.

CDMA is a digital wideband, *spread spectrum* technology that transmits multiple independent conversations across single or multiple 1.25 MHz widebands of radio spectrum. Each voice, data or fax transmission is assigned a unique spreading code that distinguishes it from other calls that share the same spectrum. The CDMA system features large cell radii and the highest capacity of any wireless technology, at least in theory [3].

Traditional uses of spread spectrum have been for military operations. Because of the wide bandwidth of a spread spectrum signal, it is very difficult to jam, difficult to interfere with, and difficult to identify. This is in contrast to technologies using a narrower bandwidth of frequencies. Since a wideband spread spectrum signal is very hard to detect, it appears as nothing more than a slight rise in the "noise floor" or interference level. With other technologies, the power of the signal is concentrated in a narrower band, which makes it easier to detect.

The maximum distance between the station and the users is a function of several factors, such as terrain profile, antenna gain, cable loss, and others. However, given the power used in typical CDMA systems and the frequency band, it is possible to achieve coverages of over 40 km in radius in open areas. Using the Okumura-Hata model [3] for various antenna heights and gain, a number of cell radii may be achieved as illustrated in Table 1 below (assuming a 1.9 GHz band).

TABLE 1. Cell size in function of environment, antenna heights and gain

Environment	**Small city**	**Small city**	**Suburban**	**Suburban**	**Open area**
BS antenna height (m)	30	50	50	50	100
BS antenna gain (dB)	20	20	20	20	20
Terminal antenna height	1.5	1.5	1.5	1.5	1.5
Terminal antenna gain	0	0	0	6	6
Max cell radius (Km)	**4**	**5**	**10**	**28**	**> 40**

2.1.1 CDMA Concept

CDMA is seen by many as a natural solution to overcoming the time and frequency limitations of other multiple access methods. CDMA systems work by allowing all signals (code channels) to combine and interfere (in a controlled manner) with one another in the same frequency band. The key to successfully combining transmissions is signal encoding. Each *code channel* is coded in such a way that it possesses a certain signature that allows its recovery. The signature is recognized by the intended receiver and the messages can be deciphered.

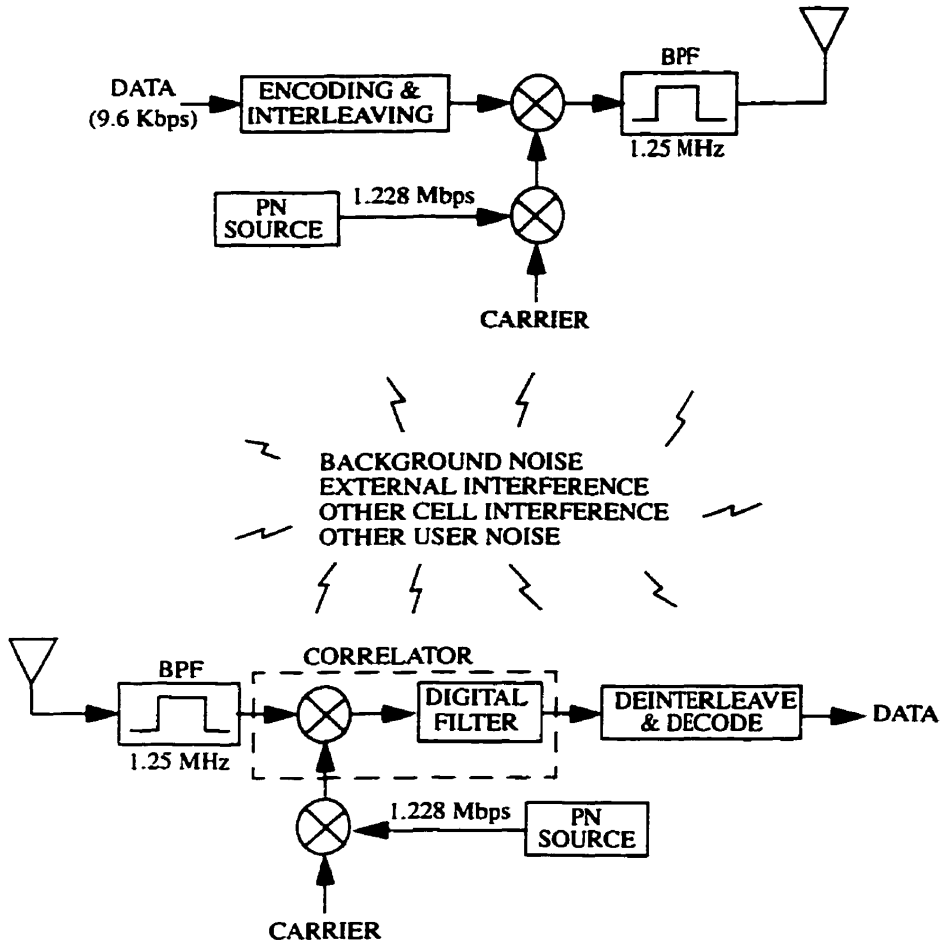

FIGURE 6. View of the CDMA concept

The signature is unique to each user, allowing a user's call to be distinguished from the many other calls being simultaneously transmitted over the same broadcast spectrum. The diagram in Figure 6 shows the concept of multiple users using the same channel. In the North American CDMA standard [IS-95], each signature is based on the use of two types of digital codes, Walsh codes and Pseudo Noise (PN) sequences. Assume that a call has a standard rate of 9600 bits per second as illustrated in Figure 6. This is then spread to a transmitted chip rate of about 1.23 Megachips per second. Spreading means that digital codes are applied to the data bits associated with users in a cell. These data bits are transmitted along with the signals of all the other users in that cell. When the signal is received, the codes are removed from the desired signal, separating the users and returning the call to a rate of 9600 bps.

2.1.2 Benefits and Advantages of CDMA

The current implementations of CDMA offer many benefits and advantages, making them very appealing to manufacturers and operators of cellular systems, as well as end-users. Some CDMA features are listed in Table 2.

TABLE 2. Advantages of CDMA

Feature	Benefits
Spread Spectrum	• Increased system capacity • Digital coding • Inherent privacy
Power Control	• Low interference • Extended talk/standby time • Increased system capacity • Smaller mobiles
Variable Rate Vocoder	• Improved voice quality • Voice activity detection • Reduced interference
Frequency Reuse	• Increased system capacity • Voice channel capacity • Frequencies need not be assigned to individual cells • All frequencies are available for use in all cells
Mobile Assisted Hand-off	• Improved hand-offs (soft and softer) • Improved voice quality
Multipath Processing	• Rake receivers • Minimized multipath fading • Improved voice quality

2.1.3 CDMA Power Control

Interference creates a practical limit to the capacity of a CDMA system. In CDMA, each user is a noise source in a shared spectrum channel. The higher the power level at which a user is transmitting, the higher the interference present in the shared channel. It is in everybody's best interest to keep the mobile and base station power levels as low as possible. Because CDMA is an interference limited system, power control is a crucial element in extracting the maximum possible capacity from the system.

The goal of power control is to keep each mobile unit at the absolute minimum power level necessary to ensure acceptable service quality. Mobile units that transmit excessive power increase interference to other units. The problem is particularly important in the near-far context, where a mobile close to the base station may mask the signal of a mobile far from the base. Through precise power control of mobile units, it is possible to equalize the received powers at the base and realize the following benefits:

- maximized system capacity, through minimizing intra-cell interference
- increased battery life, through minimizing the transmitted power of each mobile.

A number of power control techniques exist, as described in [4]:

Reverse Link Open Loop Power Control

This technique is used to compensate for sudden changes in the mean input power measured at the mobile station, and provides a mechanism to adjust to slow fading of the RF signal.

Reverse Link Closed Loop Power Control

This operation is performed to compensate for asymmetrical path loss (i.e., multipath, Rayleigh fading). The base station periodically monitors the strength of the signal received from the mobile station and then sends commands to the mobile to either increase or decrease the power level at which it is transmitting in order to achieve the desired level.

Reverse Outer Loop Power Control

This method allows the base station to control the output power level of the mobile based on the Frame Error Rate (FER) for the reverse traffic channel.

Forward Link Power Control

In the downlink, power control may be used to increase the system capacity [Lee91]. Without power control, the base station transmits the same amount of power to all portables. However, those mobiles which are near the base station receive a stronger signal from their own base and less interference from other bases. By reducing the power transmitted to these portables, yet satisfying the SIR requirements, the overall capacity is increased [Alavi82]. The decision to raise/decrease power is based on two indicators:

- periodic frame quality measurement sent from the mobile unit, and
- specific unit requests.

By monitoring the Frame Error Rate (FER), the mobile unit can send a request to the base station to increase the transmit power level when it senses an increased error rate. This method of power control is applied in a cyclic pattern to ensure that each mobile unit receives optimal power.

2.1.4 CDMA System for the Integration

The motivation for considering CDMA in a voice/data integrated wireless access network is three-fold. First, besides supporting several simultaneous transmissions, spread-spectrum techniques are known to effectively combat selective fading, jamming, and other in-band interference (narrowband or broadband). Second, they do not require stringent synchronization (at the bit-level) and provide selective addressing/reception and capture capability. Third, they enable the addition of new users to the network with a graceful degradation in the performance of the currently served users. The last two features of CDMA are very well suited for application in integrated voice/data networks. The CDMA techniques proposed for the integrated wireless access network follow those of the IS-95 standard. A pilot is used for the downlink which serves for chip-level synchronization. The bits, covering a significantly larger time interval, are thereby synchronized as well.

2.2 IWAN

The Integrated Wireless Access Network (IWAN) represents the research focus in the area of mobile and personal communications of the Canadian Institute for Telecommunications Research, an inter-university collaborative research effort supported by the Natural Sciences and Engineering Research Council of Canada and several industrial affiliates. In this network, Direct-Sequence Code Division Multiple Access (DS-CDMA) is employed. It is a model for a third-generation wireless access system that can support a wide variety of wireless services at different and variable transmission rates in both indoor and outdoor environments [Mermelstein93]. Its design has been motivated by the anticipated need to support a significantly increased market for data and multimedia services while also accommodating the currently more important market for voice services. Cost-effective use of the available spectrum in the PCS band near 1.9 GHz suggests that all services share the spectrum in a manner that permits the allocation of bandwidth on demand on a per-call basis. By not reserving radio resources for particular services, but allocating resources to individual calls as needed, maximal resource utilization consistent with the quality requirements of the particular services is ensured.

2.2.1 Services

The voice and data users have different traffic characteristics and requirements. Voice requires real-time delivery, i.e., negligible delay, but can accommodate moderate bit error rates. Data does not require a real-time delivery, i.e., it can be queued, but requires low bit error rates. The admission and flow control policy must take these features/requirements into account.

The performance measures are the average blocking probability of the voice calls and the average delay and packet loss probability of data messages as functions of the offered voice and data traffic loads.

Stream vs. Packet Protocol. The call admission algorithm found in [Kandala96] and already implemented in IWAN accommodates stream calls. At the outset, stream-based services are either admitted to the system or blocked based on their resource requirements, as shown in Figure 7.

On the other hand, the flow control protocol presented in this thesis is suitable for packet-based calls. Some incoming packet sources can still be active even when the channel is at capacity as shown in Figure 8. In this scenario, some packets (D_1 and D_2) are transmitted while others (D_3) are delayed and rescheduled.

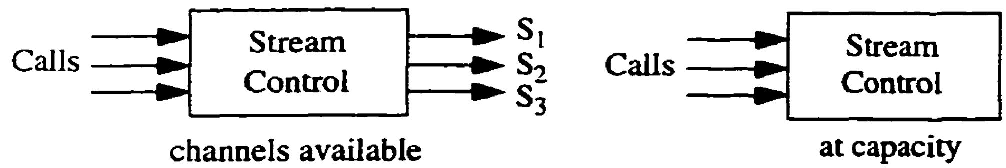

FIGURE 7. Call admission protocol for the stream-based services

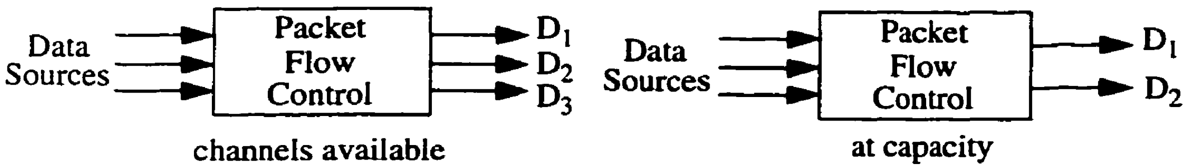

FIGURE 8. Flow control protocol for the packet-based services

Model for Voice traffic. For stream-based services, a call is continuously transmitted over a relatively long duration. The population of voice users is assumed to consist of maximum of M_v users. The traffic generated by each voice user is modeled as a two-state discrete-time Markov chain with the same parameters. Although voice is a stream-based service, it is modeled as being intermittent with a certain duty factor α, in which conversational speech is characterized by periods of activity called talkspurts (ON) and periods of silence (OFF).

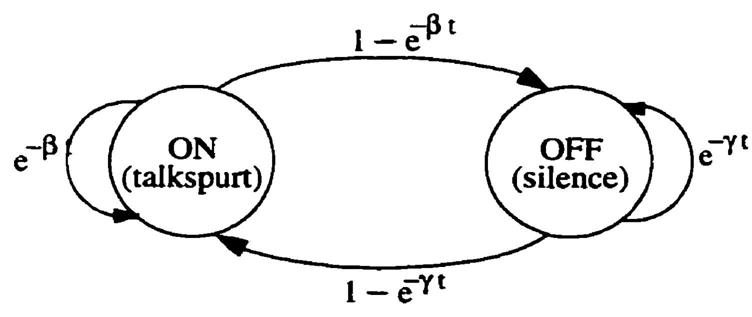

FIGURE 9. Two-state Markov model for voice activity

The transition probabilities between the two-states (ON/OFF or idle/active periods) at any packet slot are γ_v and β_v, as shown in Figure 9. Thus, the steady-state probability of k active voice users is

$$P\{\text{k active voice users}\} = \binom{M_v}{k}\left(p^v_{active}\right)^k \cdot \left(1 - p^v_{active}\right)^{M_v - k}, \tag{1}$$

where

$$p^v_{active} = \frac{\gamma_v}{\gamma_v + \beta_v}, \qquad p^v_{idle} = 1 - p^v_{active}. \tag{2}$$

The mean duration of the idle and active periods are $1/\beta = 0.25$ sec and $1/\gamma = 0.305$ sec (in packet slots); the integer parts should be taken here to ensure that an integer number of packets results. The probability that the duration of a voice call is l packet slots is

$$P\{\text{voice calls lasts for l slots}\} = (1 - \beta_v)^l \beta_v. \tag{3}$$

Mobile terminals incorporate a voice-activity monitoring technique to exploit this burst source model. This implies that the transmitter is not active or is transmitting at reduced power (and rate) during silent periods in human speech [Guo94], [Lee91], as is done in IS-95 systems.

The voice calls are of variable rate with a peak information rate of 8 Kb/s and a voice activity of $\alpha = 0.25 / (0.25+0.305) = 0.45$ or 45%. Variable rate coding involves use of a lower rate than the specified peak rate when the signal characteristics permit, such as coding the background noise at a lower rate in the absence of speech activity. The power control channel as well as other overhead increases the peak source rate to 9.6 Kb/s. The processing gain for the 5 MHz IWAN for 9.6 Kb/s voice calls is 512. For the 1.23 MHz IS-95, it is 128 [IS-95].

Model for Data traffic. The model for data traffic considered in this project is appropriate for downlink in CDMA wireless network scenarios; for example, it is suitable for the link from the base station to the mobiles for cellular networks. The key requirement for the applicability of the data traffic analysis is the concentration of data traffic at the base station prior to transmission using CDMA.

A packet data call consists of a burst of packets with a short service duration. The amount of data to be transmitted during a packet call is exponentially distributed with mean $1/\mu_d$.

The number of arriving data packets (from all data users) is assumed to be Poisson distributed with mean rate λ_d (in packets per time slot). The packet length is fixed and the transmission time of one packet is D. It is assumed that the base station is equipped with a buffer of size K (in packets) for the data packets accumulated (concentrated) there. All data packets are first stored in buffers and then transmitted at the next time slot.

If a packet is blocked by the flow control scheme or incurs an error on the channel, it remains in the buffer to be retransmitted after T seconds. Although a very simple model, this is useful in obtaining insight into the effect of flow control and its benefits. The data model is adequate for services such as file transfer, e-mail, and store-and-forward facsimile. The results can be extended to other data models. A Poisson model is believed to represent short-message service (SMS) very well [Sampath97]. Although the analysis gets complicated in this case, the qualitative results from the simple data model still hold. Interactive data services can be modeled as a queue of packets at each source with an arrival process into the queue. All results from the fixed data model directly apply in this case, with an additional stability condition to ensure that none of the queue lengths become unbounded.

Since voice calls have priority over data users, the number of successfully transmitted data packets strongly depends on the number of transmitted voice packets at each time slot. In other words, the number of servers available for data packets is dependent on the number of servers already occupied by voice calls. The processing gain for the 5 MHz IWAN for 9.6 Kb/s packet data calls is 512. For the 1.23 MHz IS-95, it is 128 [IS-95].

Packet data from Asynchronous Transfer Mode (ATM) networks is carried up to a 512 Kb/s transport rate. Lower rate synchronous services, including facsimile transmission to/from wireline fax terminals, are carried at transport rates of 4, 8, 16 or 64 kb/s according to the needs of the particular service. A 20% overhead is typically used to maintain synchronization and power control for idle intervals. Terminals lose access if no data is transmitted for an interval longer than a specified threshold.

Model for Interference. Direct-Sequence Code Division Multiple Access (DS-CDMA) is employed by all users in the network. The same frequency band is shared by voice and data traffic. Voice and data traffic have the same data rate; thus, the same pool of CDMA codes is used. This assumption can be relaxed and a truly multirate system can be considered, but is beyond the scope of this thesis. Each user (data or voice terminal) employs a distinct code for transmission.

It should be noted that the data rates given above are not the actual line rates. The highest data rate considered for the time being is 76.8 kb/s. With a chip rate of 9.83 Mchips/s, the processing gain would be 128 for the 76.8 kb/s service. Lower rates are supported at half multiples of the higher rate. These rates are 76.8, 38.4, 19.2, 9.6, 4.8, 2.4 and 1.2 kb/s. The lower rate services may be handled in two ways. One approach would be to double the symbol duration but reduce the transmitted power by half - the continuous symbol transmission (CST) method. This would achieve the same E_b/N_0 for all services. Another approach would be to transmit at the same power and the same processing gain, but at half the symbol rate - the discontinuous symbol transmission (DST) method. In other words, with DST, the lower rate services are transmitted at only half the symbol positions. Capacity has been computed assuming the CST approach. The capacity of the DST technique is expected to be lower than that of CST. However, DST is appropriate for receiver architectures that conserve power. The performance of DST is under investigation.

2.2.2 IWAN Project Goals

The principal goal of the IWAN project, as mentioned in [Mermelstein97], is the development of a set of design principles to maximize the number of calls that can be accommodated in a given bandwidth subject to the performance requirements of each call. The calls are assumed to minimize their resource requirements by using as low a transmission rate as is appropriate for the momentary signal characteristics. Effective bandwidth management techniques ensure that as many calls as possible are admitted, without exceeding the network capacity. Congestion control is used to terminate calls, if necessary, to ensure that adequate quality is maintained for the remaining calls. The wireless network is interconnected both to the public switched telephone network (PSTN) and to the broadband packet (ATM) network.

2.2.3 Characteristics of Transmissions

Most services can be classified as requiring one of two types of transmission, stream or packet, according to their sensitivity to retransmission delays in the case of uncorrectable transmission errors. Real-time services such as speech or video with strict delay limitations are transmitted as stream services, divided into transmission frames. Sufficient error-correction capability is provided so that at most a very limited number of retransmissions is required to achieve the desired grade of service. To achieve the very low error rates required for data transmission, packet transmission is used with an ARQ protocol. The time-varying nature of the wireless channel makes the need for such retransmissions sufficiently rare that, on average, retransmission requires fewer resources than operating constantly at the SNR required for a sufficiently low packet error rate.

Transmission capacity in an interference-limited wireless environment varies with the propagation conditions as well as the traffic in the surrounding cells. While it is possible to define a maximum number of calls that can be accommodated, appropriate to a situation with no traffic in the neighborhood and all calls transmitted at the lowest available rate, such a situation is rarely encountered in practice. The most general approach to prevent system overload is to monitor the radio resources available after ensuring that all calls meet their required grade of service. In CDMA systems with fast power control in both directions, power control provides the necessary information for control of congestion and call admission. To gain admission, a call's resource requirements (possibly unequal) must be satisfied in both directions:

- On the uplink, if all the calls experience acceptable signal-to-interference ratio conditions with a protection margin adequate for the requirements of a new call, that call can be admitted.
- On the downlink, the corresponding measure is the total power transmitted by the base station. If a sufficient power margin is available so that neither the power requirements

in the cell to which the new call is directed nor the power requirements in the neighboring cells would be exceeded, the downlink segment is admitted as well.

2.2.4 Design Principles of the IWAN

The design of the IWAN hinges on the need to conserve radio resources while maintaining the grade of service required for each service. Additional delay is permitted to compensate for the higher error rates of the wireless environment whenever this is acceptable to the user in terms of the quality of the service.

To retain maximum flexibility in selecting the transmission rate most appropriate to a particular call, the processing gain is allowed to vary inversely with the transmission rate, both varying by a factor of two. This permits use of a fixed chip-rate appropriate to the available bandwidth. For time-varying calls, the processing gain is selected according to the peak transmission rate specified. To maintain the same energy/chip, yet reduce the average transmitted power at the lower rates, chips are dropped by a pseudo-random selection in time.

One consequence of using fewer chips per bit for calls with higher peak rates is that the receiver sees signal powers proportional to the bit-rate. The range of bit-rates that can be simultaneously received in a practical receiver remains to be determined. Each branch of the receiver is intended to be reconfigurable to all rates. Thus, the complexity of the base station receiver will be controlled by the number of correlation branches, which corresponds to the maximum number of calls that can be simultaneously accommodated in one cell or sector.

It is important that calls with time-varying transmission rates be able to change rate rapidly without regard to the network conditions. Thus, the momentary rate of a call is selected by the source or interface with no interaction taking place with the system controller. For voice services, this ensures that no speech is lost or degraded when a party starts talking after a silent interval. For video services, the transmission can accommodate the higher coding rate necessitated by the presence of motion in the image field immediately, i.e., with a delay less than the frame interval.

Capacity analyses for mobile cellular applications with voice service have been presented in [Gilhousen91]. Capacity estimates as a function of increasing bandwidth were presented in [Jalali94].

2.2.5 Downlink part of IWAN

Pilot-signal assisted coherent reception is assumed. The user data is encoded using a rate 1/2 convolutional code followed by a bit interleaver. The interleaved bits are spread onto the in-phase and quadrature channels using different PN sequences and are transmitted using QPSK. The PN codes used within the same cell for the different users are orthogonal irrespective of the data rates used. Soft decision Viterbi decoding is employed at the receiver. The total power available for downlink transmission is divided between the pilot signal and all active mobiles. A frame length of 5 ms is considered. The real and imaginary components of the channel impulse response are estimated by averaging the correlator outputs over the power control sampling interval of T_p ms (24 coded symbols when T_p=1.25 ms) [Jalali94].

2.2.6 System Model

Traffic is classified in terms of throughput/delay requirements. The generic characteristics of each traffic type are described below. Speech needs to be assigned traffic channels permanently to avoid any delay. Data packets, on the other hand, can be delayed.

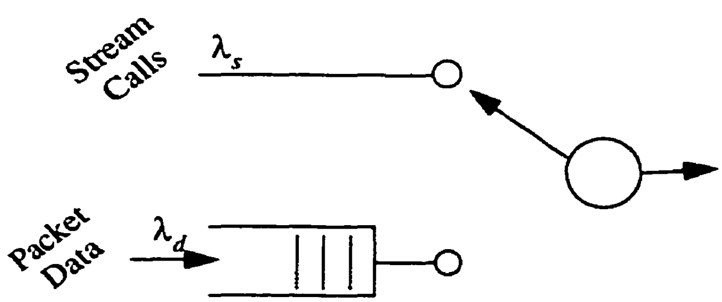

FIGURE 10. Voice has priority and no buffering

In this system, voice has priority over data with no buffering for voice (see Figure 10). Each voice call alternates between talkspurt and silence, which are assumed to be exponentially distributed with different mean lengths. Voice calls are generated at a constant rate λ_s and the average call duration is $1/\mu$. Data packet lengths or, equivalently, data packet transmission times are constant. The arrival processes of voice calls and data packet arrivals are assumed to be Poisson processes with different mean arrival rates.If the number of stream calls is higher than the capacity of the channel, any arriving voice calls are discarded. An infinite buffer is assumed for data. Some studies were done [Kandala96] to find the call-loss rate when no packet data are present. One of the objectives of this research is to find the effect of the integration of packet data with voice on the call-loss rate. Hence, the main focus will be on packet data transmission when stream calls are admitted to full capacity. A question is now raised regarding the feasibility of transmitting data packets when the base transmitted power is fully used, i.e., when the total transmitted power exceeds a certain threshold.

2.2.7 BER

The BER required to provide acceptable quality in speech calls is assumed to be 10^{-3}. For packet data calls, a BER of 10^{-6} is considered. A mobile is in outage if the probability that the BER exceeds 10^{-3} (or 10^{-6} for packet data) is greater than 1%. The average capacity of a cell is defined as the average number of mobiles per cell that can be accommodated in a 5x5 square grid when the mobiles are distributed spatially uniformly. The required E_b/N_0 to achieve a BER < 10^{-3} for the downlink is 5.0 dB at a Doppler of 2Hz when transmit diversity of order 3 and fast power control are used [Jalali94]. A 7.5 dB E_b/N_0 is suggested by [Sampath95] and [Guo95] to achieve a BER < 10^{-6} in the packet data case.

2.2.8 Power Control and Resource Management

For spread spectrum systems, power control is necessary to combat the near-far effect and has been shown to increase capacity. If, rather then using constant power, the transmitters can be controlled in such a way that the received powers from all users are roughly equal, then the benefits of spreading are realized. If the received power is controlled, then the subscribers can occupy the same spectrum, and the hoped-for benefits of interference averaging accrue. In addition to capacity increase, power control can also be used to achieve target quality of service in a multimedia environment [Sampath95].

For multimedia systems, both the power of the users and their transmission rates may be considered as controllable resources. In a mixed traffic environment, the performance of each class is a function of their resource budget. For example, a small portable transmitting data may have very tight power limits due to battery size, but very loose delay constraints. On the other hand, a voice user will have strict rate requirements, as compared to power requirements. In order to achieve their objective, they can alter their power and/or rate of transmission. In the context of CDMA, different transmission rate requirements can be modeled as variable processing gains or variable forward error correction (FEC) code rates. Typically, it is desirable to keep the chip rate for all classes of users the same, so that they all occupy the same bandwidth [Wyrwas94].

Different services have different quality of service (QoS) requirements, maximum power and/or minimum rate constraints. In order to achieve the required QoS, they can alter their power and/or rate of transmission. Rates can be varied using different processing gains. It is up to the power control mobile to find a power assignment that satisfies the QoS requirements of all users under the existing fading conditions and current mobile locations.

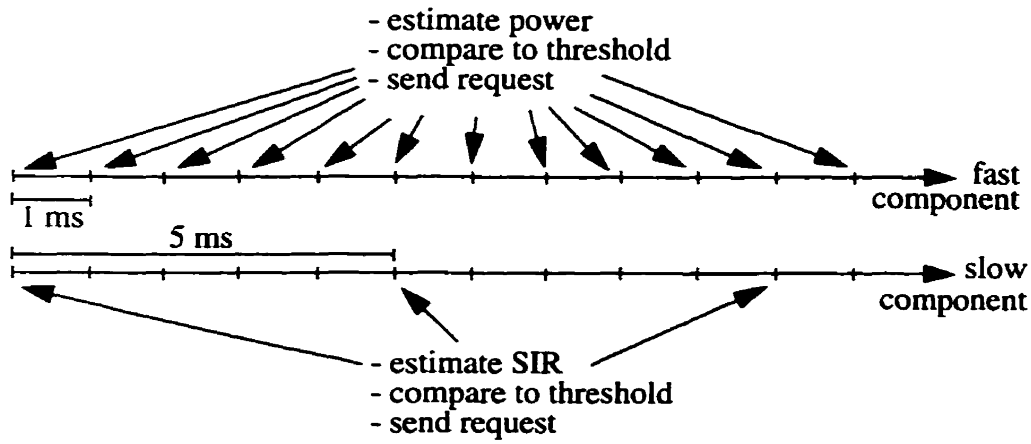

FIGURE 11. Power control components

The scheme for power control considered here consists of two components - a slow and a fast component as shown in Figure 11. For the fast component of the power control algorithm, the mobile estimates the received signal power at intervals of every 1 ms, compares the received signal power to a threshold, and transmits a power control command bit to the base station to request an increase or decrease in the transmitted power. The fast component of the power control algorithm is intended to mitigate fast fading at low Doppler frequencies. For the slow component of the power control algorithm, the mobile estimates the average received signal-to-interference ratio (SIR), compares the estimate to a threshold and transmits a request to the base station, for an increase in the transmitted power if SIR is below a threshold ($(SNR)_{low}$), and for a decrease in the transmitted power if the SIR is above another threshold ($(SNR)_{high}$). Refer to Table 3 for different SNR requirements. The slow power control commands are only sent when the SIR is outside the above two thresholds and are sent at the frame rate (5ms). The power is always adjusted by a fixed step size Δ for all mobiles at all rates. The parameter Δ is chosen to be 0.5 dB. Exploratory experiments of assigning a higher value of Δ to high rate cases did not lead to any improvements. In the simulations, however, the fast power control need not be simulated, as the effects of the fast power control can be included in the required E_b/N_0 [Jalali94]. In a wideband system with a large number of users, the variations due to fast

TABLE 3. Required SNR vs. BER

	BER < 10^{-3}	BER < 10^{-6}
$(SNR)_{low}$	6.5 dB	7.0 dB
$(SNR)_{high}$	7.5 dB	8.0 dB
Δ	0.5 dB	0.5 dB

power control from different users will cancel each other out when computing the average SIR. In other words, the fast variations will manifest themselves as an increase in the average interference by a certain amount. One way of taking the fast variations into account is to introduce the probability distribution on the required E_b/N_0. The required E_b/N_0 by a mobile will be generated according to a probability distribution, where that probability distribution is obtained by simulating the fast power control off-line. Alternatively, instead of incorporating the probability distribution, a margin of 1 dB is added to the required E_b/N_0. If the power requirements from the mobiles in a cell cannot be met due to the limitations of the maximum available power at the base station, the powers of all the mobiles are proportionally reduced.

2.2.9 Channel and Propagation Model Assumptions

Although the fading is not Rayleigh, the Rician fit is also quite poor at low signal levels which are of interest to us. Therefore, in order to take into account the worst case situation, the channel is modelled by a single Rayleigh fading path for 5 MHz systems. An artificial multipath of order 3 is considered to provide adequate fading mitigation, i.e., three omni-directional antennas are present at the base station.

As far as the shadow fading and path loss exponents are concerned, no consensus is found in the literature [IWAN]. Therefore, the generally accepted model of log-normally distributed shadow fading with standard deviation of 8 dB is used. A path loss exponent of 4 is considered.

2.2.10 Power Allotment

When the base receives a power control command from the mobile j, the transmitted power to mobile j is changed by an amount of Δ_j dB, where

$$\Delta_j = \begin{cases} \Delta & \text{if j requests more power} \\ -\Delta & \text{if j requests less power} \\ 0 & \text{if j does not make any request} \end{cases} \tag{4}$$

Hence, the new transmitted power that is to be used is given by

$$p_j' = 10^{(\Delta_j/10)} p_j. \tag{5}$$

Let $P = \sum_j p_j'$. If P is less than the maximum available power M, then the base transmits to the jth mobile with power p_j'. If P exceeds M, however, then all the power requirements cannot be satisfied. Hence, the powers of all the users are re-computed using the formula

$$p_j'' = p_j' \times \frac{M}{P} \qquad \forall j. \tag{6}$$

Thus, the actual transmit power to mobile j is given by

$$p_j'' = min\{\frac{P_{max}}{P}, 1\}\ p_j'. \qquad (7)$$

which limits P to P_{max} and reduces power proportionally so as not to exceed P_{max}. Note that this power assignment does not assume priority to any of the calls and, hence, all the powers are reduced proportionally when the total power requests exceeds the maximum available power. It is possible to prioritize the calls and assign the powers based on those priorities.

2.2.11 ARQ

In our system, selective-repeat ARQ between the base station and the mobile is used. The transmitter only resends (or repeats) those packets that are negatively acknowledged as shown in Figure 12. Since, ordinarily, code packets must be delivered to the user in correct order, a buffer must be provided at the receiver to store the error-free received packets following a received packet detected in error.

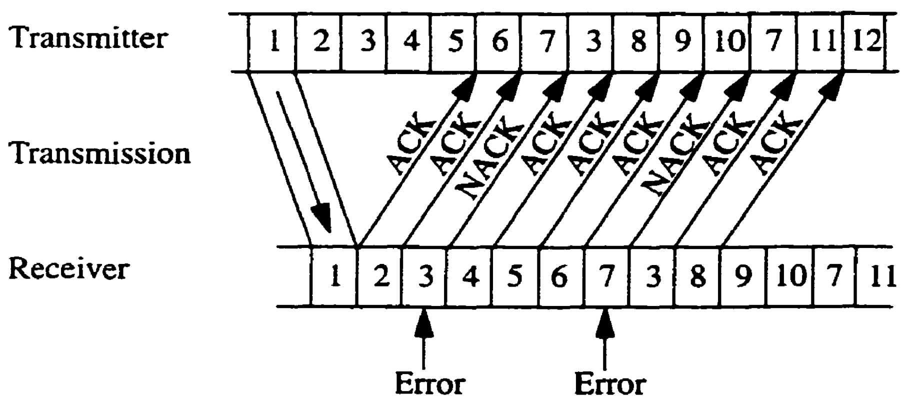

FIGURE 12. Selective-repeat ARQ

When the first negatively acknowledged packet is successfully received, the receiver than releases the error-free received packets *in consecutive order* until the next erroneously received packet is encountered. Sufficient receiver buffer must be provided; otherwise, buffer overflow may occur and data may be lost.

3. Flow control for CDMA based wireless data

CDMA voice systems have been studied in great detail (e.g., [4] and [2]). The throughput, delay, and stability characteristics of CDMA random access for data have also been evaluated. It is only in recent years that CDMA systems with multiple services have received attention [Guo94], [Sampath95]. In integrated voice/data CDMA systems, like IWAN, the capacity for data calls can be increased by allowing data transmissions in periods of low voice load and curtailing data transmissions when the voice load is heavy. While such a strategy does improve capacity, data users have to wait to transmit, and thus incur packet delay. Flow control schemes of this nature have been proposed in [Dziong96], [Yang94],[Guo94], [Mandayam95], and [Sampath97].

The chapter is divided as follows: first, the problem, its justification, and motivation are explained in detail. Then, a comparative analysis between the flow control proposed here and previous methods is given.

3.1 Problem Statement

ATM is used to support multi-media services and for CDMA, in particular, it provides efficient switching network for call control and resource management. The architecture for services provided at the ATM layer consists of the following five service categories [ATM95]:

- Constant Bit Rate (CBR)
- real time Variable Bit Rate (rt-VBR)
- non-real-time Variable Bit Rate (nrt-VBR)
- Available Bit Rate (ABR)
- Unspecified Bit Rate (UBR)

The CBR and VBR service categories were intended to support applications with precisely defined requirements for throughput and delays. By contrast, the UBR and ABR service categories were intended to support traditional computer communications applications like file transfer that are bursty and do not require tightly constrained delay and delay variation. These service categories increase the bandwidth utilization without effecting the QoS of the CBR and the VBR connections.

The ABR service is defined for the support of applications that may require minimum bandwidth. It guarantees a low or zero cell loss ratio and a fair share of the available bandwidth to an end system that adapts its traffic in accordance with the feedback received from the network. Such a service is usually required by data applications that cannot predict their own bandwidth requirements and expect to get some share of the available bandwidth. By contrast, the UBR service category offers best-effort service with no QoS guarantees. It does not include the notion of a per-connection negotiated bandwidth and does not make numerical commitments with respect to the cell loss ratio or the cell transfer delay experienced on the connection. It is designed for data applications that want to use any left-over bandwidth and are not sensitive to cell loss and delay. A UBR connection is not rejected on the basis of bandwidth shortage and is not policed by the network.

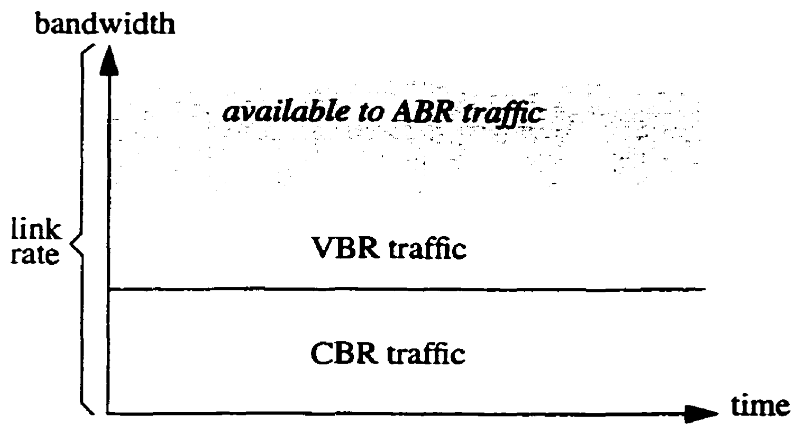

FIGURE 13. Link Bandwidth Sharing

The ABR service allows applications to utilize the bandwidth not used by the CBR and VBR services, by continually adjusting their instantaneous transmission rates to the time-varying capacity available for ABR. The relationships are shown in Figure 13.

A congestion control scheme is essential for the support of ABR traffic to utilize the available bandwidth without causing congestion and to divide that bandwidth among the ABR connections in a fair manner. Computing the fair share of an ABR connection on an outgoing link of a switch requires an algorithm for fair allocation. Another algorithm is required at the end-system to adjust its rate to the feedback received from the network. The proper source behavior is important to guarantee the low cell loss rate. If a source does not behave as expected, it could increase the loss rate to all ABR users.

3.2 Objectives

Power control and multipath diversity are among the major issues in system design which significantly affect the capacity of a CDMA system. Capacity is defined as the number of simultaneous variable rate calls that can be supported without exceeding a given probability of outage. Other factors which have a significant impact on the capacity are the system bandwidth and the path loss exponent.

The IWAN goal is total integration of voice/data/video at different and variable bit-rates on a wireless multimedia network [Mermelstein93] as shown in Figure 14. The integration of voice and video services in microcellular CDMA systems was reported in [Kandala96]. The voice calls considered in the system have variable rates with a peak information rate of 8 Kb/s and a voice activity factor of 0.45. The power control channel as well as other overhead increases the peak source rate to 9.6 Kb/s. The video calls, on the other hand, are of fixed rate of 64 Kb/s with full activity and a proportional overhead on 64 Kb/s resulting in a source rate of 76.8 Kb/s. Video calls are not used in this study.

The immediate goal addressed in this thesis is the integration of voice and packet data. The associated multimedia terminals may operate in different modes- (a) voice only, (b) packet data only or (c) both. In this discussion, a terminal is assumed to operate either in mode (a) or (b), but not (c). No reservations are made for traffic in any mode. Calls are admitted according to their resource requirements and the available transmitter power at the base station and its neighboring cells. The admission criteria are set to minimize blocking, yet limit the possibility of overload if the call is admitted. Specific objectives include the efficient provisioning of packet data services, the integration of these services with voice in a bandwidth-on-demand, fair-sharing manner, and the study of the short-time transmitted power statistics to determine the available capacity for packet data transmission when stream calls have been admitted to capacity.

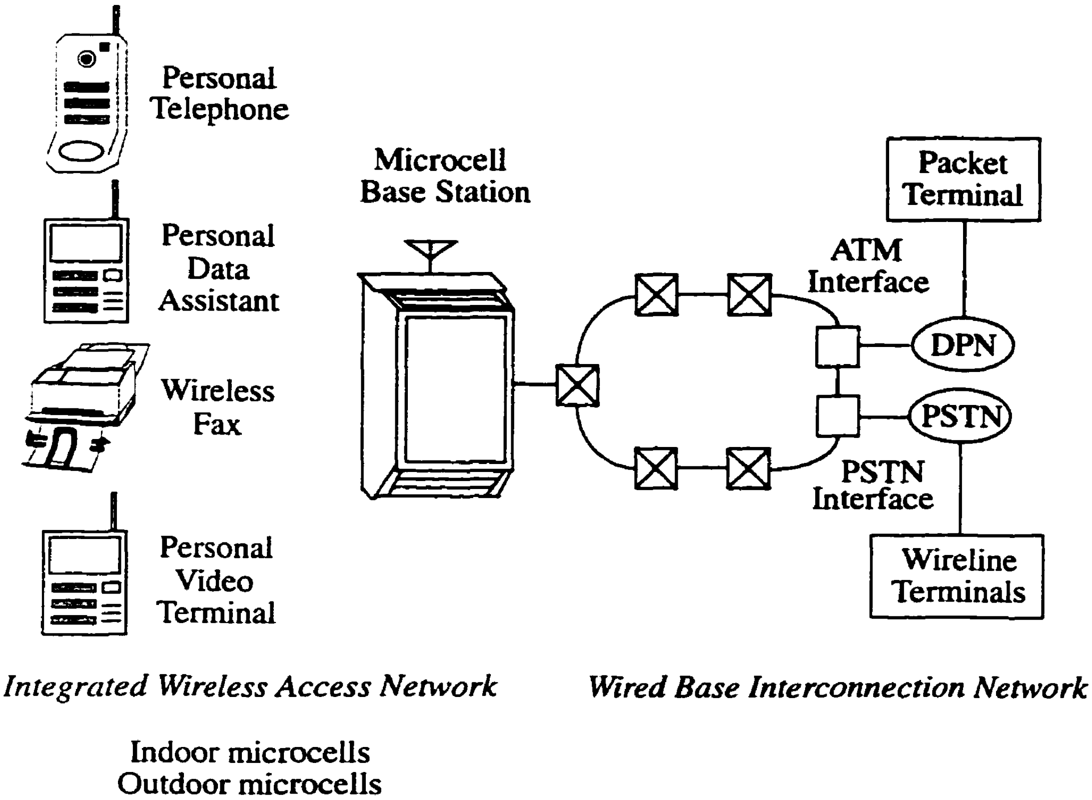

FIGURE 14. Integration of services over shared bandwidth

We consider flow control in the downlink direction for a microcellular CDMA system carrying integrated packet level-data and voice services with the objective of providing maximum traffic capacity without service degradation due to overload. The power transmitted by the base and the interference seen by the receiver at the base station provide ready indications of the traffic conditions in each cell.

3.2.1 Design Objectives

The purpose of this thesis is to provide a flow control mechanism that satisfies the following key design objectives:

1. For many higher layer protocols, a single lost ABR cell can cost the entire message to be abandoned and retransmitted. The flow control algorithm should, therefore, ensure virtually zero cell loss (due to buffer overflow) for Available Bit-Rate (ABR) traffic.
2. The algorithm should be able to operate with a small buffer size requirement, low management complexity, and low signalling overhead.
3. It should be highly responsive to changes in the available bandwidth. Almost 100% link utilization should be achieved if the ABR sources always have data to transfer.
4. The available bandwidth should be divided fairly among all competing active ABR virtual circuits.
5. The performance should be insensitive to the incident traffic pattern of the controlled traffic.
6. The algorithm should be scalable with distance, link speed, and number of connections.

3.3 Motivation for Integration

Voice telephony requires continuous bit-stream-type service with no delays, while data are best suited for discontinuous packetized transmission and can tolerate moderate delays. Voice services can tolerate higher error rates than data can. These basic differences have implications from lower layer system design all the way up to network management [Raychaudhari94]. Design of good dynamic control algorithms is also of great importance for good system performance [Mermelstein93]. These include algorithms for admission control, resource allocation, error recovery, and flow control at the burst or packet levels. The main idea behind the integration of voice and data traffic in a single network has been the use of hybrid circuit/packet switching techniques ([Coviello75] and [Wu88]).

Available Bit Rate (ABR) has been attracting attention as a new service category of ATM. ABR is a suitable class for supporting highly-bursty data applications on ATM networks.

These applications have cell loss requirements to achieve good throughput. In ABR, bandwidth is not statically allocated. Instead, the Peak Cell Rate (PCR) and Minimum Cell Rate (MCR) are specified at the call set-up, and the cell emission rate of each connection is controlled between the PCR and MCR based on feedback information from the network. This feedback control makes it possible to effectively utilize bandwidth without cell loss.

By designing an efficient flow control protocol, the capacity is better exploited; and one can potentially transmit useful volumes of packet data even when no fixed mean data transmission rate can be guaranteed. Also, when the momentary interference conditions are low, resulting, perhaps, from a concentration of mobiles near the base station, substantial packet transport capacity may be available. The exploitation of this residual capacity without degrading admitted stream services requires continuous rapid monitoring of the interference conditions.

3.4 Previous Approaches to Flow Control Protocols

Many publications have dealt with bandwidth utilization in CDMA networks, e.g., ([Gilhousen91] and [Yang94]). Most of them have focused on evaluating the system capacity under stationary, uniform, and homogeneous traffic conditions. An interesting exception is the study presented in [Liu94], where two traffic admission algorithms are compared under nonuniform traffic conditions. In the first algorithm, each base station utilizes only the local signal-to-interference ratio (SIR) for traffic admission decisions. The second algorithm utilized the SIR's from neighboring stations as well. In both algorithms, the SIR's are assumed to be known and fixed power control is assumed.

In [Dziong96], a flow control algorithm was proposed for the integration of packet data calls over voice calls in the IWAN. This study differs from [Liu94] in many aspects, the most important being the estimation of the mean and variance of the interference based on measurements and source declarations, inclusion of adaptive power control, and analysis under nonstationary traffic conditions.

Also in [Dziong96], three information flow categories are proposed. They allow one to create several alternatives for traffic admission algorithms which will differ in efficiency and complexity:

1. *Fixed Strategy (FS):* The admission decision is based on traffic source declarations only. The bandwidth manager ensures that the declared traffic parameters do not exceed the *thresholds* $A_T(A_d)$. The *thresholds* are defined so that the information BER constraint is met for any source and propagation distribution

$$A_d \leq A_T(A_d) \Rightarrow P\{BER_s > \varepsilon\} \leq \gamma. \tag{8}$$

The obvious advantage of the scheme is its simplicity. On the other hand, the algorithm has to be designed for the worst case of source location and propagation distributions which are not under the base station control so the bandwidth utilization can be far from optimum.

2. *Local Adaptive Strategy (LS):* The admission decision is based on both the source traffic declarations and the channel characteristics measured at the local base station H. In this case, the *threshold* A_T adapts to the current propagation conditions at the local base station

$$A_d \leq A_T(A_d, H) \Rightarrow P\{BER_s > \varepsilon\} \leq \gamma. \tag{9}$$

It is clear that the adaptive capabilities of this scheme can provide better bandwidth utilization. At the same time, monitoring of the transmission conditions increases the complexity of the algorithms.

3. *Global Adaptive Strategy (GS):* This is an extension of the local adaptive strategy, where the decision is also based on information from the neighboring base stations, G

$$A_d \leq A_T(A_d, H, G) \Rightarrow P\{BER_s > \varepsilon\} \leq \gamma \tag{10}$$

In this case, the level of traffic accepted by the local base stations takes into account the propagation conditions in a larger region. This is achieved at the expense of the overhead associated with the exchange of information between the base stations.

In [Yang94], data users have lower priority over voice users. Voice packets are transmitted first and use all the CDMA codes they need (provided their number is smaller than the multiple access capability for voice calls). If, after these allocations, there still are codes available for the data users, then the data packets are transmitted; otherwise, no data packets are transmitted at all and they are queued or lost (when the buffer size is exceeded). From this, we see that it makes sense to choose a *threshold* admission policy for data. The *threshold* value depends upon the voice traffic load, admission policy, and multiple reception model. According to this admission policy, the *threshold* data arrival rate λ_t which corresponds to a prespecified tolerable average data delay (or packet loss probability) is found and compared to the sum of the rate λ_i of data that is currently in the system and the rate of new data arrivals λ_a; in particular, the policy $R^d(\lambda_i)$ is

$$R^d(\lambda_i, \lambda_a) = \begin{cases} 1 \text{ (accept)}, & \text{if} \quad \lambda_i + \lambda_a \le \lambda_t \\ 0 \text{ (reject)}, & \text{if} \quad \lambda_i + \lambda_a > \lambda_t \end{cases} \tag{11}$$

On the other hand, in order to guarantee specified Grade of Service levels, [Guo94] suggests that the number of calls allowed access to the network must be limited. When a call (either voice or data) requests access to the network, the bit error rate after setting up the call is calculated for voice and data calls. If the bit error rate is higher than the specified grade of service, the call should be rejected or put into a queue and held until the network passes into a state with an acceptable bit error rate.

In [Mandayam95], an access protocol for short data messages in an integrated voice/data CDMA system was studied. It was assumed that voice traffic has priority and is less tolerant to delay than data traffic. Hence, the transmission of data traffic could be controlled to guarantee acceptable QoS for voice. The data messages were of short duration and as a result, closed loop power control for them was not possible. The access protocol was an extension of that found in [Viterbi94], which is for data-only systems. A slotted system was considered for data. At the end of every slot, the base station would control the transmission of data users by increasing/decreasing data permission probability. Measured QoS for voice was the criterion. An approximate model wherein a one-dimensional Markov chain of just the persistence state was considered and solved explicitly. Imperfect power control for voice was accounted for.

One other interesting technique proposed by [Sampath97] to control data access is by first predicting the number of active voice users in the next slot, based on past measurements; then, computing the permission probability for data in the next slot so that the target outage probability is met. The case of perfect prediction was first considered. This scheme, although physically unrealizable, provides an upper bound on possible gains through access control and serves as a benchmark for more practical access control schemes considered later. Another scheme considered is based on MMSE prediction of the number of active voice users. The MMSE prediction scheme requires knowledge of the voice activity parameters λ and μ for its implementation. These may not be accurately known. Also, if an adaptive model for the cumulative voice load is used, then the permission probabilities will have to be recomputed each time the model is changed and would add to the complexity. Another difficulty is that at the end of every slot, explicit load information or the permission probability itself has to be conveyed to the mobile. This can result in unacceptable overhead.

[Sampath97] presented a simple, real-time access control scheme originally proposed by Viterbi for a data-only system [Viterbi94] and based on the protocol presented by [Mandayam95]. The control ensures adequate performance for all users by dynamically varying the permission probability for data users. The permission probability is set through a nonnegative integer parameter called the persistence state. If, in the nth slot, the persistence state $T(n)$ is j, each data user, independent of other users, transmits with probability π^j and refrains from transmitting with probability $1 - \pi^j$. The value of π $(0 < \pi < 1)$ is fixed and known to the data users. The persistence state is broadcast to all of the users by the base station. The persistence state in the $(n+1)$st slot is assigned as follows:

$$T(n+1) = \begin{cases} T(n) + K, & \text{if } S(n) \geq \Omega \\ T(n) - 1, & \text{if } S(n) < \Omega \end{cases} \tag{12}$$

where $\Omega \leq 1$ is the *threshold* used to trigger access control. The combined choice of K, π, and Ω is critical in obtaining the desired performance from the control scheme, as discussed in Section IV (E) of [Sampath97]. The case of $K = 1$ and short-message data services is treated in [Mandayam95] and [Viterbi94].

The steps in access control based on prediction at the end of the *n*th slot are as follows:

1. Measure $V(n)$, the number of active voice users in the *n*th slot.
2. Predict the number of active voice users in the $(n+1)$st slot

$$\hat{V}(n+1) = \begin{cases} V(n+1) & \text{for perfect prediction} \\ E\{V(n+1) \mid V(n), V(n-1), V(n-2), \ldots\} & \text{for MMSE prediction} \end{cases} \tag{13}$$

3. Compute permission probability for data.

3.5 Description of Flow Control Algorithm

The main objective of this project is to design close to optimal flow control algorithms, with respect to bandwidth utilization under QoS constraints, for each of the generic alternatives and to assess the trade-offs between the call-capacity utilizations and the algorithm complexities.

First, consider a cell with N_V voice calls and N_D packet data calls operating well under capacity, i.e., none of the mobiles are in outage and the total base transmitter power is well below the maximum power that is available. Let there be a new call arrival into this cell. As none of the existing calls is experiencing outage and the total transmitted power is below the maximum allowable power M, it is therefore reasonable to admit this call. The newly admitted call will increase the total transmitter power from the base of this cell, increasing the amount of interference to the mobiles in the adjacent cells. This will lower the available SIR at these mobiles. If the SIR at all these mobiles is above the SNR needed to provide the desired grade of quality, then the effect of this new arrival is negligible. However, in some cases the resulting SIRs will be below the required SNR, thus necessitating increased power to be sent to these other mobiles, which in turn increases the transmit power of the adjacent bases.

Each voice call uses a specific preassigned CDMA code. In particular, new voice calls enter the system until the interference level in the channel passes a *threshold*. New call attempts may be blocked and cleared on the basis of the global information available. There is no buffering of voice calls. The new data packets join the backlog pool before their first transmissions are attempted, and each packet is independently transmitted. During transmission data users use their CDMA codes.

Basically, a flow control process is modeled as a discrete-time single server queuing system where a new data call joins existing calls. According to this protocol, a packet call joins the data buffer until reaching the head of the buffer which is referred to as the transmit buffer. Hence, when a data packet enters the transmit buffer, the transmitter will sense the level of the total transmitted power. If this power is fully loaded, the packet will wait until the following flow control decision is permitted. Decisions are made every 10 msec, time to transmit one packet data. The fact that the total transmitted power level allows for packet transmission does not guarantee its correct transmission, since the probability exists that the interference level can vary depending on the activity of stream calls, call duration and call arrival rate. Having failed to transmit successfully a packet, the packet is retransmitted when the transmitter again detects a low power level. However, using the 2nd order statistics of the total transmitted power, the number of retransmissions can be minimized significantly, as prediction of the power level is computed. Depending on the predicted level, a certain number of packets are transmitted.

3.5.1 Services' Priorities

Clearly, the fact that voice traffic has precedence over the data traffic will introduce an additional delay component in the overall average packet delay from the time the packet is accepted in the transmitted buffer pool to the time the packet is successfully transmitted. Consequently, in our analysis, we will seek to determine the total average packet delay as well as the average buffer occupancy as a function of the statistical characteristics of voice and the retransmission delay.In Figure 15, the service priorities are shown. The algorithm provides priority for the voice traffic at the expense of lower data throughput and higher delay for data packets. The priorities are as follows:

1. New voice calls
2. Buffered packet data
3. New packet data calls.

The *"Voice Call Processing"* or call admission algorithm and the *"Data Call Processing"* or flow control algorithm boxes are shown in figures 16 and 17. The call admission process is used to block calls that will degrade the QoS. On the other hand, flow control includes call blocking, packet discarding, packet blocking and packet scheduling. These algorithms are explained in Sections 3.5.2 and 3.5.3, respectively.

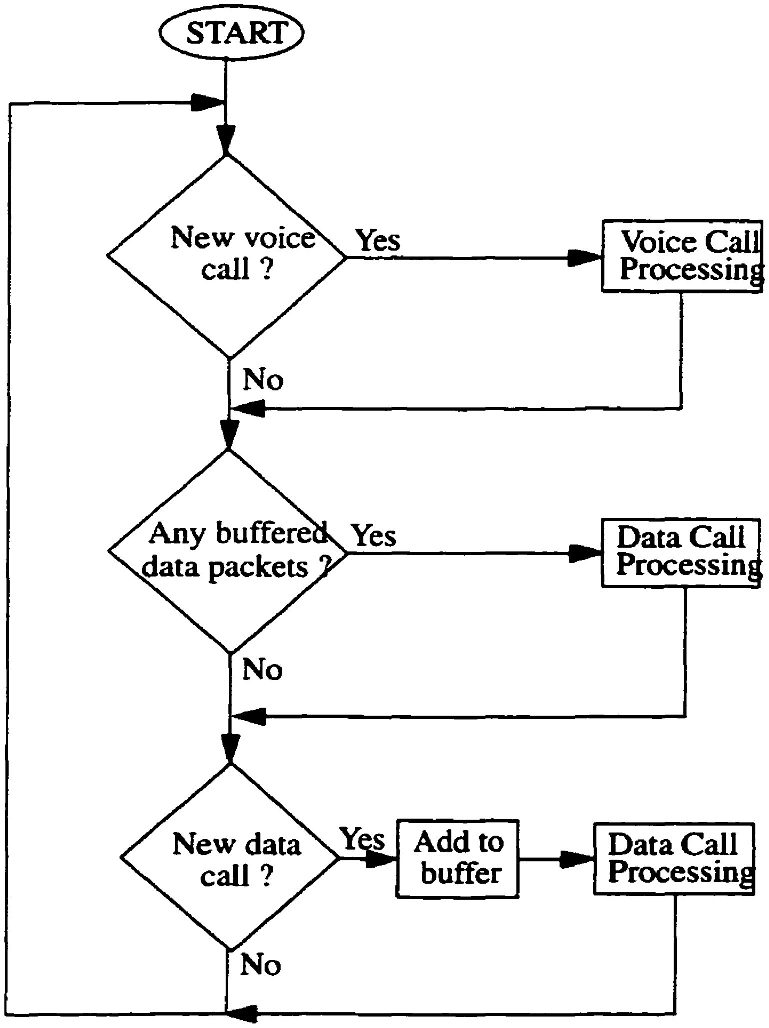

FIGURE 15. Priority flow chart

3.5.2 Voice Call Processing

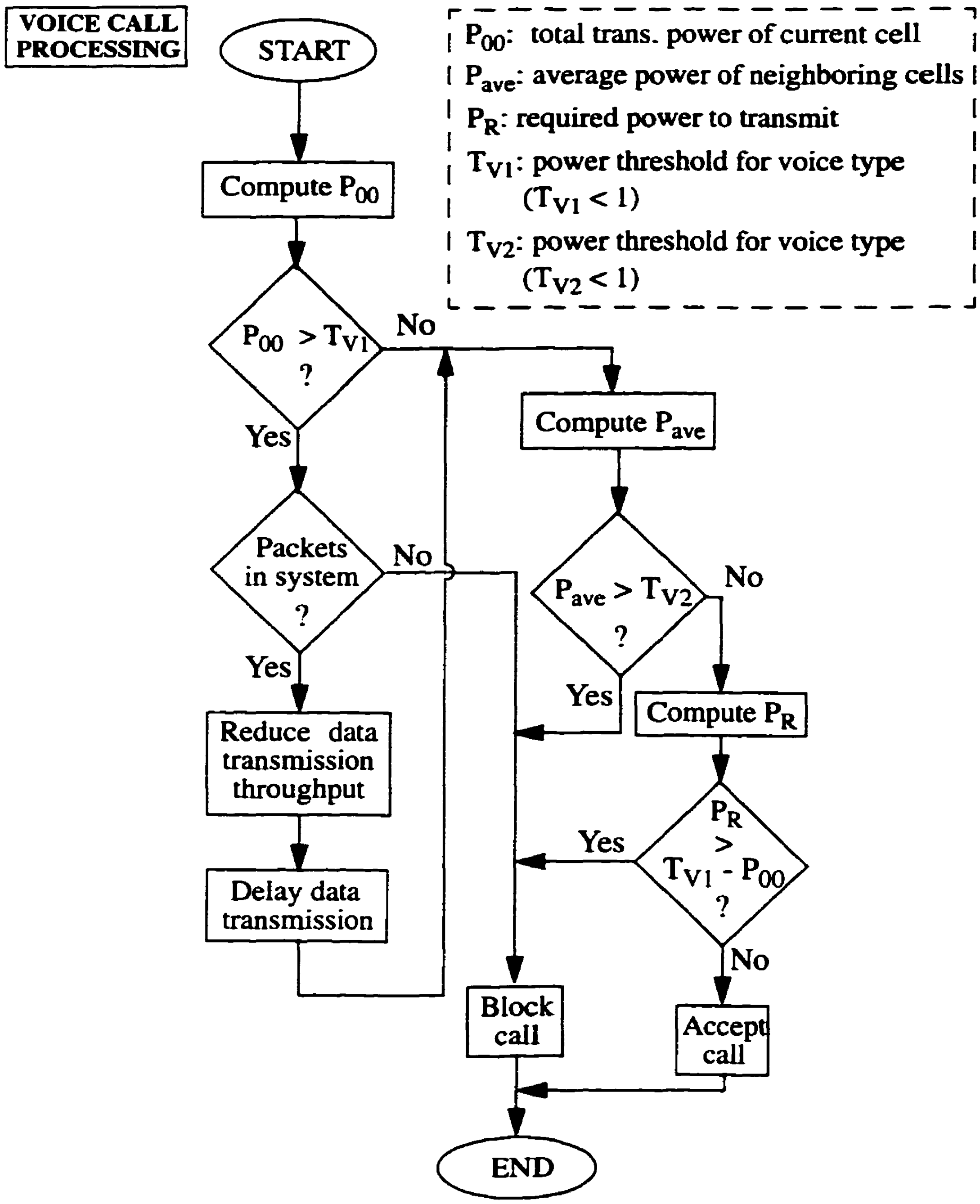

FIGURE 16. Voice call processing

Figure 16 presents the voice call processing algorithm as proposed in this voice/data integration.

Let the index of the cell which receives a new voice call be denoted by (2,2) and the maximum allowable power be M. In the following, let q_i be the operating points on the transmitted power distributions of the homogeneous traffics and T_{V1} and T_{V2} be the power thresholds given by,

$$T_{Vi} = M - q_i \tag{14}$$

and

$$q_1 > q_2 \tag{15}$$

The following four criteria are used for voice call admission:

1. If the total base power P_{00} is more than T_{V1} go to step 2, else go to step 3.
2. If the system contains data packets, reduce data transmission throughput to lower P_{00} below T_{V1}, delay their retransmissions, and go to step 3, else block the voice call.
3. Compute the average neighborhood Power P_{ave} based on

$$P_{ave} = \sum_{\substack{(i,j) \in A \\ (i,j) \neq (2,2)}} \alpha_{ij} P_{ij}, \tag{16}$$

   where A = {1,2,3} x {1,2,3} is the surrounding cells area, α_{ij} is a scaling factor based on the distance from the other base transmitters to the cell that receives the call, and P_{ij} is the transmitted power of the base in the cell (i,j).

   If $P_{ave} < T_{V2}$, go to 4, else block the voice call.
4. Compute the amount of power that must be transmitted to the mobile requesting the new call. If this power is less than M - P_{00} - q_i, then allow the voice call, else block the voice call.

Exploratory experiments indicated that a choice of 94% of full transmitted power for T_{V1} and 95% for T_{V2} would eliminate oscillations in the base transmitter power, yet yield a low blocking probability.

3.5.3 Data Call Processing

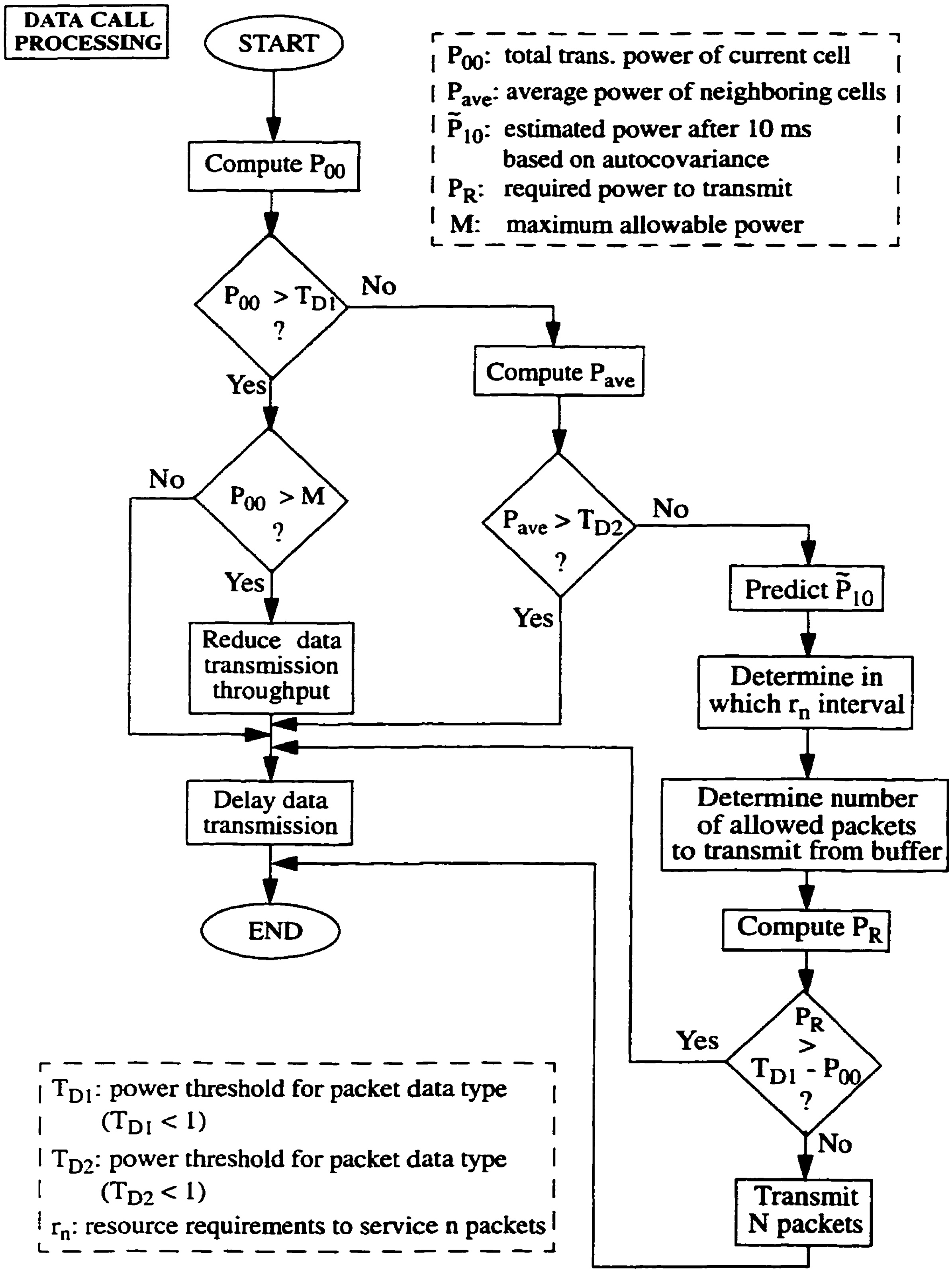

FIGURE 17. Adaptive data call processing

Figure 17 presents the data call processing algorithm as proposed in this voice/data integration.

Again, let the index of the cell which receives a new voice call be denoted by (2,2) and the maximum allowable power be M. In the following, let q_i be the operating points on the transmitted power distributions of the homogeneous traffics and T_{D1} and T_{D2} are power thresholds set using the following:

$$T_{Di} = M - q_i \tag{17}$$

and

$$q_1 > q_2 \tag{18}$$

The following four criteria are used for data flow control:

1. If the total base power P_{00} is more than T_{D1} go to step 2, else go to step 3.
2. If P_{00} is greater than the maximum allowable power and the network contains data packets, reduce data transmission throughput to lower P_{00} below T_{V1} and delay their retransmissions. Else just delay the retransmissions of the data packets.
3. Compute the average neighborhood Power P_{ave} based on

$$P_{ave} = \sum_{\substack{(i,j) \in A \\ (i,j) \neq (2,2)}} \alpha_{ij} P_{ij}. \tag{19}$$

   where A = {1,2,3} x {1,2,3} is the surrounding cells area, α_{ij} is a scaling factor based on the distance from the other base transmitters to the cell that receives the call, and P_{ij} is the transmitted power of the base in the cell (i,j).

   If $P_{ave} < T_{D2}$, go to 4, else delay the retransmissions of the data packets.

4. Using second order statistics, estimate the probability that the total transmitted power after 10 ms, $\bar{P}_{10}$, stays below the maximum allowable power level M. Then, depending on $\bar{P}_{10}$, the number N of allowed packets to transmit is determined, as explained in Section 3.5.4. Finally, compute the amount of power that must be transmitted to the

mobile requesting the new call. If this power is less than $M - P_{00} - q_i$, then transmit N packets, else delay the retransmissions of the data packets.

Exploratory experiments indicated that a choice of 0.90 for T_{D1} and 0.94 for T_{D2} would reduce oscillations in the base transmitter power, yet yield an acceptable data throughput.

The results are shown and analyzed in the following chapter.

3.5.4 Integration using Prediction

The proposed flow control is based on the prediction of the power requirements over the next packet transmission interval using a 2nd order statistic, the autocovariance. The fast Rayleigh fading and the slow shadow fading allow the use of autocovariance-based prediction. The total transmitted power of a base station is predicted over the next 10 msec, the time to transmit one packet. If the probability of exceeding the maximum allowable power M is sufficiently low, the extra power required to accommodate the data transmission is unlikely to result in outage and, therefore, data flow is permitted. The normalized autocovariance function, $R(\tau)$, of the transmitted power under constant traffic conditions allows such predictions and is given by

$$R(\tau) = \frac{1}{N-\tau} \sum_{i=0}^{N-\tau} P(i) \cdot P(i+\tau) , \tag{20}$$

where $P(i)$ is the total transmitted power function.

During the jth update interval, the maximum expected relative change δ_j in the total transmitted power after 10 ms can be estimated by comparing $R(j+10\ ms)$ with $R(j)$, i.e.,

$$\delta_j = \frac{R(j) - R(j+10\ \text{msec})}{R(j)} \quad , \quad j = 0, 1, 2, \ldots \tag{21}$$

Since the autocovariance is not stationary, $R(j)$ is continuously estimated and δ_j is updated every jth interval. This relative change δ_j is assumed to be independent of the mobile's distance from the base station.

If the power required to support a new packet data is P_D, then the total power required P_{00}' in the cell where the packet is transmitted is

$$P_{00}' = P_{00} + P_D. \tag{22}$$

where P_{00} is the current cell power.

Also, the total power required P_{ave}' in all neighboring cells as described in [Jalali94] is

$$P_{ave}' = P_{ave} + 0.05 \times \frac{P_{00}'}{\text{number of active data mobiles}}. \tag{23}$$

Hence, in the *jth* update interval, *n* packets are transmitted if

$$r_{n+1,j} < max\{P_{00}', P_{ave}'\} < r_{n,j}, \tag{24}$$

where $r_{n,j}$, the resource requirements to service *n* packets during the *jth* update interval, are illustrated in Figure 18.

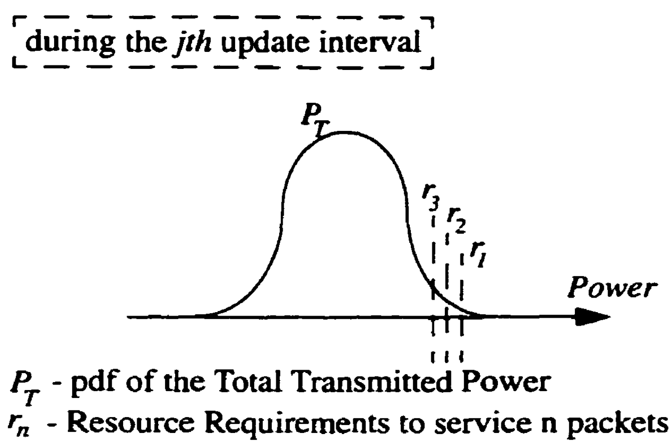

FIGURE 18. Flow control strategy based on prediction

The power required increases nonlinearly with the momentary transmission rate. However, for the range of *n* considered in the simulations, $r_{n,j}$ may be approximated by the following linear function:

$$r_{n,j} \le M - (n \times \delta_j), \tag{25}$$

where *M* is the maximum allowable power.

Depending on the probability that the transmitted power will exceed some power values, a certain number of data packets will be processed as shown in Figure 18; e.g.,

if the total transmitted power is between $r_{1,1}$ and $r_{2,1}$, 1 packet will be serviced,
if the total transmitted power is between $r_{2,3}$ and $r_{3,3}$, 2 packets will be serviced,
and so on...

4. Performance evaluation of packet data flow control

This chapter details the results obtained for the proposed algorithm in the downlink process under three simulation scenarios:

a) packet data calls only,

b) voice calls only, and a

c) mixture of voice and packet data calls.

A 'C' program is used to simulate this system. The assumptions and simulation parameters are first explained. The results of the simulations are then presented. Finally, the proposed adaptive flow control algorithm is compared to a non-adaptive algorithm.

4.1 Assumptions

Prior to discussing the simulation results, we first state the major assumptions of the proposed network model.

For the downlink computations, a cell layout of 5x5 cells is considered. All cells are considered equally populated by randomly positioned mobiles and the outage probability is estimated for the mobiles within the central cell. Starting with a low population, the number of mobiles is incremented until the outage probability is exceeded. At this point, the capacity is estimated as one less than the number of mobiles ($C = N_{outage} - 1$).

Mobile calls are assumed to request admission following a Poisson distribution with adjustable mean arrivals for voice and packet data. The mobiles are assumed to be moving at a speed of 2 Km/hr following a random walk in uniformly distributed angular directions. The standard deviation of the assumed log-normal fading is 8 dB and its spatial correlation function is assumed to be of exponential form $\exp(-d/d_0)$, where d is the Euclidean distance in space and d_0 is 6 meters.

In [IS-95], each cell can have up to six sectors, each of which holds 64 Walsh functions. Therefore, the maximum number of users in a cell is limited to 384. In our case, we assume that this limit will never be exceeded: either because data and voice users share the same Walsh functions, or the overall capacity limit of the system will be reached before the maximum number of users is reached.

4.1.1 Voice and Data Characteristics:

The characteristics of the voice and data traffic are described in Table 4. A maximum of one data packet can be sent to one terminal in any transmission interval.

TABLE 4. Stream and packet calls

	Services	Priority & delay	Length	Arrival	Activity
Stream	voice (video were considered in [Kandala96])	no delay but calls may be dropped	exponential duration with a mean of 4 min	Poisson with mean arrival rate of 45 msec	speech activity at 45% according to Markov model (mean of 0.25 sec for talk-spurt and 0.305 sec for silence)
Packet	data	can be delayed but retransmission has higher priority then newly arrived packets	constant length packets with exponentially distributed number of packets per call	Poisson with mean arrival rate of 75 msec	100% activity during a call

4.1.2 System & Traffic Characteristics:

- All cells are considered equally populated by randomly positioned mobiles.
- Transmitted power should not exceed 99% of its capacity.

- The call admission algorithm for the stream calls is the one described in [Kandala96].
- The flow control algorithm for packet data calls is the one proposed by this research and described in Section 3.5.
- Both voice and packet data calls are divided into frames of 10 msec.
- The arrival rates of the calls are set to ensure that the number of users in the system is always close to capacity. To ensure a sufficiently low rate of call blocking, they may need to be increased.
- The Erlang capacity is determined by adjusting the mean call arrival rate so that the blocking is just below 2%.

The simulations presented in this thesis are based on a simplified analytical model of the system characteristics. While, in reality, it is argued that arrival processes are not exactly Poisson distributed, it is believed that the results presented based on this assumption are a valid basis for system design.

4.2 Prediction Parameters

As explained in Section 3.5.4, the autocovariance, $R(\tau)$, is evaluated on the total transmitted power of the cell containing voice calls only. Figure 19 shows a typical total transmitted power profile at full speech capacity with its probability density function (pdf). Using this power function, $R(\tau)$ is computed and is displayed in Figure 20. The fast change in the autocovariance function is due to fading. This results in a low correlation of the power. However, the slowly varying, strongly correlated portion is due to speech activity. A close-up in Figure 21 shows in detail the expected behavior of the power within 10 msec, the time to transmit one packet.

TABLE 5. Autocovariance values

time τ (in msec)	Autocovariance $R(\tau)$
0	10.76
10	10.563

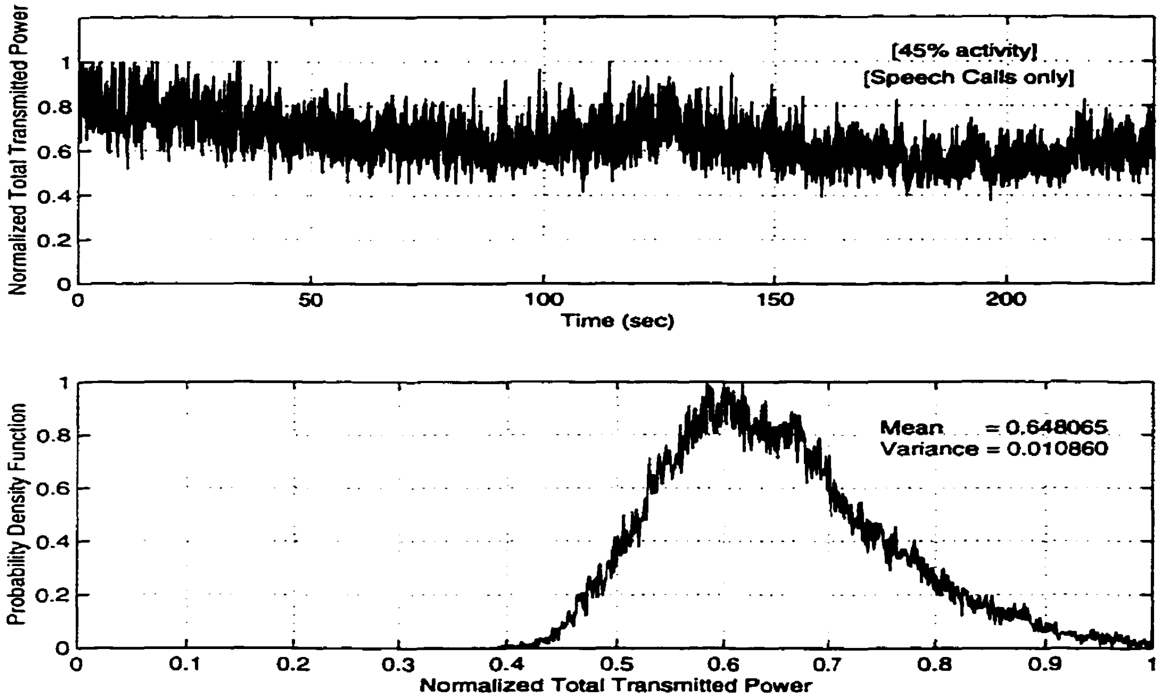

FIGURE 19. Typical total transmitted power with 45% activity and PDF

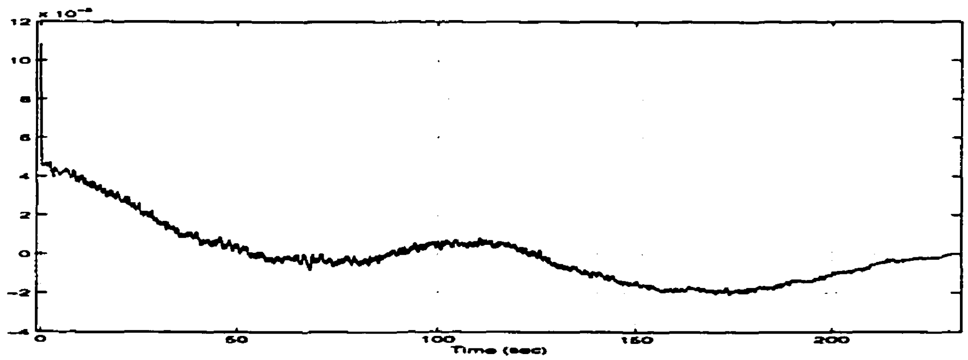

FIGURE 20. Autocovariance of the total transmitted power of speech only

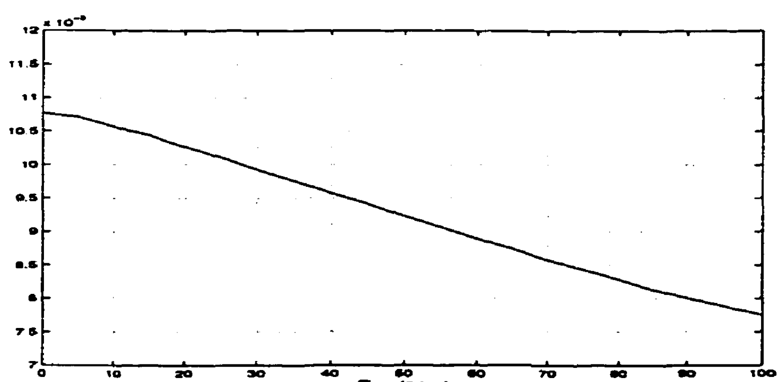

FIGURE 21. Close up of autocorrelation

The average maximum expected relative change for this case is found to be

$$\delta = \frac{10.76 - 10.563}{10.76} = 0.0183 \text{ or } 1.83\%. \tag{26}$$

Simulations with half of full power speech calls only were run and an average maximum expected relative change, δ, was found to be 1.2%. For networks where data packets are not serviced using an ABR policy, other average values of δ should be considered. Using this value of δ, the resource requirements, r_n, that allow the prediction of the total transmitted power to service n packets, can be found using

$$r_n = M - (n \times \delta), \tag{27}$$

with M being the maximum allowable power.For M = 92% of the power and δ = 1.83%, some r_n are computed using (27) and are shown in Table 6.

TABLE 6. Resource requirements for M=0.92 and δ=0.0183

$r_1 = 0.902$	$r_6 = 0.810$	$r_{11} = 0.719$	$r_{16} = 0.627$
$r_2 = 0.883$	$r_7 = 0.792$	$r_{12} = 0.700$	$r_{17} = 0.609$
$r_3 = 0.865$	$r_8 = 0.774$	$r_{13} = 0.682$	$r_{18} = 0.591$
$r_4 = 0.847$	$r_9 = 0.756$	$r_{14} = 0.664$	$r_{19} = 0.572$
$r_5 = 0.829$	$r_{10} = 0.737$	$r_{15} = 0.646$	$r_{20} = 0.554$

δ is updated every 0.5 seconds to optimize the flow control of data. A more frequent update is possible resulting in a computational increase. However, the estimation interval is the previous 8 seconds of total transmitted power values to ensure enough data for the autocovariance evaluation.

4.3 Capacity Assessment

To obtain comparative results, we model the channel by a single Rayleigh fading path for 5 MHz systems. Artificial multipath is used to provide diversity. In environments where sufficient natural multipath is observed, the amount of artificial multipath provided may be reduced.

The capacity estimation procedure for uniform traffic is described in [Jalali94]. First, the required SNR at the receiver is computed so that an acceptable quality of service can be provided. Mobiles are then incrementally introduced until the outage condition is exceeded. Average capacity is one less than the value of the number of mobiles at the onset of outage. The same procedure can be extended to mixed traffic. We further assume that the cells are square shaped with each side equal to 100 meters and that 12.8% of the total power available (i.e., the maximum power that can be transmitted from the base) is allocated to the pilot signal. This figure is computed so that the required pilot-SNR is achieved at least 95% of the time as discussed in [Jalali94]. We consider three omni-directional antennas to provide a space-diversity of 3, and we further assume that the distribution of the relative powers of the three paths is uniform. Packet data calls require BER < 10^{-6} and voice calls require BER < 10^{-3}. Table 7 summarizes all the channel parameters that are used in all simulations shown in this chapter. A propagation model with one path was chosen, i.e., no multipath was considered. The presence of multipath will result in a diversity gain [Jalali94].

TABLE 7. Channel conditions

Item	Value
Path loss exponent	4
Number of transmit antennas	3
Random user path power distribution	1
User path power distribution	1
Number of channel paths received at mobile's antenna	1
Number of path combinations	1
Path power distribution	1
Pilot signal fraction	0.128
Standard deviation of the fading	8.0 dB

4.3.1 Packet Data Calls Only

In this first scenario, only packet data calls are considered, i.e., all of the mobiles connecting with the base station are data mobiles. Table 8 displays some parameters used for this study. The average number of packets per call is chosen to be high to simulate a system under full capacity. In Figure 22, we see that the total transmitted power is quite stable despite the random introduction of new calls and the termination of old ones. This is the result of a robust flow control algorithm with good prediction. The cell (2,2) is the cell under study, while all other neighboring cells are considered to simulate the interference.

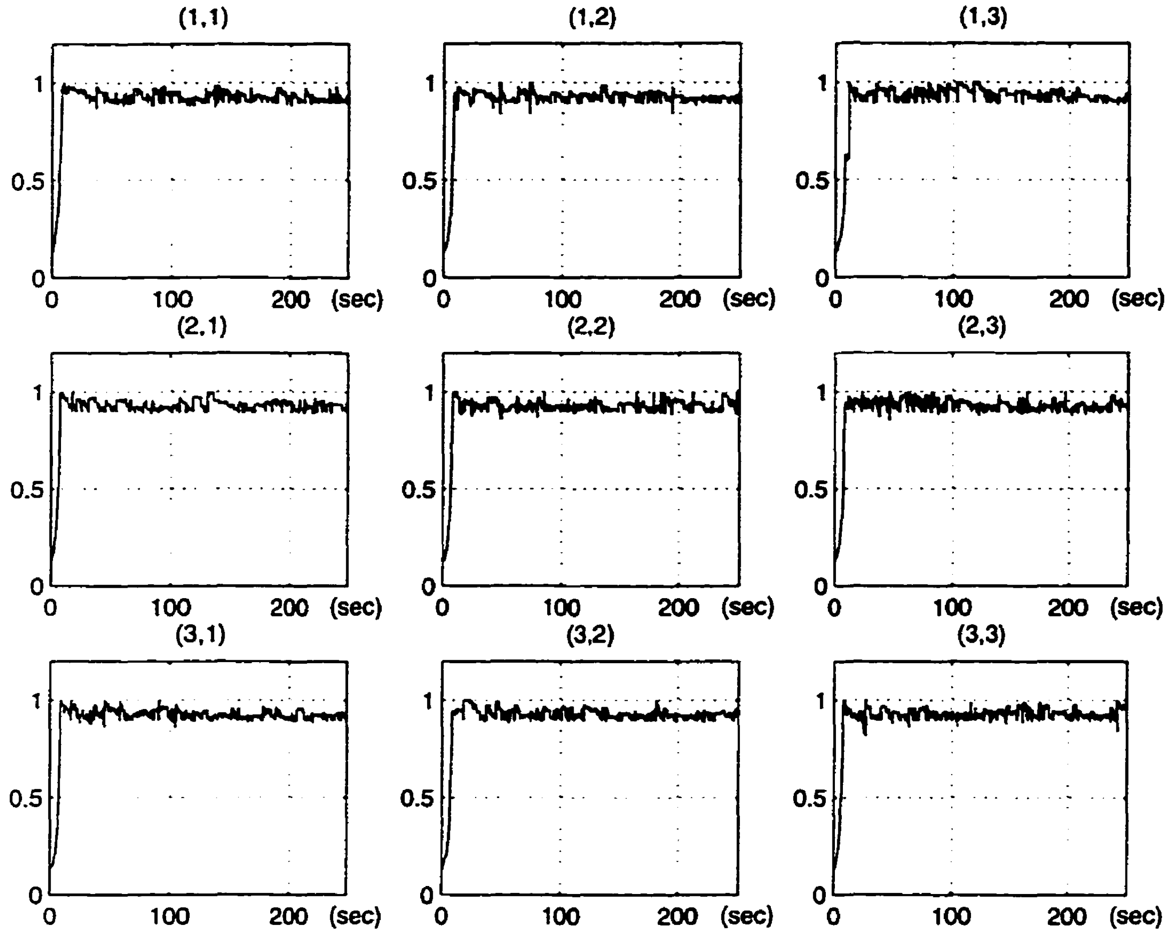

FIGURE 22. Total transmitted power with uniform packet data traffic

TABLE 8. Simulation parameters for packet data calls only

Item	Value
Bandwidth	5 MHz
Total simulation time	250 sec
Cell structure	5x5 cells
Cell side	100 meters
Speed of the mobiles	2 Km/hr
Total number of data mobiles per cell	180 mobiles
Inter-arrival rate of new packet data calls	75 msec
Average number of packets/call	15000
Retransmission delay	10 msec
Transmission packet time	10 msec

TABLE 8. Simulation parameters for packet data calls only

Item	Value
Processing gain for packet data calls	512
Threshold T_{D1} (refer to Section 3.5.3)	90 %
Threshold T_{D2} (refer to Section 3.5.3)	94 %
$(SNR)_{low}$	7.0 dB
$(SNR)_{high}$	8.0 dB

Figure 23 displays the throughput of this system versus the delay per packet. We note that 99% of the time, the delay is less than 160 msec. This delay includes the first transmission of the packet, the retransmissions, and the negative acknowledgment messages.

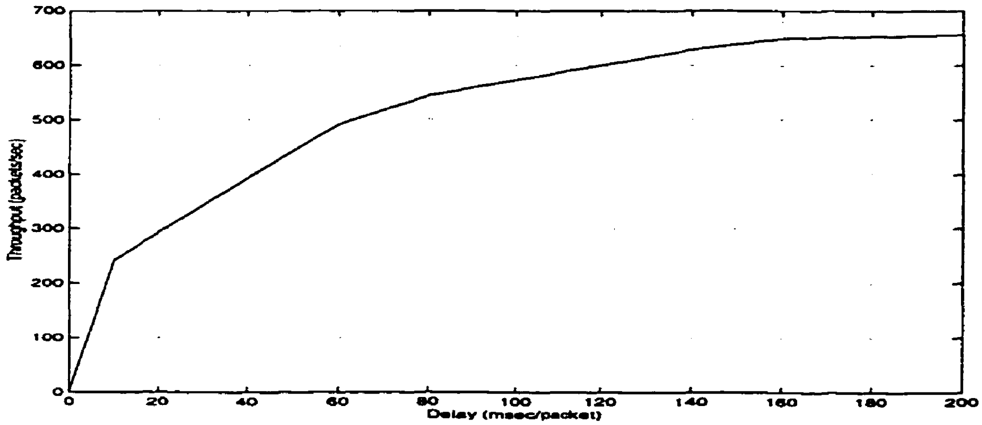

FIGURE 23. Throughput vs. delay with uniform packet data traffic

At a 90% loading, the average delay is about 120 msec. A comparison of throughputs using adaptive and fixed algorithms is discussed in Section 4.4.

In Figure 24, the number of active data mobiles monitored every 5 msec is shown. The first 20 seconds corresponds to the initialization of the system with packet data calls. The average number of active data mobiles at a given time in each cell is about 71 mobiles, i.e., about 71 packet data calls can be active at the same time. The small deviations from these numbers in the plot of Figure 24 are due to the fact that some packets are being transmitted, others are retransmitted, other packet data calls are being completed, new packet data calls are entering the system, and so on. The distribution of transmitted data packets (first transmissions and retransmissions) at one time is shown in Figure 25.

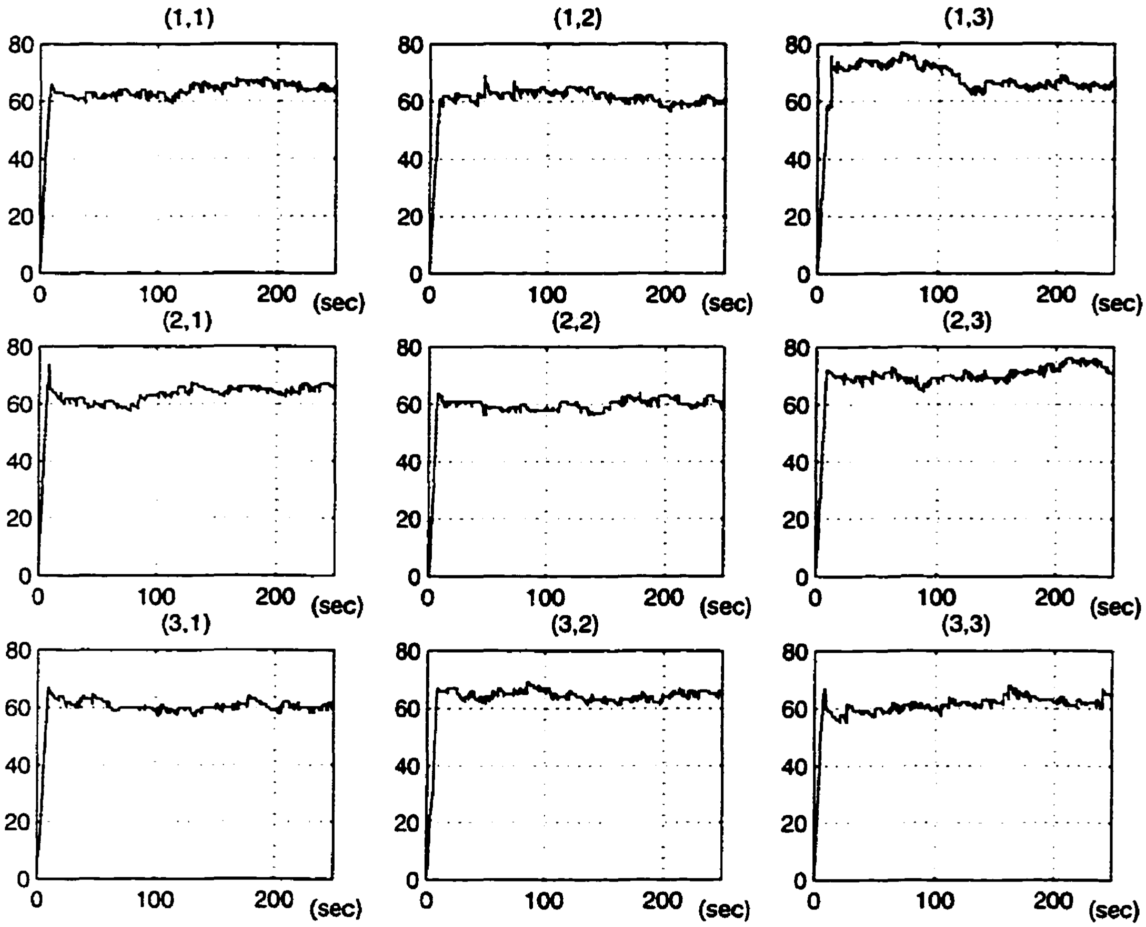

FIGURE 24. Number of active data mobiles with uniform data traffic

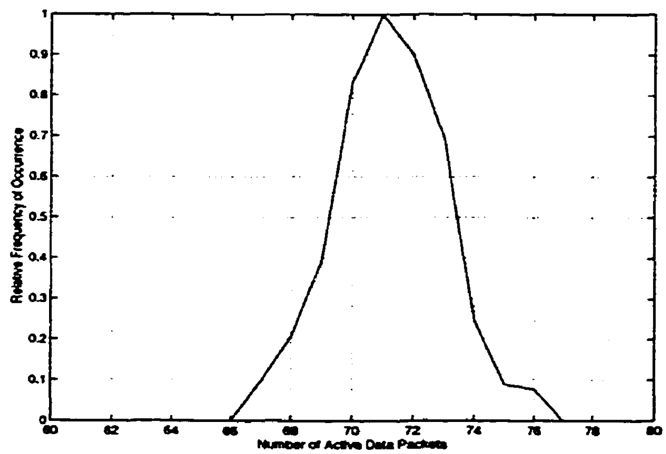

FIGURE 25. Distribution of transmitted data packets with uniform data traffic

In Figure 26, the number of buffered data calls is shown to be quite stable as packet data calls are monitored through the proposed flow control algorithm described in

Section 3.5.3. This plot includes active and non-active calls. Non-active packet calls are calls that have been admitted to the queue, but have not yet started transmission. The buffer is filled in the first 20 seconds by the initialization process and then new packet data calls enter the system following a Poisson distribution with an average rate of 75 msec. Note that a mobile can only receive one packet at a time, i.e., it can only be connected with one call at a time, and the packets are sent serially on a first-come first-served basis. However, retransmissions have a higher priority over buffered packets.

Now, given that a data packet is erroneously received the first time, the probability of multiple retransmissions for that packet is shown in Figure 27.

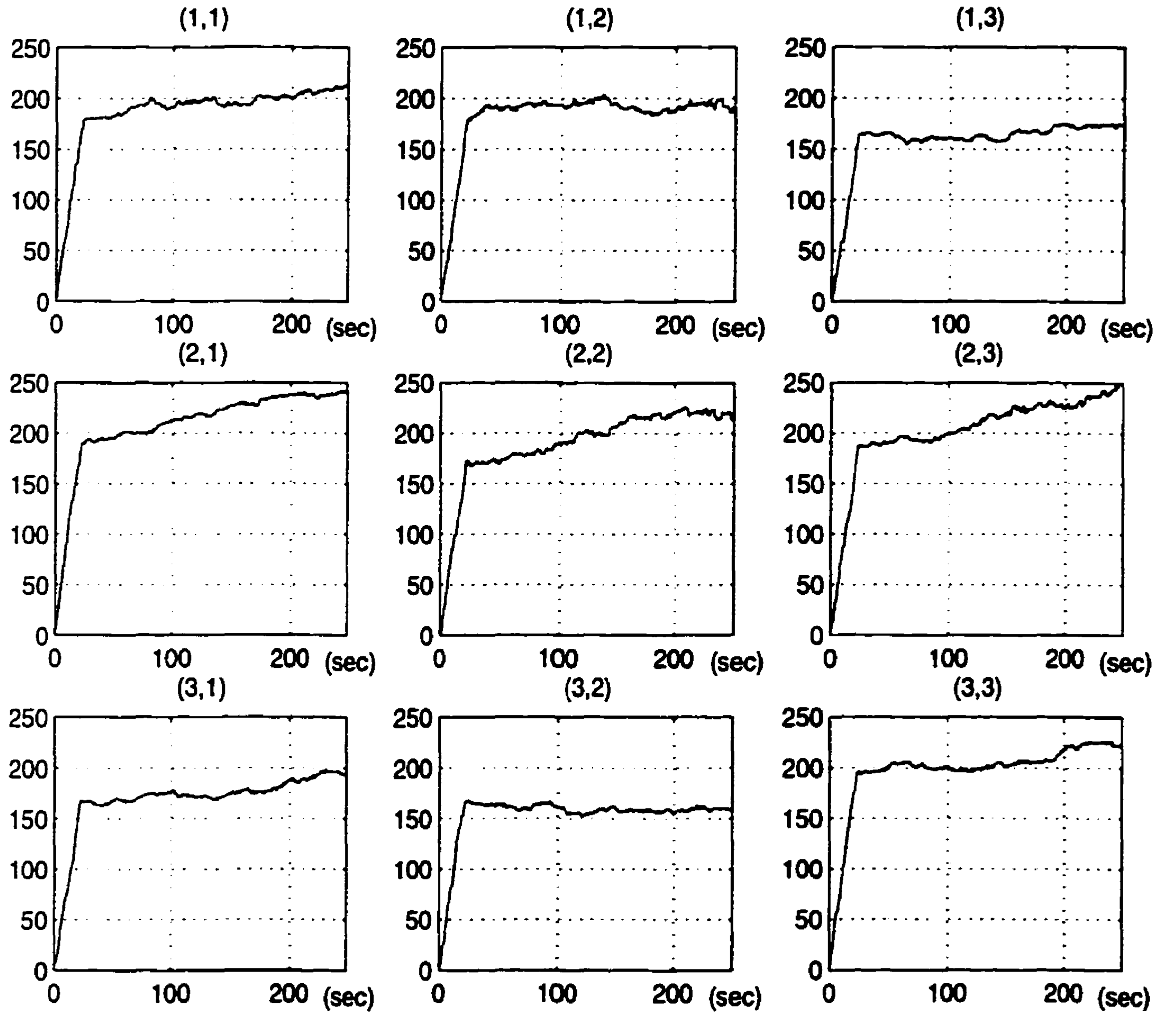

FIGURE 26. Number of buffered data calls with uniform data traffic

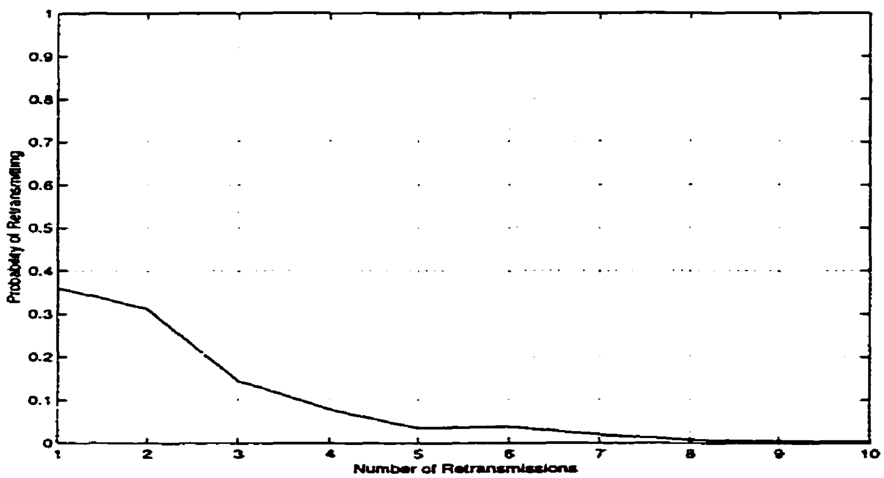

FIGURE 27. Probability of retransmissions after 1st negative acknowledgment

In Figure 28 and Table 9, the retransmissions of data packets are shown. Retransmissions are requested when either the mobile or the base station is in outage, i.e., when the mobile has a SNR lower than $(SNR)_{low}$, or when the total transmitted power at the base station exceeds the maximum allowable power. Note that for this simulation, it was found that about 67 mobiles are active at the same time, i.e., about 4% of the active mobiles request retransmissions 6% of the time, 12% of the active mobiles request retransmissions 0.7% of the time, and so on.

These figures will, however, change as speech calls are introduced due to the fact that packet data calls will be services on an ABR basis in that case. Therefore, more retransmissions will be required to ensure a BER less than 10^{-6} for the data packet services. However, the loss of voice frames is not accounted for here, since it does not have an influence on the overall voice traffic and thus will not affect the flow control of data.

TABLE 9. Frequencies of retransmissions with packet data calls only

# of retransmissions	% of active mobiles	% of occurrence
0	0	93.2880
1-4	~ 4	5.8940
5-9	~ 12	0.7400
10-19	~ 23	0.0710
20-39	~ 45	0.0056
40-59	~ 75	0.0014

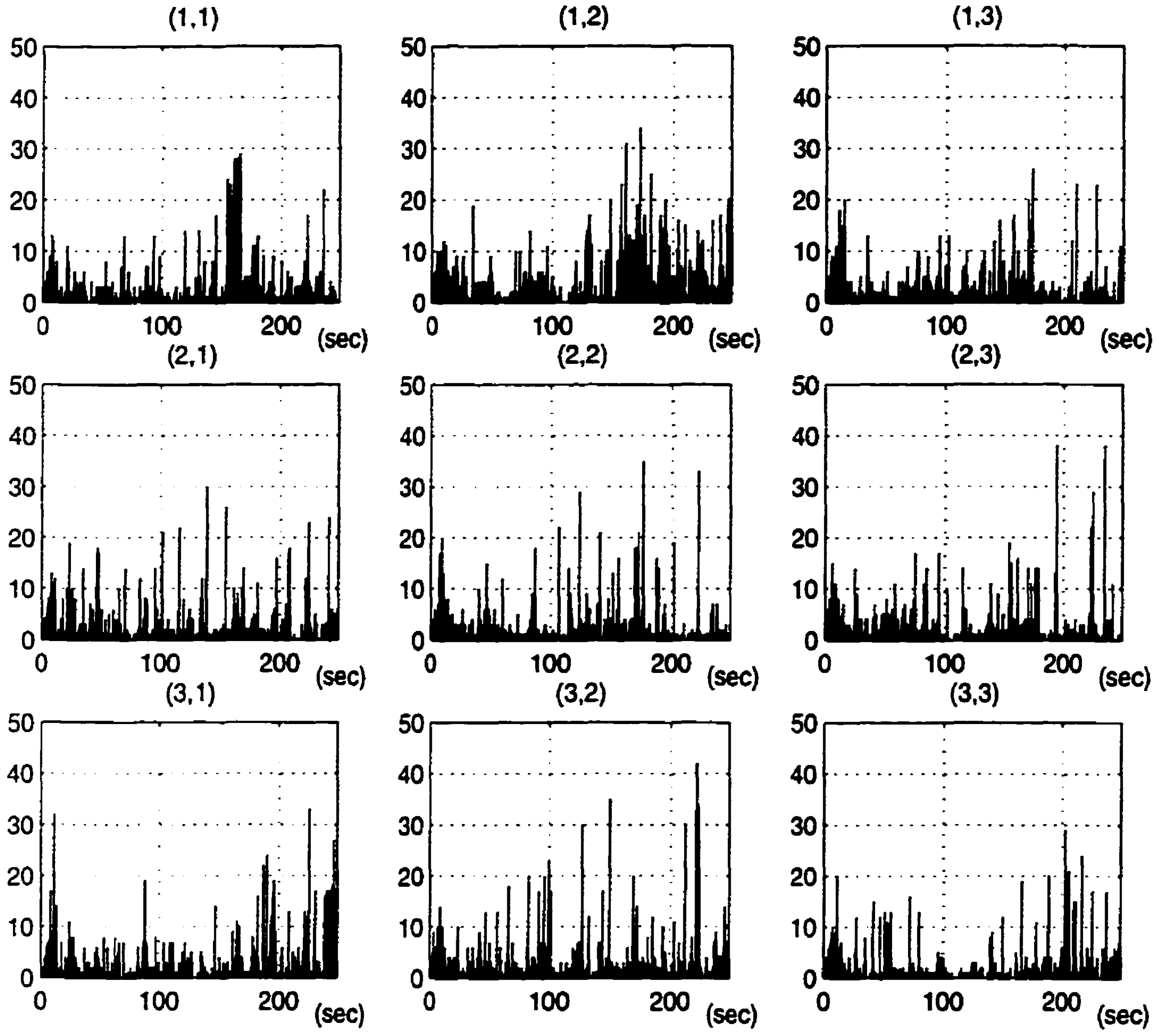

FIGURE 28. Number of packet data retransmissions with uniform data traffic

4.3.2 Speech Calls Only

In this scenario, only voice calls are considered, i.e., all of the mobiles connecting with the base station are speech mobiles. Table 10 displays some parameters used for this study. This case has been studied by S. Kandala and reported in [Mermelstein95]; however, for the sake of completeness of this study, some results are shown here. In Figure 29, the total transmitted power is shown under full capacity. Note that this is the scenario used when integrating packet data calls as an ABR system is implemented. The call admission protocol used here is the one described in [Kandala96]. The cell (2,2) is the cell under study, while all other neighboring cells are considered to simulate the interference.

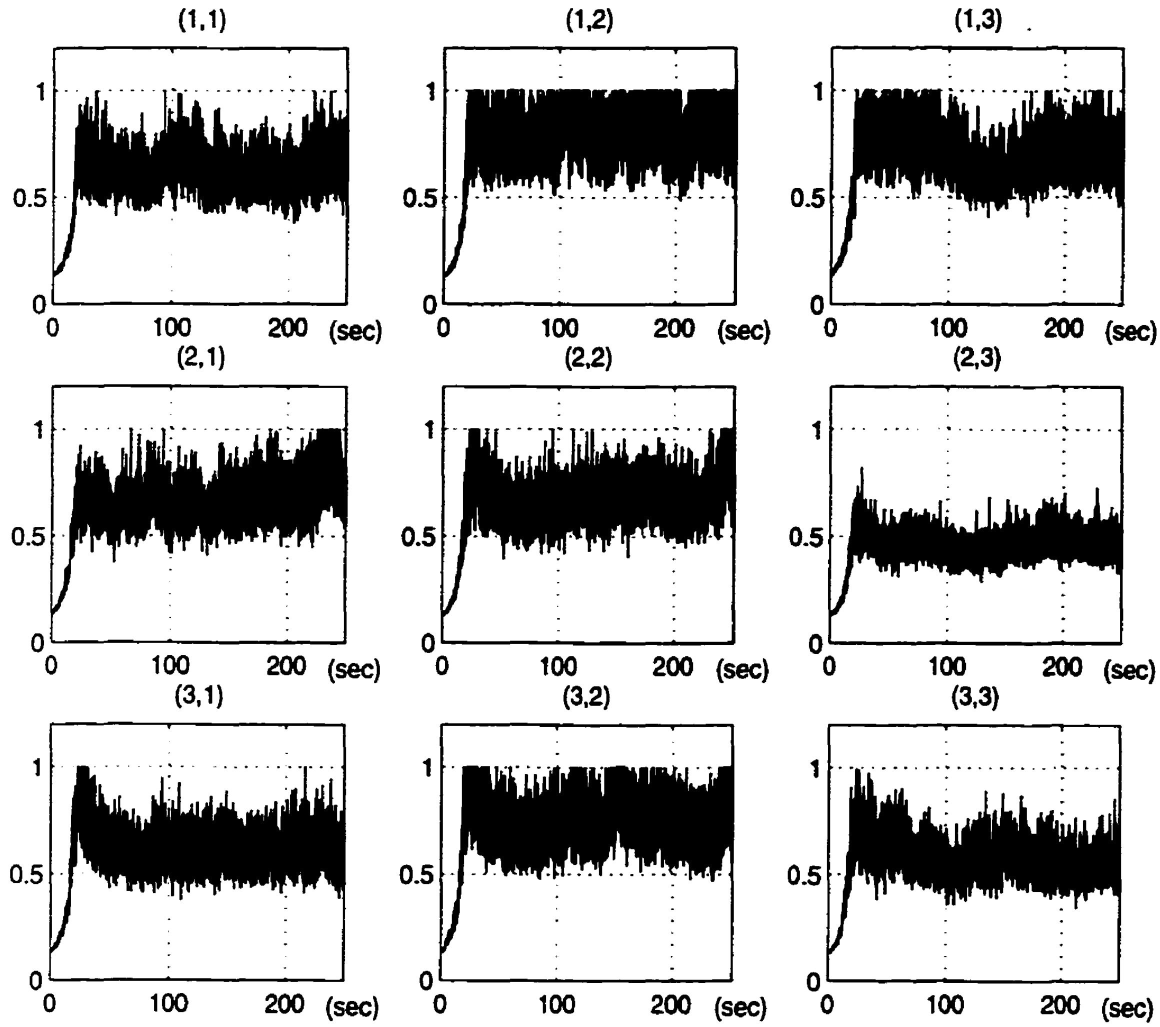

FIGURE 29. Total transmitted power with uniform speech traffic

TABLE 10. Simulation parameters for voice calls only

Item	Value
Bandwidth	5 MHz
Total simulation time	250 sec
Cell structure	5x5 cells
Cell side	100 meters
Speed of the mobiles	2 Km/hr
Total number of speech mobiles per cell	180 mobiles
Inter-arrival rate of new voice calls	45 msec
Average duration of voice calls	4 minutes
Average duration of talkspurt periods	0.25 sec
Average duration of silent periods	0.305 sec
Processing gain for voice calls	512

TABLE 10. Simulation parameters for voice calls only

Item	Value
Threshold T_{V1} (refer to Section 3.5.2)	94 %
Threshold T_{V2} (refer to Section 3.5.2)	95 %
$(SNR)_{low}$	6.5 dB
$(SNR)_{high}$	7.5 dB

In Figure 30, the number of active speech mobiles monitored every 5 msec is shown. The first 20 seconds corresponds to the initialization of the system with voice calls. The average number of active speech mobiles at one time in every cell is about 150 mobiles which is more than double the number of active data mobiles with uniform packet data traffic previously reported. This is due to the fact that, in speech calls, there is only about 45% speech activity, hence the average power usage per call is lower. As in the case of data, the small fluctuations of the number of active calls in these plots are due to the fact that some voice calls are being terminated, others are being entered into the system, and others are just blocked. The throughput of this system is shown in Table 11 and plotted in Figure 35.

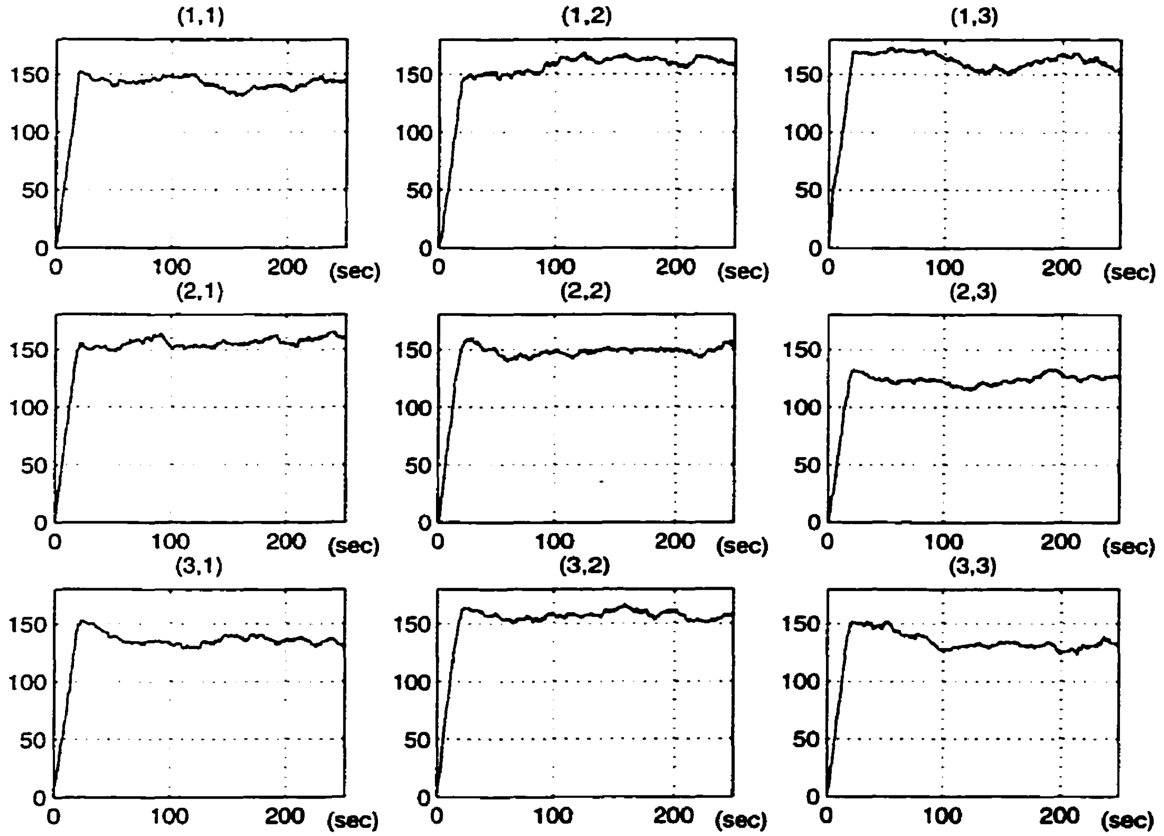

FIGURE 30. Number of active speech mobiles with uniform speech traffic

TABLE 11. Throughput for speech calls only

Number of voice calls	Throughput (Kb/sec)
165	685
155	643
140	581
130	540
70	291

4.3.3 Mixed Services

In this scenario, mixed traffic is considered, i.e., the mobiles connecting with the base station are either speech or data mobiles. Table 12 displays some parameters used for this study. The total number of speech mobiles is much higher than the total number of packet data mobiles to simulate an ABR system under full capacity. Figure 31 shows the total transmitted power when both speech and packet data calls are considered. By comparing Figure 31 with Figure 29, we note that the capacity enhancement is better exploited as the power is more fully used. The cell (2,2) is the cell under study, while all other neighboring cells are considered to simulate the interference.

TABLE 12. Simulation parameters used for mixed traffic

Item	Value
Bandwidth	5 MHz
Total simulation time	250 sec
Cell structure	5x5 cells
Cell side	100 meters
Speed of the mobiles	2 Km/hr
Total number of speech mobiles per cell	180 mobiles
Inter-arrival rate of new voice calls	45 msec
Average duration of voice calls	4 minutes
Average duration of talkspurt periods	0.25 sec
Average duration of silent periods	0.305 sec
Total number of data mobiles per cell	20 mobiles
Inter-arrival rate of new packet data calls	0.4 sec
Average number of packets/call	1500
Retransmission delay	10 msec
Transmission packet time	10 msec
Processing gain for voice and packet data calls	512

TABLE 12. Simulation parameters used for mixed traffic

Item	Value
Threshold T_{V1} (refer to Section 3.5.2)	94 %
Threshold T_{V2} (refer to Section 3.5.2)	95 %
Threshold T_{D1} (refer to Section 3.5.3)	90 %
Threshold T_{D2} (refer to Section 3.5.3)	94 %
$(SNR)_{low}$ for voice calls	6.5 dB
$(SNR)_{high}$ for voice calls	7.5 dB
$(SNR)_{low}$ for packet data calls	7.0 dB
$(SNR)_{high}$ for packet data calls	8.0 dB

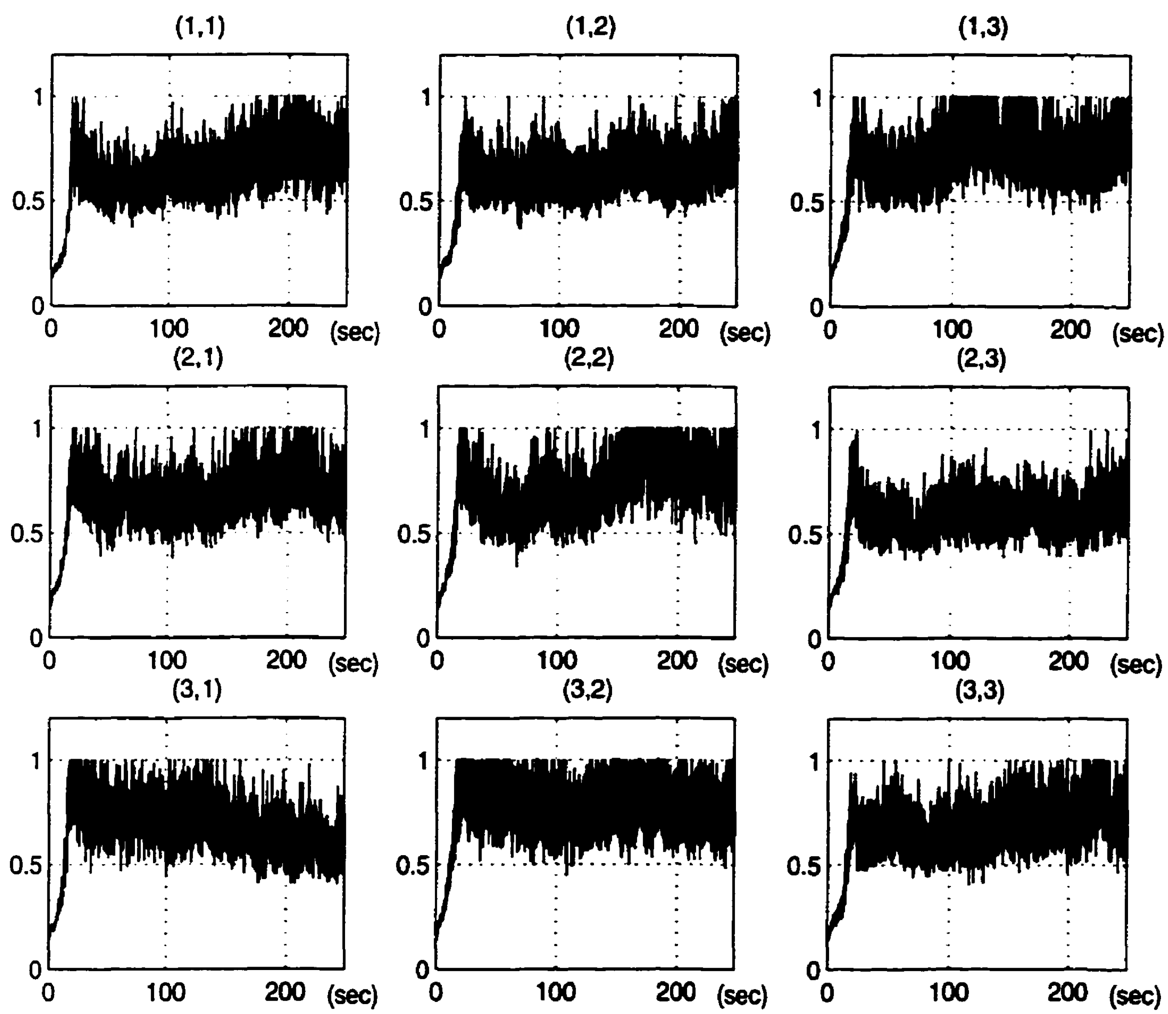

FIGURE 31. Total transmitted power with mixed traffic

Note that the speech calls parameters are kept the same as described in Section 4.3.2, while the packet data conditions are reduced to simulate an ABR system.

In Figure 32, the number of active speech mobiles is shown. We note that this figure is very similar to the one presented in the case of uniform speech traffic (Figure 30). The average number of active speech mobiles is about 150 mobiles. The fact that this number is kept constant emphasizes the need for a good flow control that will service data packets even when the system is under full capacity.

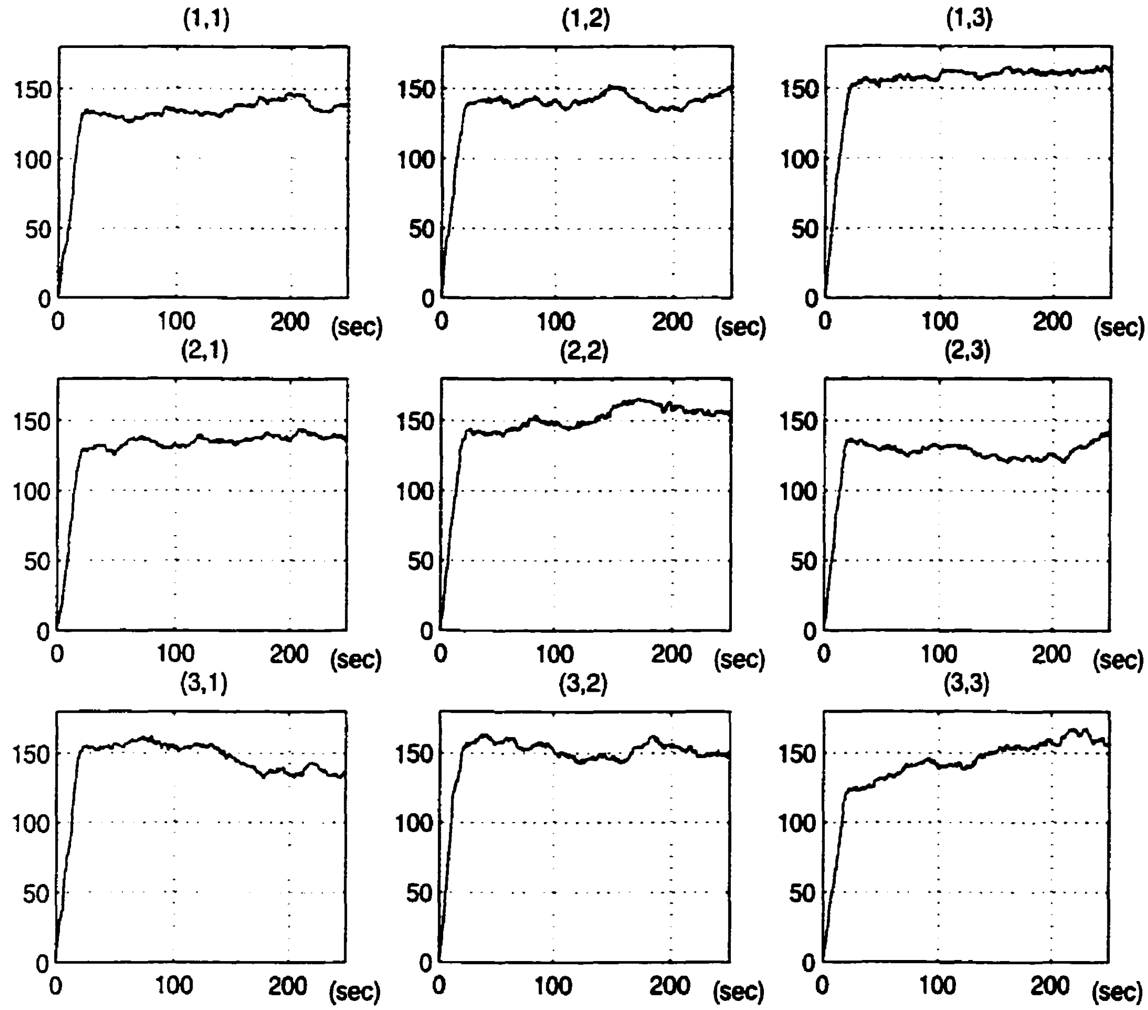

FIGURE 32. Number of active speech mobiles with mixed traffic

The number of active data mobiles with mixed traffic is shown in Figure 33. Note that the first 20 msec are due to initialization of the system. We can see that an average of 2.5 packet data calls are served at one time when the system is full with speech calls. However, a maximum of 7 additional packets were serviced when the inter-arrival rate of new voice calls was increased to 60 msec. The distribution of the transmitted data packets is shown in Figure 34.

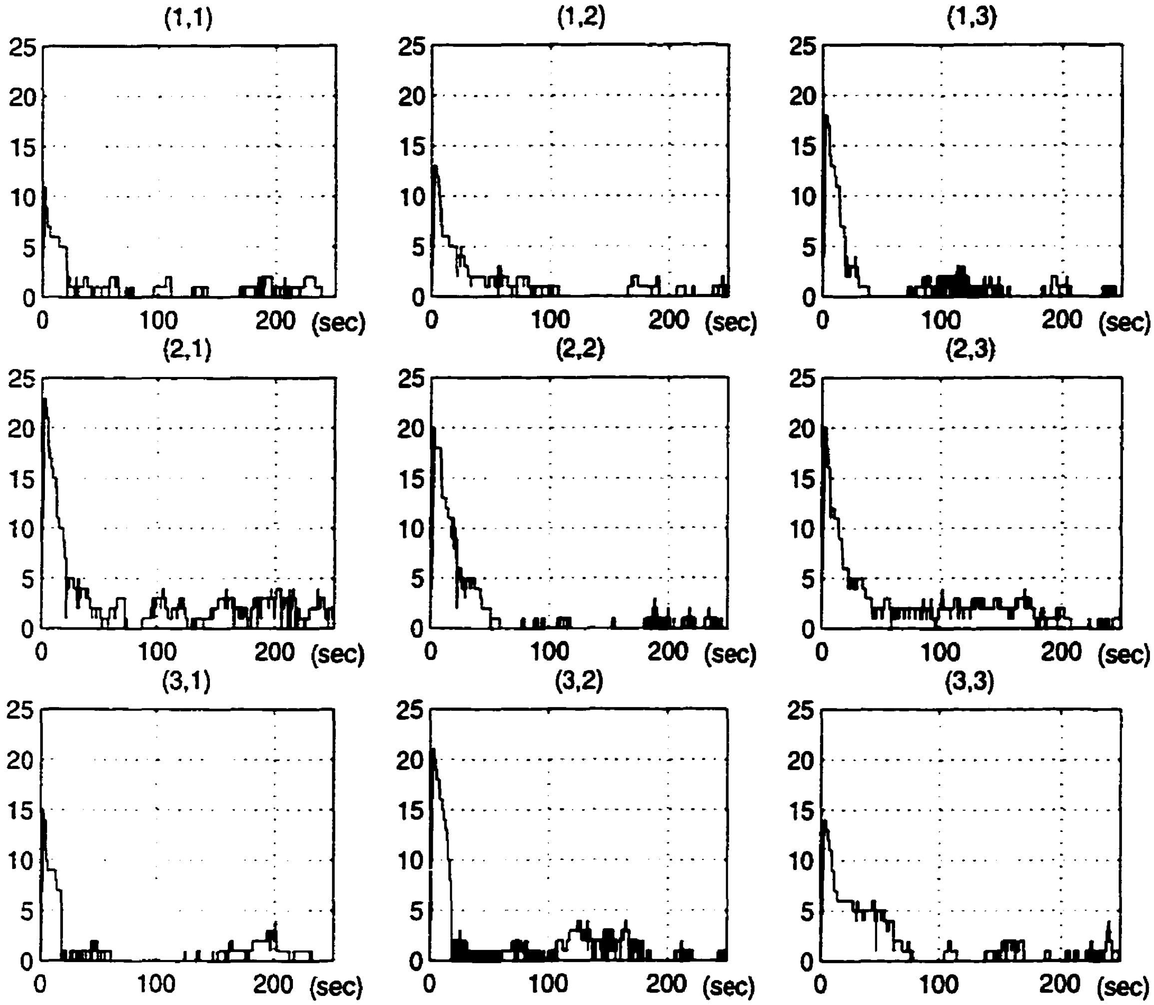

FIGURE 33. Number of active data mobiles with mixed traffic

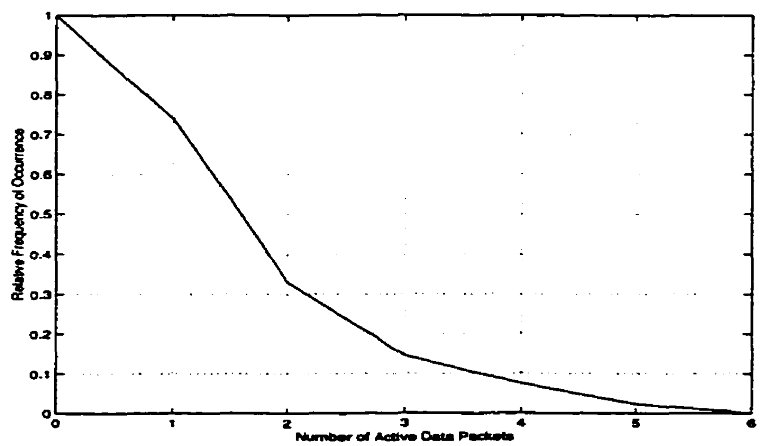

FIGURE 34. Distribution of transmitted data packets with mixed traffic

The results shown in this simulation correspond to systems reserved 100% for voice calls. However, if we consider systems less loaded by speech calls, we get an intermediate combination of packet data calls and voice calls as shown in Figure 35.

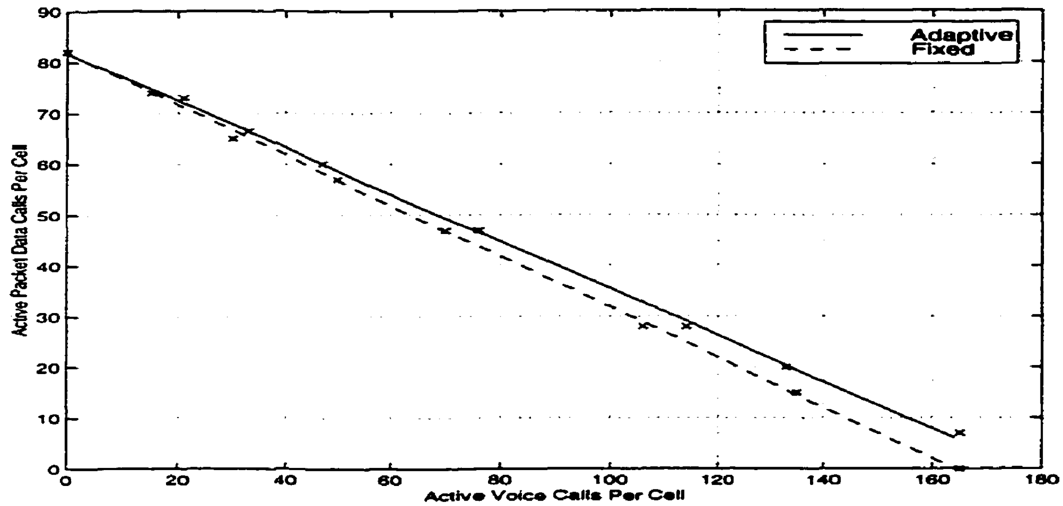

FIGURE 35. Call capacity in a mixture traffic

The proposed algorithm produces the solid line on the graph. The dotted curve results from the non-adaptive (or fixed) algorithm described in Section 4.4. Dynamic simulations under full loading reveal a noticeable variation in the number of packet data calls supported over the range 30-40 and relatively smaller variations in the simultaneously supported voice calls in the range 75-95. Table 13 shows the total throughput for the adaptive mixed traffic case.

TABLE 13. Throughput for the adaptive mixed traffic

Number of voice calls	Number of packet data calls	Total throughput (Kb/sec)
165	7	741
133	20	712
114	28	697
76	47	691
47	60	675
33	67	672
21	72	663
0	82	656

These results indicate that admitting one voice call instead of a data call will increase overall throughput. Evidently, the silence periods in voice calls allows more ABR data packets to be transmitted. Correspondingly, the throughput is minimum when the cell is completely populated with data mobiles.

4.4 Adaptive vs. Non-Adaptive Flow Control Algorithm

The adaptive flow control algorithm is the one proposed in this research and implements packet level admission control as well as call level control. The packet level control exploits the variations in transmitted power and receiver signal-to-interference ratio over intervals comparable to the packet duration. On the other hand, a non-adaptive flow control algorithm has no packet level flow control, only call level admission control as shown in Figure 38.

The advantage of using adaptive flow control is a better throughput control and, therefore, a better utilization of the available bandwidth, which is a prime objective as shown in Figure 37. In the proposed adaptive flow control, *N*, the number of allowed packets to transmit, depends on the different resource requirement ranges and the level of the predicted total transmitted power over the next 10 msec. On the other hand, in the non-adaptive algorithm, *N* is the difference between the maximum allowed packet data calls and the current number of active packet data calls.

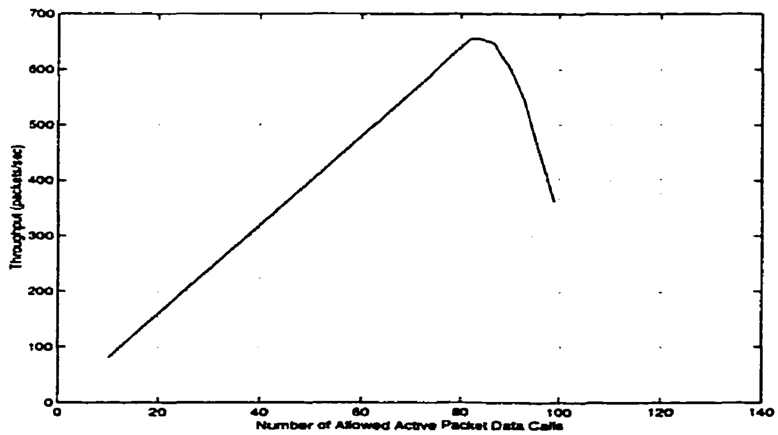

FIGURE 36. Throughput of a non-adaptive flow control algorithm

Simulations were done to determine the difference in packet data throughputs due to these different algorithms. Figure 36 shows the disadvantage of a non-adaptive flow control protocol: the throughput is decreased when the system is overloaded.

Due to the sharing of the spectrum and the lack of a constraint on the number of active data mobiles, adaptive flow control, which can cope efficiently with nonuniformity and nonstationarity of traffic patterns and propagation characteristics, is a key element of an Integrated Wireless Access Network bandwidth management system as shown in Figure 37. The average number of packets per call is 5000. Even if the system is overloaded, adaptive flow control assures a steady and high throughput.

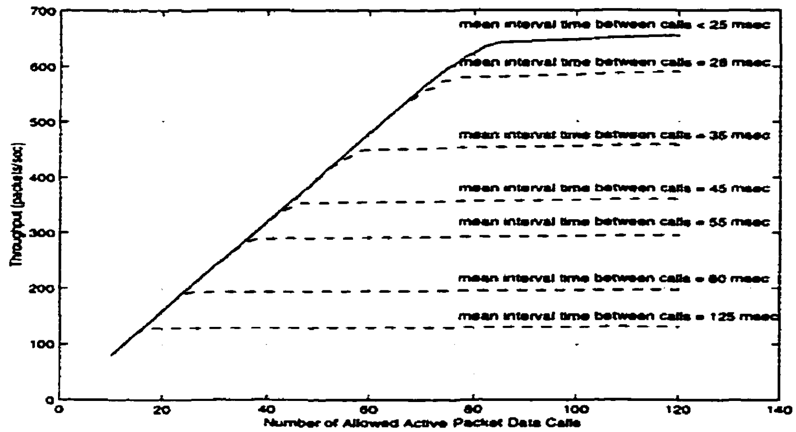

FIGURE 37. Throughput of an adaptive flow control algorithm

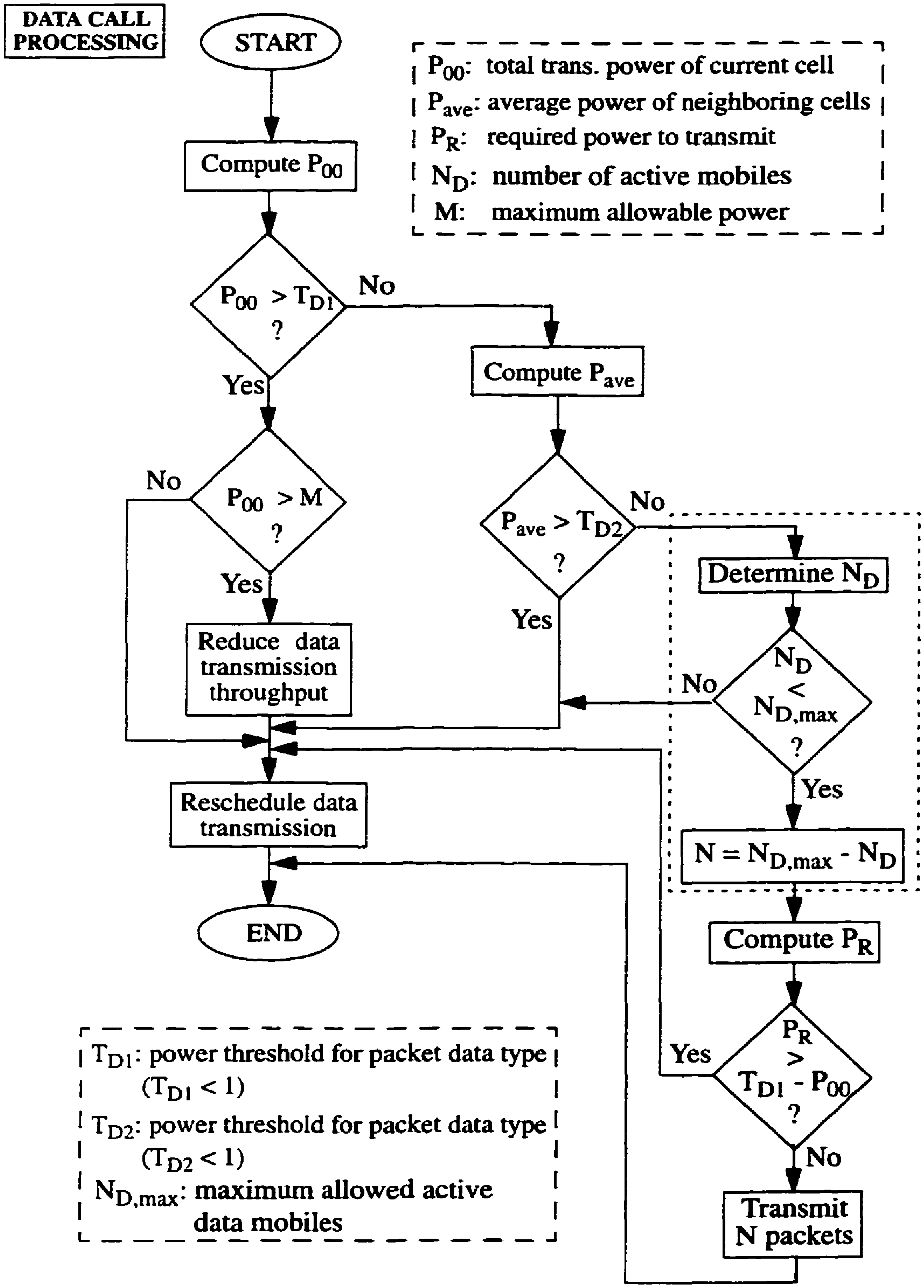

FIGURE 38. Non-adaptive data call processing

5. Conclusions

Bandwidth management generally includes algorithms for call admission, flow control, and congestion control. This work has focused on flow control for integrated voice/data wireless CDMA networks based on an understanding of this relationship. In fact, all users in a cell of wireless CDMA networks occupy the same entire allocated spectrum to communicate at different transmission rates; therefore, the task of bandwidth management in wireless CDMA networks relies to a great extent on strategies for rate management.

5.1 Implementation of Flow Control

This research project explored packet-level access and scheduling for data-voice integration on the Integrated Wireless Access Network (IWAN) multicellular CDMA system. Different stream and packet services are transported on a bandwidth on demand basis. Packet data carried on an available basis does not require bandwidth reservation, while other resources are reserved.

The main result is the successful development of a common framework for different services which to a large extent allows flow control to be decoupled from the power control operation and the two to be separately optimized. No matter what the details of the power control algorithm are, as long as it can specify a measure of the available resources, flow control can utilize that measure to prevent admission of an excessive number of calls. It prevents overload and degradation in throughput as shown in Figure 36 and 37. The smoothness of the total transmitted power throughout the simulations demonstrates the robustness of the flow control in activating and blocking the data packets that contribute in an increased system throughput.

Flow control is controlled by bandwidth management logic that monitors the existing distribution of signal-to-interference ratios experienced by both the base and the mobile terminals. Packet data calls are activated if the estimated incremental signal to interference ratio does not violate the outage requirements in both directions.

The flow control mechanism enforces elasticity in traffic capacity both spatially and temporally. The traffic admitted to any cell is increased if the traffic in the neighboring cells happens to be low. Reduced voice traffic permits more packet data calls to be active. No reservations are needed for any type of traffic and the available bandwidth is utilized in a most efficient manner. The actual capacity of such a system will depend on the cell geometry, the propagation conditions, and the spatial distribution of the mobiles within the cell. The power-control based flow control shows good flexibility in adapting to the local conditions. The adaptive flow control algorithm is able to increase the capacity by a maximum of 7 packet data calls in 82, or 8.2%. By varying the thresholds T_{D1}, T_{D2}, T_{V1}, and T_{V2}, these figures might vary slightly. Higher thresholds introduce more interference, and lower values activate less calls in the system. These thresholds should be set according to the specifications and requirements of the system.

Another advantage of flow control is the arbitration of a good throughput delay trade-off. If a user is sending a high average message rate (in our studies this is equated with throughput), the resulting delays may be intolerably long. On the other hand, the user would not want to sacrifice too much throughput in order to achieve low delay. Related to this is the notion of fairly dividing network resources between competing network users.

Hence by providing flexible interfaces, new services requiring transmission rates in the available range can now be introduced without modifying the way existing services are handled. With distributed antenna structures, further gains in peak transmission rates may be achieved. At a bandwidth of 5 MHz, one can expect to integrate speech services at a rate as low as 4 Kb/s with data services at rates up to 512 Kb/s. On a bandwidth of 15 MHz, data rates up to 1.5 Mb/s are considered achievable.

5.2 Autocovariance in the Downlink Power Prediction

Different and variable rate calls corresponding to transmission of signals representing different media can be transmitted over a common access system without reserving capacity for any type of call. Total transmitted power is the common resource continuously allocated to the supported calls. Note that, under high loading, a voice call may not be admitted due to its more severe resource requirements, but a data packet might be admitted for transmission. Different priorities can be assigned to the two service types if the consequences of blocking one type of service are considered more serious.

Analytical capacity limits are difficult to obtain due to the time-varying nature of the channel and the sources. Dynamic simulations indicate useful capacity limits. Several packet data calls can be supported on top of voice traffic at bandwidths of 5 MHz and above. Given appropriate traffic models, the results allow the estimation of the number of subscribers in each service type that can be accommodated.

The total transmitted power is a stochastic variable. Its variance over 10 msec intervals depends on the number and rate of admitted terminals. The momentary value of the transmitted power provides insufficient information for call admission because its validity cannot be guaranteed for any time in the future. However, a combination of the power and its estimated probability over a small interval comparable to the estimated call duration provides a more complete picture.

With packet-level flow control, one can potentially transmit useful volumes of packet data even when no fixed mean data transmission rate can be guaranteed. Using constant bit rate (CBR) speech transmission with 100% activity, the system can accommodate about 67 mobiles. With available bit rate (ABR), we can accommodate up to 82 data mobiles receiving data at the same average rate of 8 Kb/sec. As more delay is allowed in the transmission of data services, throughput is higher.

Resource requirements of a packet to be transmitted successfully can be estimated from the distributions of the transmitted power near capacity, using short-term statistics like the autocovariance, when only speech calls are active.

5.3 So, what will you carry in 2010?

Our vision is that the personal terminal/communicator will be a single pocket-size device that will be a pager, telephone, computer, memo pad, address book, mail-box, television, etc. It will include a hands-free two-way audio interface and the primary input interface to it will be speech. It will also include a camera for visual input of all sorts. The supporting network will be essentially ubiquitous and seamless, with the physical location of personal files and contacts distributed and transparent to the user.

The industry is trying to anticipate what customers will want next. The real value is going to come when the device can be matched to the activity. That could mean voice recognition and interaction, such as talking to a laptop while driving to a meeting, asking it for directions as you go. If that's what consumers want in their portables, it won't be far off.

References

[Alavi82] H. Alavi and R. Nettleton, "Downstream Power Control for a Spread Spectrum Cellular Mobile Radio System," in *Proc. GLOBECOM'82*, pp. 3.5.1-3.5.5, 1982.

[Ariyavisitakul92] S. Ariyavisitakul, "SIR-based Power Control in a CDMA System," in *Proc. GLOBECOM'92*, pp. 868-873, 1992.

[ATM95] ATM Forum Technical Committee, "Traffic Management Working Group," *ATM Forum Traffic Management Specification*, Version 4.0, October 1995.

[Baier94] A. Baier, U. C. Fiebig, W. Granzow, W. Koch, P. Teder, and J. Thielecke, "Design Study for a CDMA-Based Third Generation Mobile Radio System," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 4, pp. 733-743, 1994.

[Brady68] P. T. Brady, "A Statistical Analysis of ON-OFF Patterns in 16 Conversations," *Bell System Technical Journal*, vol. 47, no. 1, pp. 73-91, 1968.

[Coviello75] G. J. Coviello and P. A. Vena, "Integration of Circuit/Packet Switching by a SENET Concept," *National Telecommunications Conference*, pp. 42.12-42.17, 1975.

[Das97] K. Das and S. D. Morgera, "Interference and SIR in Integrated Voice/Data Wireless DS-CDMA Networks-A Simulation Study," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 8, pp. 1527-1538, October 1997.

[Dziong96] Z. Dziong, M. Jia, and P. Mermelstein, "Adaptive Traffic Admission for Integrated Services in CDMA Wireless-Access Networks," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 9, pp. 1737-1747, Dec. 1996.

[Gilhousen91] 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 Trans. Vehic. Technol.*, vol. 40, no. 2, pp. 303-312, May 1991.

[Guo94] N. Guo and S. D. Morgera, "The Grade of Service for Integrated Voice/Data Wireless DS-CDMA Networks," in *Proc. IEEE ICC'94*, pp. 1104-1110, May 1994.

[Guo95] N. Guo, S. D.Morgera, and P. Mermelstein, "Common packet data channel (CPDC) architecture for CDMA integrated wireless access networks," *Sixth IEEE International Symposium on Personal, Indoor and Mobile radio Communications*, Part vol. 1, pp. 253-258, 1995.

[IS-95] "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System", *TIA / EIA /* IS-95-A, May 1995.

[IWAN] "Transport Level Description of the Integrated Wireless Access Network," *Canadian Institute for Telecommunications Research.*

[Jalali94] A. Jalali and P. Mermelstein, "Effects of Diversity, Power Control and Bandwidth on the Capacity of Microcellular CDMA Systems," *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 952-961, June 1994.

[Kandala96] S. Kandala and P. Mermelstein. "Integrated Voice and Video Services in Microcellular CDMA Systems - Downlink Power-based Call Admission," *Third Int. Workshop on Mobile Multimedia Communications*, Princeton, N.J., Sept. 1996.

[Lee91] W. C. Y. Lee, "Power Control in CDMA," in *Proc. VTC*, pp. 77-80, 1991.

[Ling97] T.-L. Ling and N. Shroff, "A Novel Flow Control Mechanism for ABR Traffic in ATM Networks," *ICC 97*, Montreal, Canada, vol. 2, pp. 1169-1173, 1997.

[Liu94] Z. Liu and M. E. Zarki, "SIR-based call admission control for DS-CDMA cellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 4, pp. 638-644, May 1994.

[Mandayam95] N. B. Mandayam and J. M. Holtzman, "Analysis of a Simple Protocol for Short Message Data Service in an Integrated Voice/Data CDMA System," in *MILCOM'95*, San Diego, CA, November 1995.

[Mermelstein93] P. Mermelstein, A. Jalali and H. Leib, "Integrated Services on Wireless Multiple Access Networks," in *Conference Record, ICC'93*, Geneva, Switzerland, pp. 863-867, 1993.

[Mermelstein95] P. Mermelstein and S. Kandala, "Capacity Estimates for Mixed-Rate Traffic on the Integrated Wireless Access Network," *Proc. PIMRC'95*, Toronto, Ont., pp.228-232, 1995.

[Mermelstein97] P. Mermelstein and S. Kandala, "An Integrated Multiservice Wireless Access Network," *Canadian Journal of Electrical Engineering*, vol. 22, pp. 41-46, 1997.

[Raychaudhari94] D. Raychaudhari and N. Wilson, "Multimedia Personal Communication Networks (PCN): System Design Issues," in *Proc. 3rd WINLAB Workshop Third Generation Wireless Inform. Networks*, pp. 259-288, 1994.

[Salmasi91] A. Salmasi and K. Gilhousen, "On the System Design Aspects of Code Division Multiple Access (CDMA) Applied to Digital Cel-

lular and Personal Communication Networks," in *Proc. VTC'91*, pp. 57-62, 1991.

[Sampath95] A. Sampath, P. S. Kumar, and J. M. Holtzman, "Power Control and Resource Management for a Multimedia CDMA Wireless System," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, v. 1, pp 21-25,.1995.

[Sampath97] A. Sampath and J. M. Holtzman, "Access Control of Data in Integrated Voice/Data CDMA Systems: Benefits and Tradeoffs," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 8, pp. 1511-1526, October 1997.

[Shim94] C. Shim, I. Ryoo, J. Lee, and S.B. Lee. "Modelling and Call Admission Control Algorithm of Variable Bit Rate Video in ATM Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 2, pp 332-344, February 1994.

[Soroushnejed95] M. Soroushnejed and E. Geranrotis, "Multi-Access Strategies for an Integrated Voice/Data CDMA Packet Radio Network," *IEEE Transactions on Communications*, vol. 43, no. 2, pp. 934-945, February 1995.

[Viterbi94] A. J. Viterbi, "Capacity of a Simple Stable Protocol for Short Message Service over a CDMA Network," in Blahut, editor, "*Communications & Cryptography*," pp. 423-429, KAP, 1994.

[Wireless97] "Wireless/Mobile Communications," *White Paper*, vol. 5, no. 3, February 1997.

[Wu88] C. S. Wu and V. O. K. Li, "Integrated Voice/Data Protocols for Satellite Channels," *Mobile Satellite Conference, jet Propulsion Laboratory*, pp. 413-422, May 1988.

[Wyrwas94] R. Wyrwas et al., "Multiple Access Options for Multimedia Wireless Systems," in *Third WINLAB Workshop on Third Generation Wireless Information Networks*, pp. 303-311, 1994.

[Yang94] W.-B. Yang and E. Geraniotis, "Admission Policies for Integrated Voice and Data Traffic in CDMA Packet Radio Networks," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 4, pp. 654-664, May 1994.

Bibliography

[1] D. J. Goodman, *Wireless Personal Communications Systems*, Addison-Wesley, 1997.

[2] V. K. Garg, K. Smolik, and J. E. Wilkes, *Applications of CDMA in Wireless/Personal Communications*, 1st ed. Upper Saddle River, NJ: Prentice-Hall, 1997.

[3] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall PTR, 1996.

[4] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Reading, MA: Addison-Wesley, 1995.

[5] D. Bertsekas and R. Gallager, *Data Networks*, Second Edition, Prentice Hall, 1992.

[6] J. G. Proakis, *Digital Communications*, New York: Mc Graw-Hill, 1989.

[7] S. Lin and D. Costello, *Error Control Coding*, Prentice-Hall, 1983.

[8] L. Kleinrock, *Queuing Systems, vol. 1: Theory*, Wiley-Interscience Publication, 1975.