Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

División de Electrónica, Computación, Información, y Comunicaciones

Programa de Graduados

Performance Analysis of cdma2000 MAC Layer with Voice/Data Integration

THESIS

Presented as a partial fulfillment of the requeriments for the degree of

Master of Science in Electronic Engineering

Major in Telecommunications

Fernando Agustín Domínguez Ferrer

Monterrey, N.L. June 2003

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

División de Electrónica, Computación, Información, y Comunicaciones

Programa de Graduados

The members of the thesis committee recommended the acceptance of the thesis of Fernando Agustín Domínguez Ferrer as a partial fulfillment of the requeriments for the degree of Master of Science in:

Electronic Engineering

Major in Telecommunications

THESIS COMMITTEE

César Vargas Rosales, Ph.D.
Advisor

José Ramón Rodríguez Cruz, Ph.D.
Synodal

Artemio Aguilar Coutiño, M.Sc.
Synodal

Approved

David Garza Salazar, Ph.D.
Director of the Graduate Program

June 2003

To my dear parents,
To my sister Mariela del Carmen,
I value your love, effort and support.
To my dear mother Dora María,
You are always in my heart.

Acknowledgments

I thank Instituto Tecnológico y de Estudios Superiores de Monterrey for the facilities provided. I mention specially my advisor, César Vargas Rosales, Ph. D., for his teaching, contribution and suggestions which make this work possible. I am grateful with José Ramón Rodríguez Cruz, Ph. D., and Artemio Aguilar Coutiño, M. Sc., for their interest and comments that have helped to enhance this work. To all my friends from the Center for Electronics and Telecommunications (CET).

This thesis is also dedicated to my friends and brothers from the AJEF, Pedro Armando, Carlos Fernando, Arturo, and of course, my dear Karla Paola. This work is also yours.

I want to mention specially my dear *brasileira* family , Eliana, Renata, Paula, Elizy, Junior, Gil. I miss you a lot.

FERNANDO AGUSTÍN DOMÍNGUEZ FERRER

Instituto Tecnológico y de Estudios Superiores de Monterrey
June 2003

Abstract

The growing demand for high speed packet data services and multimedia applications over mobile personal communication networks has set new design requirements and objectives for the Third Generation (3G) of air interface protocols. The asymmetric and bursty nature of multimedia packet data traffic, and the variability of data rates, packet sizes, and quality of service requirements make conventional voice-oriented channelization and access protocols inefficient. New Medium Access Control (MAC) and common channel messaging procedures need to be defined to efficiently support concurrent multiple packet and circuit-switch-based services and their related signaling messages.

The system cdma2000, which is an evolution from IS-95, is one of the strongest contenders for wireless data networking and provides next generation capacity while maintaining backward compatibility. cdma2000 also includes a sophisticated MAC feature which can concurrently support multiple data and voice services, thus effectively supporting very high data rates.

In this thesis we study the performance of the MAC states defined in cdma2000. The proposed model has the capability of proportionally combining three parameters: channel utilization, waiting time and the saving in the signalling overhead. We study the effect of integrating a voice service and two data services with different characteristics. Since the true nature of the wireless data is yet unknown, we use a mix of Poisson distributed voice packets and Pareto distributed data packets based on the model proposed in [7]. We derive analytical expressions and also obtain numerical evaluations to study the nature of the performance curve.

Resumen

La creciente demanda en los servicios de datos de alta velocidad y aplicaciones multimedia en redes de comunicación personal ha fijado nuevos requerimientos de diseño y objetivos para la Tercera Generación (3G) de protocolos inalámbricos. La naturaleza asimétrica y de ráfagas (burst) del tráfico de paquetes multimedia, la variabilidad de velocidades de datos, tamaños de paquetes y requerimientos de calidad de servicio, hacen que los protocolos tradicionales orientados a voz sean ineficientes. Es por esto que se requieren de nuevos esquemas de Control de Acceso al Medio (MAC).

cdma2000, que es una evolución de IS-95, es uno de los contendientes mas fuertes para el establecimiento de redes inalámbricas de datos y proporciona la capacidad para la Tercera Generación. cdma2000 incluye un control sofisticado MAC que puede soportar concurrentemente servicios múltiples de voz y datos.

En esta tesis, estudiamos el desempeño de los estados MAC definidos en cdma2000. El modelo propuesto tiene la capacidad de combinar proporcionalmente tres parámetros: utilización del canal, tiempo de espera y ahorro en la retransmisión de señalización. Se estudia el efecto de integrar un servicio de voz y dos servicios de datos con diferentes características. Debido que hasta ahora, la verdadera naturaleza del tráfico inalámbrico no es conocida, usamos una mezcla de paquetes de voz (con distribución de Poisson) y paquetes de datos (con distribución de Pareto). Se derivan expresiones analíticas y se obtienen evaluaciones numéricas para estudiar la naturaleza de la curva de desempeño.

Contents

List of Figures

List of Tables

Chapter 1

Background Concepts

This chapter begins with a review of the different spectrum sharing methods. The sharing of spectrum is required to achieve high capacity by simultaneously allocating the available bandwidth to multiple users. Code Division Multiple Access (CDMA) is based on the use of wideband spread spectrum techniques that enable the separation of signals that are coincident in time and frequency. The CDMA system (IS-95 standard) has been deployed worldwide because of its superior voice quality, robust performance, and large air interface capacity.

The design of the next generation air interface will build on this proven second generation air interface and will be backward compatible to current CDMA networks to protect the large capital investment in the already deployed base of networks. The third generation evolution of IS-95, known as cdma2000, uses a wideband CDMA technology to provide a graceful transition from second generation IS-95 systems. cdma2000 also includes sophisticated MAC features which can concurrently support multiple data and voice services.

1.1 Spectrum Sharing Concepts

Before we start with an overview of CDMA, we should review some basic concepts pertaining to spectrum sharing. Recall that there are three major techniques used for radio frequency (RF) spectrum utilization: (a) Frequency Division Multiple Access (FDMA); (b) Time Division Multiple Access (TDMA), and (c) Code Division Multiple Access (CDMA), [1].

First, frequency division multiplexing divides the bandwidth of the air interface between the Mobile Station (MS) and the Base Station (BS) into multiple analog channels; each radio frequency channel occupies one part or a larger frequency spectrum (see Figure 1.1). For example, some second generation systems divide a larger frequency spectrum of 10 MHz into multiple 30 kHz channels. This technique is called Frequency Division Multiple Access (FDMA).

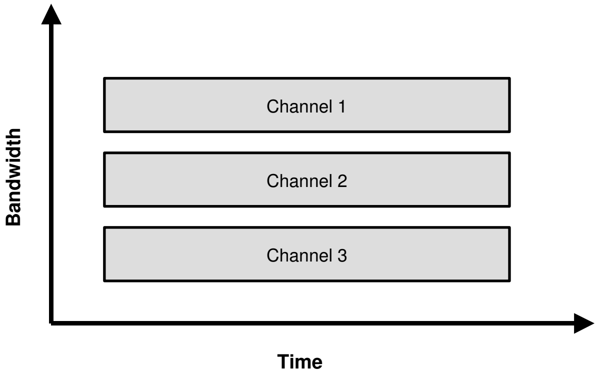

Figure 1.1: Frequency Division Multiple Access (FDMA)

Time Division Multiple Access (TDMA) operates on an RF channel but divides this analog channel into time slots, which contain digital traffic (see Figure 1.2). With TDMA, a user is given a digital time slot, and the slots are rotated among the users on a periodic basis. For example, user A might be assigned time slot 1 on a specific RF channel, user B could be assigned time slot 3 on the same channel, and so on. Each user is assured of having these slots available at a known time, which means the user's mobile knows the exact time to send (and stop sending traffic).

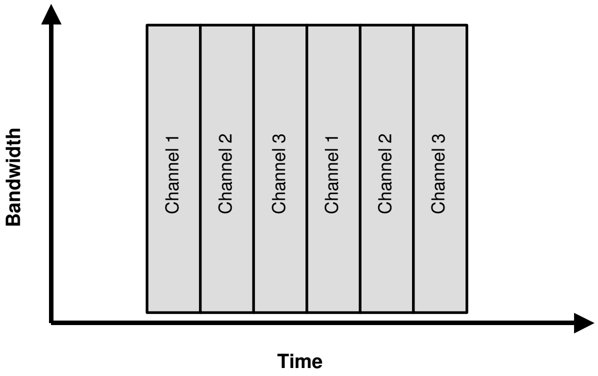

Figure 1.2: Time Division Multiple Access (TDMA)

The third major sharing technique is called Code Division Multiple Access (CDMA), depicted in Figure 1.3. This technology does not divide the time spectrum nor the frequency spectrum into pieces. Rather, CDMA places all users onto the same frequency spectrum at the same time. This concept uses a technique developed many years ago called spread spectrum, which means traffic is transmitted (spread) over the entire spectrum (typically across a wide bandwidth of 1.23 MHz). Each user is identified on the channel with a unique code. This code is used at the transmitting site to encode the traffic and it may also be used to spread it across the frequency spectrum. At the receiver, the code is used to extract the user's traffic.

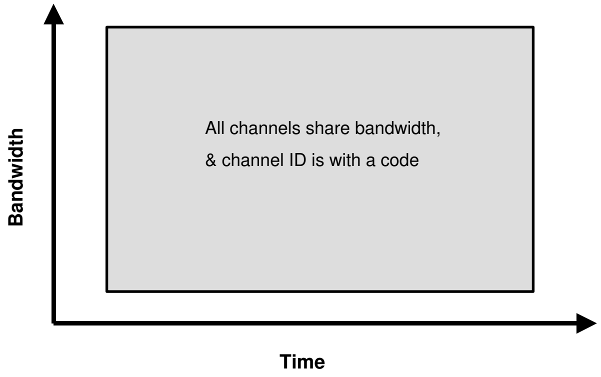

Figure 1.3: Code Division Multiple Access (CDMA)

1.2 Code Division Multiple Access (CDMA)

As we said before, CDMA is a form of spread spectrum, a family of digital communication techniques that have been used in military applications for many years. The core principle of spread spectrum is the use of noise like carrier waves, and, as the name implies, bandwidths much wider than that required for simple point to point communication at the same data rate, [2]. The goal of spread spectrum is a substantial increase in bandwidth of an information bearing signal, far beyond that needed for basic communication. The bandwidth increase, while not necessary for communication, can mitigate the harmful effects of interference, either deliberate, like a military jammer, or inadvertent, like co-channel users. The interference mitigation is a well known property of all spread spectrum systems. However the cooperative use of these techniques in a commercial, non-military, environment, to optimize spectral efficiency was a major conceptual advance.

Spread spectrum systems generally fall into one of two categories: Frequency Hopping (FH) or Direct Sequence (DS). In both cases synchronization of transmitter and receiver is required. Both forms can be regarded as using a pseudo-random carrier, but they create that carrier in different ways. Frequency Hopping is typically accomplished by rapid switching of fast settling frequency synthesizers in a pseudo-random pattern. CDMA uses a form of Direct Sequence.

Direct Sequence is, in essence, multiplication of a more conventional communication waveform by a pseudonoise (PN) ±1 binary sequence in the transmitter (see Figure 1.4).

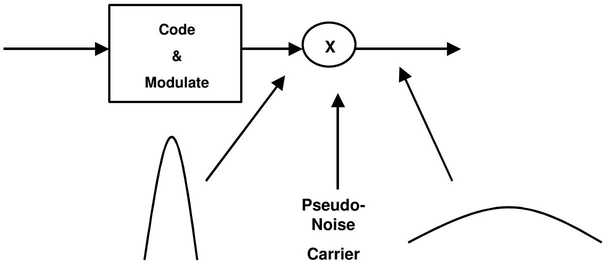

Figure 1.4: Multiplication by PN

We are taking some liberties with the details. In reality spreading takes place prior to any modulation, entirely in the binary domain, and the transmitted signals are carefully bandlimited. A second multiplication by a replica of the same ±1 sequence in the receiver recovers the original signal (see Figure 1.5). An important feature of CDMA is that typically uses a wide bandwidth of 1.23 MHz in each cell of the system, which translates into a frequency reuse factor of 1. So, the frequency planning concept is eliminated.

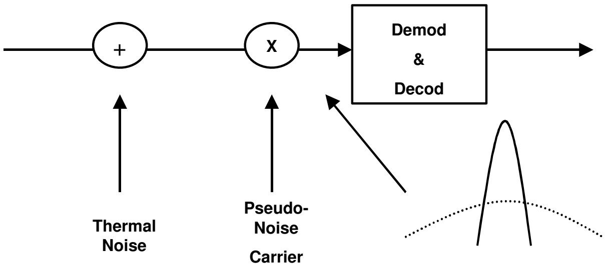

Figure 1.5: Recovering the Original Signal

We can resume that CDMA is a spread spectrum technology, which means that it spreads the information contained in a particular signal of interest over a much greater bandwidth than the original signal. When implemented in a cellular telephone system, CDMA technology offers numerous benefits to the cellular operators and their subscribers. The following is an overview of the benefits of CDMA:

- Capacity increases of 8 to 10 times that of an AMPS analog system.
- Improved call quality, with better and more consistent sound as compared to AMPS system.

- Simplified system planning through the use of the same frequency in every sector of every cell.
- Enhanced privacy.
- Improved coverage characteristics, allowing for the possibility of fewer cell sites.
- Increased talk time for portables.

1.3 The Third Generation (3G)

Second Generation (2G) wireless systems include TDMA (IS-136), CDMA (IS-95) and GSM (Global System for Mobile Communications). GSM is the mobile radio standard with the highest penetration worldwide. These 2G systems are limited in maximum data rate. On the other hand, the percentage of mobile multimedia users has increased significantly since 2000. It is expected a high grade of asymmetry between the demand for uplink and downlink of data transmission with much higher capacity needed on the downlink. More advanced services that current voice and low-data-rate services are foreseen and will bring together the three disciplines according to three basic categories:

- Computer data with Internet access, e-mail, real-time image transfer, multimedia document transfer, and mobile computing.
- Telecommunications with mobility, video conferencing, video telephony, and wide-band data services.
- Audiovisual content with video on demand, infotainment, electronic newspapers, teleshopping, value-added Internet services, television, and radio contribution.

Third Generation (3G) mobile radio networks have been under intense research and discussion recently, [3], [4], [5]. In the International Telecommunications Union (ITU), 3G networks are called International Mobile Telecommunications-2000 (IMT-2000); and in Europe, Universal Mobile Telecommunications System (UMTS). IMT-2000 will provide a multitude of services, specially multimedia and high-bit-rate packet data. Recently, extensive investigations have been carried out into the application of a CDMA system as an air interface multiple access scheme for IMT-2000/UMTS. It appears that CDMA is the strongest candidate for the third generation wireless personal communication systems. Emerging requirements for higher rate data services and better spectrum efficiency are the main drivers identified for the third generation.

The main objectives for the IMT-2000 air interface can be summarized as:

- Full coverage and mobility for 144 Kb/s, preferably 384 Kb/s.
- Limited coverage and mobility for 2 Mb/s.
- High spectrum efficiency compared to existing systems.
- High flexibility to introduce new services.
- Worldwide roaming capability.

1.4 cdma2000 System

The cdma2000 Radio Transmission Technology (RTT) is a wideband, spread spectrum radio interface that uses Code Division Multiple Access technology to satisfy the needs of 3G wireless communication systems, [6]. The service requirements are satisfied for indoor office, indoor-to-outdoor/pedestrian, and vehicular environments.

Within the standardization committee TIA TR45.5, the subcommittee TR45.5.4 was responsible for the selection of the basic cdma2000 concept. The goal has been to provide data rates that meet the IMT-2000 performance requirements of at least 144 Kb/s in a vehicular environment, 384 Kb/s in a pedestrian environment, and 2048 Kb/s in an indoor office environment. The main focus of standardization has been providing 144 Kb/s and 384 Kb/s with approximately 5 MHz bandwidth.

The evolution to cdma2000 3G services is not limited to current CDMA operators. cdma2000 is extremely attractive for TDMA operators because they already use the same core network as CDMA operators (the ANSI-41 standard). For GSM operators, the CDMA-MC to GSM MAP standard (IS-833) defines how the cdma2000 air interface can operate on the GSM MAP network, making this a technically feasible and economical approach to offering 3G services in a timely manner. In addition to mobile applications, cdma2000 may also be deployed in a fixed Wireless Local Loop (WLL) environment. cdma2000 is not constrained to only the IMT band; it is defined to operate in all existing allocated spectrum for wireless telecommunications, thereby maximizing flexibility for operators. Furthermore, cdma2000 delivers 3G services while occupying a very small amount of spectrum, protecting this precious resource for operators.

The key design characteristics of cdma2000 are [4]:

- Backward compatibility with IS-95A/B.
  - Support of IS-95-A/B signaling.
  - Support of IS-95A/B services as well as new services.
  - Spreading bandwidths compatible with IS-95A/B deployments.
- Fully supports handoff to and from existing systems.
- Support of different RF channel bandwidths of the form N x 1.23 MHz where N = 1, 3, 6, 9, 12.
- Includes an advanced Medium Access Control (MAC) layer.
- Supports different QoS characteristics.

We can resume that CDMA has the solution to meet the growing and changing demands of the wireless industry. The cdma2000 evolution path offers benefits to both voice and data users with increased voice capacity and increased data speeds. For the operator, the cost of upgrades can be shared between voice and data, as there is value in having data capability within the system while doubling voice capacity at the same time.

1.4.1 IS-95 and cdma2000 Layering Structure

As shown in Figure 1.6, IS-95 has a layered structure designed to provide voice, packet data (up to 64 kb/s), simple circuit data (e.g. asynchronous data and fax), and simultaneous voice and packet data service, [6]. At the most basic level, the cdma2000 RTT provides protocols and services that correspond to the bottom two layers of the International Organization for Standardization/Open Systems Interconnection (ISO/OSI) Reference Model (i.e. layer 1, the physical layer, and layer 2, the link layer) according to the general structure specified by the ITU for IMT-2000 RTTs. Layer 2 is further subdivided into the Link Access Control (LAC) and Medium Access Control (MAC) sublayers.

Motivated by higher bandwidths and the need to handle a wider variety of services, several enhancements have been incorporated into cdma2000 (these have been highlighted in Figure 1.6). In cdma2000, a fully generalized multimedia service model is supported. This allows virtually any combination of voice, packet data, and high-speed circuit data services to operate concurrently (within the limitations of the air interface system capacity).

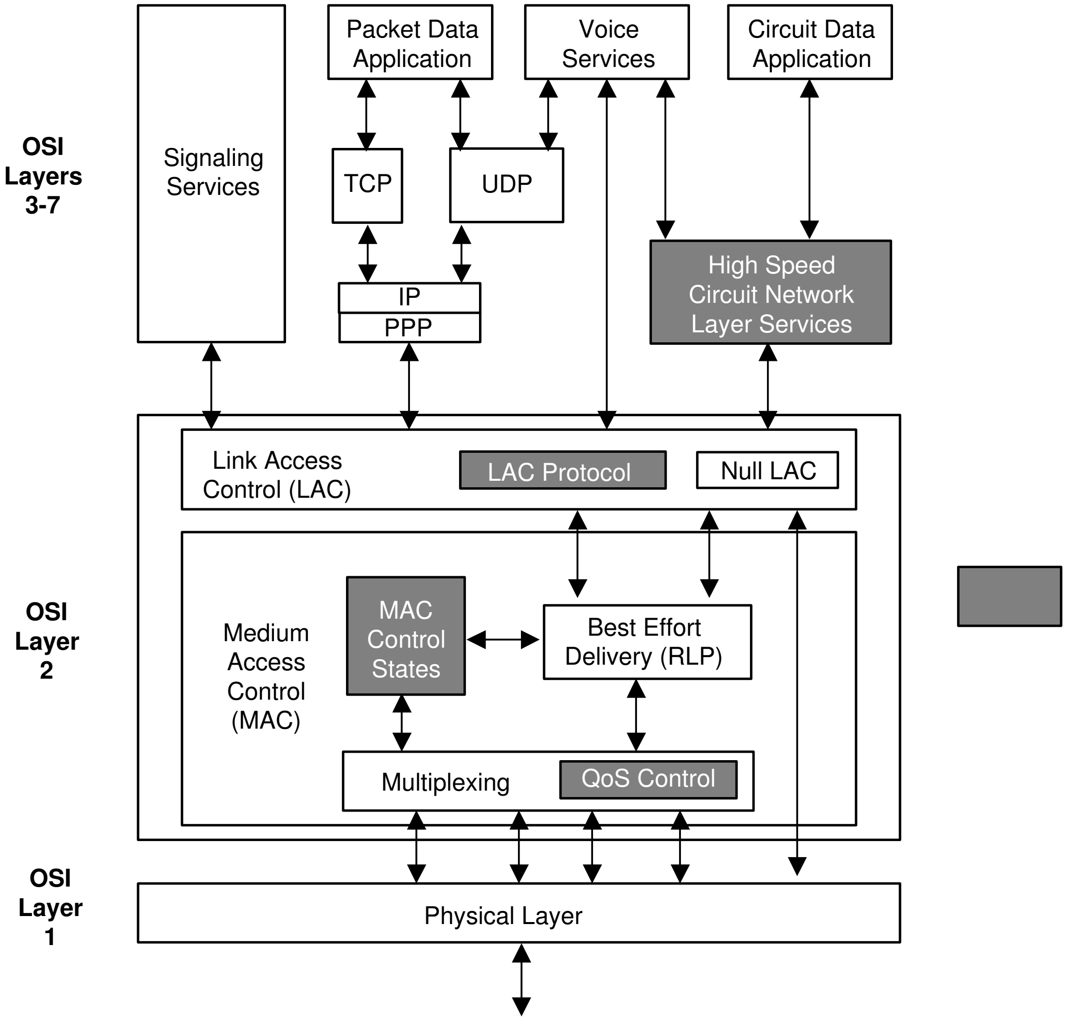

Figure 1.6: IS-95 and cdma2000 Layering Structure, [5]

The cdma2000 system also includes a sophisticated QoS control mechanism to balance the varying QoS requirements of multiple concurrent services, for example, to provide Broadband Integrated Services Digital Network (B-ISDN) or Resource Reservation Protocol (RSVP) QoS capabilities.

Most significantly, cdma2000 includes a flexible and efficient MAC entity that supports multiple instances of an advanced state machine, one for each active packet or circuit data instance. Along with the QoS control entity, the MAC realizes the complex multimedia, multiservice capabilities of next-generation wireless systems with QoS management capabilities for each active service. The cdma2000 system also introduces a true LAC protocol entity to support highly reliable point-to-point transmission over the air for signaling services and (optionally) for circuit data services. To provide a high degree of flexibility in the evolution of voice services, the cdma2000 RTT provides the framework and services to transport encoded voice in the form of packet data or circuit data traffic, in a manner that is backward compatible with previous IS-95 standards where the encoded voice is transported directly by the physical layer. In the latter example, the LAC and MAC services are considered to be null. The cdma2000 MAC sublayer provides three important functions:

- MAC Control States: Procedures for controlling the access of data services (packet and circuit) to the physical layer (including contention control between multiple services from a single user as well as between competing users in the wireless system).
- Best Effort Delivery: Reasonably reliable transmission over the radio link with a Radio Link Protocol (RLP) that provides a best effort level of reliability.
- Multiplexing and QoS Control: Enforcement of negotiated QoS levels by mediating conflicting requests from competing services and the appropriate prioritization of access requests.

1.4.2 Packet Data Service MAC Layer

In IS-95B, a new service option called the high-speed packet data service option is defined. This service option is established between the mobile and the InterWorking Function (IWF) when the mobile requests high-data-rate packet-mode service. During the service option negotiation procedure, the mobile must specify its high-speed data capability to the BS/MSC, that is, its capability in number of code channels on the forward and reverse channels. The BS then specifies the maximum number of codes on the forward and reverse channels it can provide to the mobile. High-data rate packet service is provided only when the service option is connected. The packet data MAC functions in IS-95 are based on the concept of providing packet data service over an underlying circuit-data-based call model.

In a somewhat simplified form, this MAC service can be reduced to only two states (see Figure 1.7).

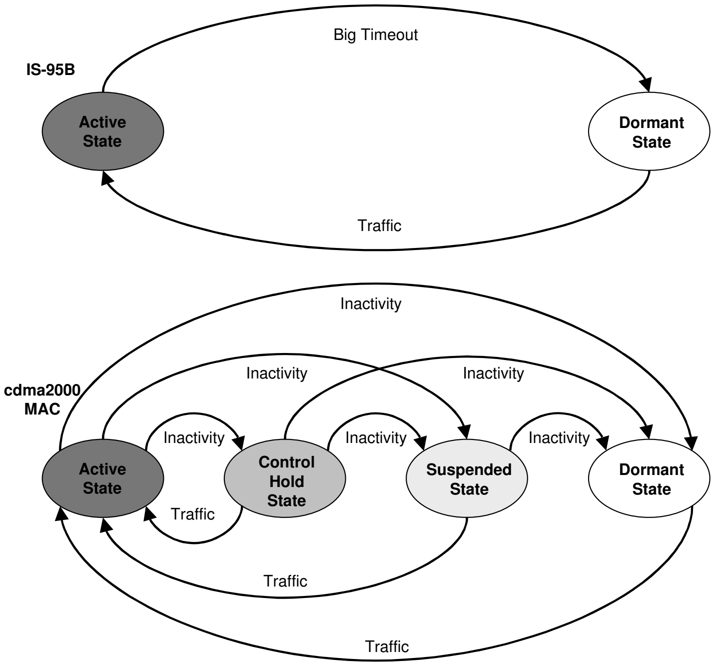

Figure 1.7: Comparison Between IS-95 and cdma2000 MAC's

- Active: In which a traffic channel is assigned to the mobile, and a link layer and Point-to-Point Protocol (PPP) connection is established between the IWF and the mobile.
- Dormant: In which no traffic channel is assigned to the call (however, the knowledge of the user's registration for packet data service is maintained along with the PPP connection).

User data can be transmitted only in the active state. However, since data is expected to be bursty with periods of inactivity, the user is allowed to remain in the active state for a short duration after a burst transmission is complete, to minimize the access time for the next burst. An inactivity timer is defined to conserve the air interface and other network resources. If this timer expires, the user is considered to be in the dormant state. In this state, there are no RF or system resources allocated to the user. The service option still remains connected, and the PPP connection between the mobile, IWF, and BS/MSC remains open during this period.

The packet data MAC functions for IS-95 have only two states, active and dormant, as described above. This simple approach works well for fairly low-speed data services with relatively low occupancy for any given user. However, this MAC model is inadequate to meet the aggressive requirements for very-high-speed data services with many competing users in third generation systems. This is due to the excessive interference caused by idle users in the active state, and the relatively long time and high system overhead required to transition from the dormant to the active state. To address these requirements, the cdma2000 system incorporates a sophisticated MAC mechanism that includes two intermediate states between the IS-95 active and dormant states (Figure 1.7).

- Control Hold. In this state, a dedicated control channel is maintained between the user and the BS on which MAC control commands (e.g. to begin a high-speed databurst) can be transmitted with virtually no latency. Power control is also maintained so that high-speed burst operation can begin with no delay due to stabilization of power control.

- Suspended State. In this state, no dedicated channels to or from the user are maintained; however, the state information for RLP is maintained, and the BS and the user maintain a virtual active set which permits either the user or the BS to know which BS can best be used (accessed by the user, or paged by the BS) in the event that packet data traffic for the user occurs. This state also supports a slotted substate that permits the user's mobile device to preserve power in a highly efficient manner.

In addition, a short data burst mode is added to the cdma2000 dormant state to support the delivery of short messages without incurring the overhead of transitioning from the dormant to the active state. Transitions between MAC states can be indicated by MAC control signaling or the expiration of timers. By carefully choosing the values for these timers, cdma2000 MAC can be adapted to a wide variety of data services and operating environments.

As we have seen the MAC structure for cdma2000 is an improvement over IS-95B MAC structure where only two states, the *active* and the *dormant* exist. The two states MAC works properly for fairly low speed data services with relatively low occupancy for any user given in IS-95B. However, the MAC is not adequate to meet the requirement of the Third Generation systems because of the long set up time and high system overhead from the dormant state to the active state. In the next Chapter we discuss all the attributes of the MAC states presented in this section. In section 3.2 we derive analytical expressions that model the performance of the MAC state transitions in order to obtain an optimal operating point.

Chapter 2

Model Description

As mentioned in Chapter 2, the MAC layer of cdma2000 provides extensive enhancements to negotiate and support concurrent services. It supports simultaneous voice/data operations without impacting voice quality or sacrificing high speed data performance. Voice can be mixed with high speed packet data, high speed circuit data, or a combination of multiple packet and/or circuit data services. In this Chapter we present analytical expressions that model the performance of the MAC state transitions. Since the true nature of the wireless traffic is yet unknown, we use a mix of Poisson distributed voice packets and Pareto distributed data packets with different shape parameters.

2.1 MAC State Transitions

The MAC states defined in cdma2000 are shown in the flow model in Figure 2.1. A data service is connected to the base station in the *active, control hold* and *suspended* states, but not in the *null* state. Transitions between these MAC states are controlled by *expiration timers.*

A node before the start of any service is in the null state and thus, in no way connected to the base station. When a node wants to initiate a session for a particular service, it sends a request to the scheduler at the base station for the allocation of a dedicated MAC channel to switch to the control hold state. The request contains all the information about the service to be initiated. In [7], the authors considered that wireless traffic would have packets from the telephony domain as well as the IP domain. However, 3G mobile radio networks will provide a multitude of services, specially multimedia and high bit rate packet data, thus, it is necessary to take into account that data services will have different characteristics (i.e. burstiness). Therefore, in this thesis we consider that the service request from the null state can be a voice service or two different data services. Our proposal can be seen in Figure 2.1. The packets in contention for the control hold state are a mix of voice packets (Poisson distributed), and two data services (Pareto distributed) with different burstiness (shape parameters).

After acquiring the MAC channel, the node goes into the control hold state, the service is established, and the MAC channel is retained till the end of the service, but it can not transmit data since the MAC channel is just a control channel.

Transitions between the MAC states are based on timer functions and the activity. By choosing proper timer values, the scheduler can switch the MAC states to accommodate more data services. When a data burst arrives it must acquire some traffic channels, the number of which depends on the bit rate requirement of that session.

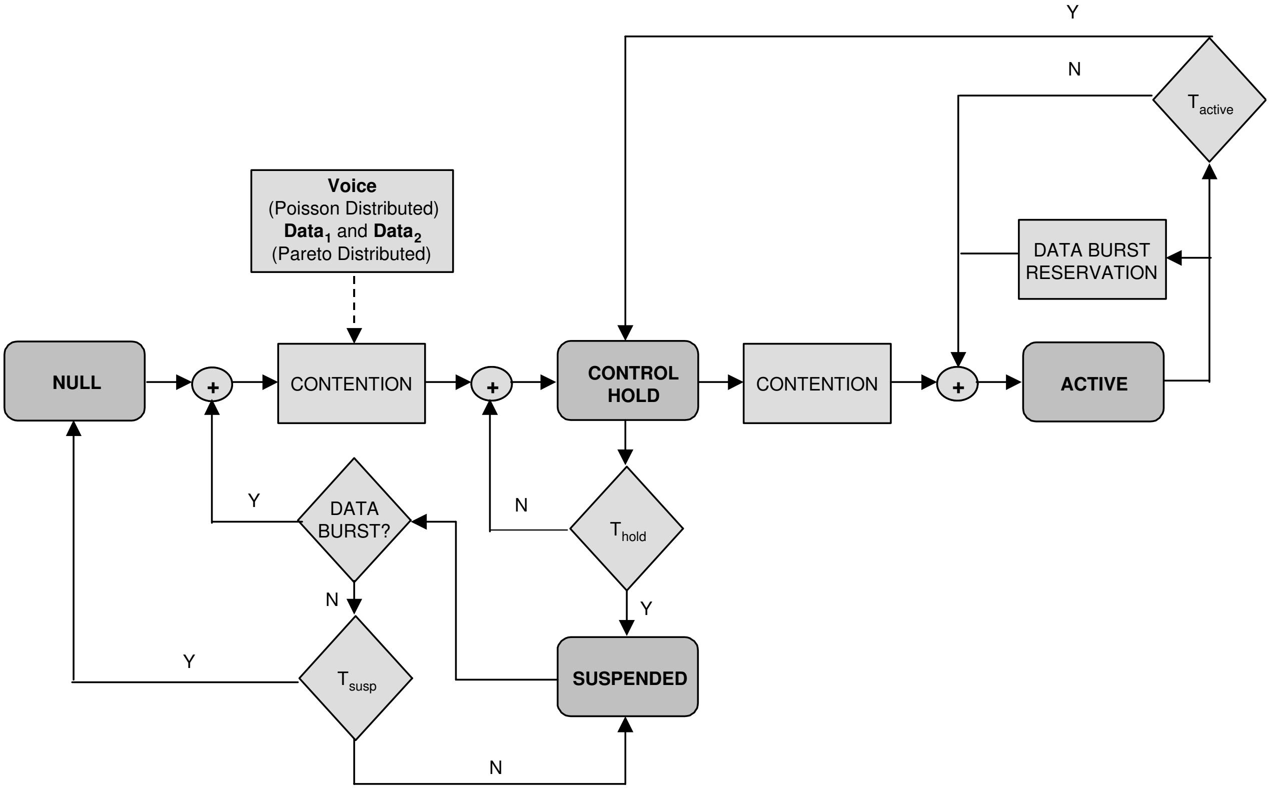

Figure 2.1: MAC Flow Model

If the scheduler can allocate the channels, the service goes into the active state where it starts transmitting data. The service remains in the active state till it has transmitted the data burst.

Now, it is important to know whether the service should release the dedicated traffic channels immediately or hold them for a certain amount of time after the burst has been transmitted. The time for which the state should be active for that service is represented by T_{active}. If there is a data burst within T_{active} then the same traffic channels are allocated.

If there is no activity within T_{active} then the traffic channels are released and the service goes back to the control hold state. This situation is illustrated in the example shown in Figure 2.2. In case 1, the second burst does not arrive within T_{active} after the first burst departs. Whereas in case 2, the second burst arrives within T_{active} and does not contends for the traffic channels.

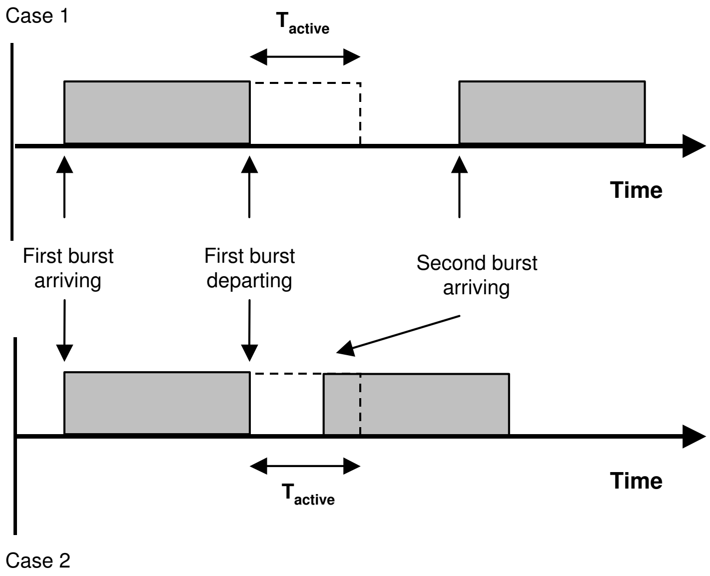

Figure 2.2: Data Bursts Arriving at Different Times

The performance optimization problem deals with the switching between the control hold state and active states. The cost factor associated with it, is the overhead due to this switching. If the cost is low, then, it is affordable to switch frequently between states, otherwise it might be desirable to remain in the active state in anticipation of new bursts and hold the traffic channels even if there is no data to be transmitted at this time. However, this results in wastage of resources that could have been used by other services.

On the other hand, if the traffic channels are released immediately after the transmission of the burst, then it has to contend for the traffic channels every time there is a burst which might incur additional delay.

Similar to traffic channels, the MAC channel can also be released if there is no activity for a long period of time, say T_{hold}, and the node can go into the suspended state as shown in Figure 2.1. In the suspended state the requirement profile is still maintained. If there is a burst from the suspended state the the node has to first get a MAC channel and follow the same procedure. From the suspended state, the node can go into the null state if there is no activity for the duration T_{susp}.

2.2 Performance Modeling

The performance of a MAC protocol is commonly measured in terms of the throughput and the waiting time. As in [7], we consider *channel utilization* (U) as the measure of throughput of the system. U is defined as the fraction of the channel capacity used. The *waiting time* (W) is as the average waiting time of the bursts before they acquire the traffic channels. Another parameter which needs to be taken into consideration is the *saving in signalling overhead* (S) prior to possibly every data exchange. Therefore, S is the fraction of the bursts arriving within time T_{active}. The three parameters U, W and S are functions of T_{active}. As in [7], in order to maximize this parameters is defined the overall system performance as a composite metric which is an additive function of the three basic parameters. We observe that $0 \leq U \leq 1$. The signalling overhead can be defined in such way that $0 \leq S \leq 1$. So, S is defined as the fraction of the bursts arriving within T_{active}. It also represents the probability that the arriving bursts finds that traffic channels are being held. Thus, U and S can be bounded, but W can be unbounded in general. We observe that the expected waiting time is at least half the frame time because any arriving burst has to wait at least till the beginning of the next frame. Thus, a function is defined which is the reciprocal of the waiting time and is bounded within the same interval as the other two parameters. This function is written as $\frac{0.5}{W}$ such that $0 \leq \frac{0.5}{W} \leq 1$.

In order to find an optimal point of operation for T_{active} we need an objective function. The objective function used in the model is the sum of the individual parameters U, S and W. Therefore, we have a lineal function that needs to be optimized. A first approximation that ensures that all the parameters collectively optimize the overall system performance is

$$maximize \left(U + S + \frac{0.5}{W} \right). \tag{2.1}$$

A scaling factor for each of these parameters is needed so that their relative contributions are normalized. Scaling functions $f()$, expressed in terms of T_{active}, are used. Therefore, the objective function for the overall system performance is given by

$$Z = f_u(T_{active})U + f_s(T_{active})S + f_w(T_{active})\frac{0.5}{W}, \tag{2.2}$$

where f_u, f_s and f_w are the scaling functions of channel utilization, saving in signalling overhead and waiting time, respectively.

2.2.1 Network Traffic Model

It is known that modeling network traffic using a Poisson or Markovian arrival process is common due to its simplicity and some favorable properties. However, over the last decade, several experimental observations showed that Poisson models fail to describe the behavior and/or statistics of computer network traffic such as Ethernet, [8] . Careful statistical analysis of data collected from experiments on the Ethernet LAN traffic for long durations, [9], has shown that the data exhibit properties of self-similarity, [10], and that there is long-range dependencies among the data. On the traffic control aspect, self-similarity implies the existence of correlation structure at a distance. Self-similarity is the conservation of a property of an object (a time series for example) with respect to scaling in space and/or time. It has been observed that the Ethernet traffic is bursty over a wide range of time scales, and can usually be generated by heavy tailed distributions with infinite variance. Pareto distribution is one such distribution with heavy tail and large burstiness.

The self-similar nature of Ethernet traffic is different both from the conventional traffic models and also from the currently considered formal methods of packet traffic. There is a considerable debate over the actual modeling of network traffic because it has serious implications on the design and analysis of networks. Just considering either of the two distributions (Poisson and Pareto) will not truly represent the nature of wireless data and therefore characterization becomes difficult. In this work we assume that wireless data will not strictly follow Poisson or Pareto distribution, but will have components from both.

As we said before, it is expected that 3G networks will provide voice and various data services that have different characteristics. Therefore, we consider a mix of voice packets (Poisson distributed), and two different data services (Pareto distributed) named $data_1$ and $data_2$.

2.2.2 Voice and Data Fractions

In this work we will consider a mix of Poisson and Pareto distributions in order to model both voice and data services. As we know, cdma2000 offers various types of data services, therefore, it would be appropriate to evaluate a system which considers all possible ratios of voice and data components for a given traffic load. Let us define the voice fraction V_f as

$$V_f = \frac{voice\ load}{voice\ load + data_1\ load + data_2\ load}, \tag{2.3}$$

where $data_1$ and $data_2$ are services with different burstiness. Then, the two different data fractions are given by

$$D_{f1} = \frac{data_1\ load}{voice\ load + data_1\ load + data_2\ load}, \tag{2.4}$$

$$D_{f2} = \frac{data_2\ load}{voice\ load + data_1\ load + data_2\ load}. \tag{2.5}$$

2.2.3 Voice Model

A voice call has periods of activity (talk spurts) and inactivity (gaps). The duration of the talk spurts and gaps is modeled as exponentially distributed with mean duration τ_t (for talk spurts) and τ_g (for gaps), respectively. *Activity* is defined as the ratio of the total length of the burst to the total time, i.e.,

$$Activity = A = \frac{\tau_t}{\tau_t + \tau_g}. \tag{2.6}$$

If the length of the talk spurts is given by the exponential random variable $\mathcal{T}$ with mean τ_t, then, using the transformation method, $\mathcal{T}$ is obtained as

$$\mathcal{T} = -\tau_t \ln(1 - U). \tag{2.7}$$

where U (channel utilization) is uniformly distributed in [0,1]. Equation (2.7) can be simplified to

$$\mathcal{T} = -\tau_t \ln(U). \tag{2.8}$$

since $(1 - U)$ is also uniformly distributed in [0,1]. The duration of a voice session is also exponentially distributed with a certain mean.

2.2.4 Data Model

For modeling data traffic we use heavy tailed Pareto distribution. We assume that the active data spurts are independent and identically distributed according to the Pareto distribution with shape parameter α and scale parameter k. The cumulative distribution function for the Pareto distribution is given by

$$F_T(t) = 1 - \left(\frac{k}{t}\right)^{\alpha}, t \geq k. \tag{2.9}$$

Recall that we have two different data services ($data_1$ and $data_2$), which have different burstiness that can be controlled by changing the shape parameter α. The length of the data spurts for $data_1$ is given by the random variable $\mathcal{D}_1$, again, using the transformation method we obtain

$$\mathcal{D}_1 = k_1 e^{-\frac{\ln(U)}{\alpha_1}}, \tag{2.10}$$

where k_1 is the location parameter and α_1 the shape parameter for the service of $data_1$. Similarly, for $data_2$

$$\mathcal{D}_2 = k_2 e^{-\frac{\ln(U)}{\alpha_2}}. \tag{2.11}$$

The duration of a data session is also Pareto distributed with certain mean.

2.2.5 Expressions for S, U and W

In this section, we derive expressions for the parameters channel utilization (U), waiting time (W) and savings in signalling overhead (S), as functions of T_{active}. The notations used in the analysis are shown in Table 2.1.

Table 2.1: Description of Parameters

Parameter	Description
λ	Service Arrival Rate
L	Mean Service Length in seconds
l	Mean Burst Length in frames
d	Mean Inactivity Length in seconds
A	Activity
T	Frame Duration in seconds
r	Average Channel Requirement
C	Number of Traffic Channels
V_f	Voice Activity Fraction
D_{f1}	$Data_1$ Activity Fraction
D_{f2}	$Data_2$ Activity Fraction

The parameter λ is the Poisson arrival rate. For voice services L (mean duration of the session), is exponentially distributed and for data services is Pareto ditributed. Similarly, the mean burst length (l) for voice and data services is also exponentially and Pareto distributed respectively. d is the mean inactivity length and is given by Equation (2.12). Activity (A), the ratio of the total length of burst over total time is given by Equation (2.6). The average channel requirement (r), can take the values 1, 3, 6, 9 or 12, this is because cdma2000 has a multi-carrier feature in which a service can be allocated more than 1 traffic channel per transmission. C is the number of traffic channels available in the system. As mentioned, we are considering that wireless data will have packets from the telephony domain as well as two different data services. Therefore, the base station will receive both voice and data multiplexed. In order to study a system with different voice and data components we define V_f as the voice component offering load in the system, and, similarly, D_{f1} and D_{f2} are the fractions of the data components.

Expressions to Find Saving in Signalling Overhead (S)

The saving in the state switching overhead is the fraction of the bursts arriving within T_{active}. The probability of a burst arriving within T_{active} is given by $1 - e^{\frac{-T_{active}}{d}}$, where d is the mean inactivity length and is given by

$$d = \frac{l(1-A)}{A}, \tag{2.12}$$

where l is the mean burst length and A is the activity given by Equation (2.6).

The saving in the switching overhead due to voice packets (Poisson distributed) is given by

$$S_v = 1 - e^{\frac{-T_{active}}{d}}. \tag{2.13}$$

For the Pareto distributed data bursts, S_d is obtained from the Pareto cumulative distribution function given in (2.9). For $data_1$ and $data_2$ we have, respectively

$$S_{d1} = 1 - \left(\frac{k_1}{T_{active}}\right)^{\alpha_1}, \tag{2.14}$$

$$S_{d2} = 1 - \left(\frac{k_2}{T_{active}}\right)^{\alpha_2}. \tag{2.15}$$

The mean of the burst length is given by $\frac{\alpha k}{\alpha - 1}$. Thus, the mean inactivity length d is given by

$$d = \frac{\alpha k}{\alpha - 1}\frac{(1-A)}{A}. \tag{2.16}$$

Thus, the values of k_1 and k_2 are obtained as

$$k_1 = \frac{d(\alpha_1 - 1)A}{\alpha_1(1-A)}, \tag{2.17}$$

$$k_2 = \frac{d(\alpha_2 - 1)A}{\alpha_2(1-A)}, \tag{2.18}$$

for $data_1$ and $data_2$, respectively.

Finally, the overall saving in the overhead due to both voice and data is given by

$$S = V_f S_v + D_{f1} S_{d1} + D_{f2} S_{d2}. \tag{2.19}$$

where $V_f S_v$ is the fraction of bursts produced by voice users (Poisson distributed), arriving within time T_{active}. Similarly, $D_{f1}S_{d1}$ and $D_{f2}S_{d2}$ are the fraction of bursts produced by the services $data_1$ and $data_2$ (Pareto distributed), with different burstiness.

Expressions to Find Channel Utilization (U)

Channel utilization is defined as the fraction of the channel capacity used. It can be expressed as a ratio of the total load offered to the total system capacity. The maximum load handling capacity is C. The load offered by voice packets is $V_f \lambda ALr$ and by data packets is $D_{f1}\lambda ALr$ and $D_{f2}\lambda ALr$.

Now, in order to obtain the overall channel utilization we have to take into consideration two cases. First, if T_{active} were 0, then channel utilization for voice and data would be given by

$$U = \frac{\lambda ALr}{C}, \tag{2.20}$$

but if T_{active} is non-zero, some bursts might not get traffic channels due to channel holding by other services and the expected channel utilization will decrease, the amount of which will depend on T_{active}. For voice traffic the loss in the channel utilization is

$$\left(1 - e^{\frac{-T_{active}}{d}}\right) \frac{\lambda ALr}{C}. \tag{2.21}$$

Therefore, the total channel utilization U_v for voice traffic is given by

$$U_v = \frac{\lambda ALr}{C} - \frac{\lambda ALr}{C}\left(1 - e^{\frac{-T_{active}}{d}}\right). \tag{2.22}$$

or,

$$U_v = \frac{\lambda ALr}{C}\left(e^{\frac{-T_{active}}{d}}\right). \tag{2.23}$$

Similarly, the channel utilization for data traffic (U_{d1}, U_{d2}) is given by

$$U_{d1} = \frac{\lambda ALr}{C}\left(\frac{k_1}{T_{active}}\right)^{\alpha_1}, \tag{2.24}$$

$$U_{d2} = \frac{\lambda ALr}{C}\left(\frac{k_2}{T_{active}}\right)^{\alpha_2}. \tag{2.25}$$

Finally, the overall channel utilization U due to both voice and data services is given by

$$U = \frac{\lambda ALr}{C}\left[V_f\left(e^{\frac{-T_{active}}{d}}\right) + D_{f1}\left(\frac{k_1}{T_{active}}\right)^{\alpha_1} + D_{f2}\left(\frac{k_2}{T_{active}}\right)^{\alpha_2}\right]. \tag{2.26}$$

Expressions to Find Waiting Time (W)

The waiting time of a burst will depend on whether channels are being held for that burst or it has to go through the contention process. S is the probability that the burst finds reserved channels. The expected waiting time in this case is $\frac{T}{2}$ (where T is the frame duration), because the burst has to wait till the beginning of the next frame. It is also necessary to find the waiting time for the burst which go through the contention process to get the desired channels. We call p_1 the probability that a burst does not get the required number of channels; p_1 is the probability of a burst leaving the queue after waiting for one frame, this is the probability of success in Bernoulli trials. The probability of waiting for the second frame is independent of the fact that it waited for the first frame, so, we can consider the various frames as independent, identically Bernoulli random variables. Therefore, p_1 is given by

$$p_1 = \sum_{i=C-(r-1)}^{C} \binom{C}{i} U^i (1-U)^{C-i}, \tag{2.27}$$

where C is the total channel capacity, r is the average channel requirement which can be 1, 3, 6, 9 or 12 depending of the bit rate requested. U is the overall channel utilization given by Equation (2.26). A known result in Benoulli trials states that $\frac{1}{p_1}$ is the expected number of experiments to be performed before success is reached. Therefore, it follows that the number of frames a request waits before leaving the queue is given by $\frac{1}{p_1}$. Combining the two cases we obtain the overall waiting time W as

$$W = \left(\frac{T}{2}\right) S + \left(\frac{1}{p_1}\right)(1-S), \tag{2.28}$$

where T is the frame duration, S is the overall saving in the overhead and $\frac{1}{p_1}$ is the expected number of frames a request waits before getting a traffic channel.

Expressions to Find the Optimal T_{active}

In order to obtain the optimal value for T_{active} we have to maximize the cost function given in Equation (2.2). The optimal value is found by equating to zero the first derivative of the objective function

$$\frac{d}{dT_{active}} \left(U + S + \frac{0.5}{W}\right). \tag{2.29}$$

As we can observe, the above expression is a transcendental equation. A possible solution would be to use an iterative algorithm so that we could maximize Equation (2.29). In this work we chose to evaluate the optimal timer by simulation experiments.

In this Chapter we presented analytical expressions in order to obtain the performance of the MAC states in cdma2000. We derive equations for the three parameters, channel utilization, waiting time and saving in the signalling overhead. We considered that the network have three different requests, a voice service and two data services. In the next Chapter we obtain numerical evaluations so we can study the behavior of the performance curves for the different parameters.

Chapter 3

Numerical Analysis

In this Chapter, the results of the system throughput equations are presented for different scenarios. To evaluate the performance of cdma2000 networks we take into consideration an environment where there are multiple service requests. The requests are made for service connection establishment at a certain service arrival rate λ. As in [7], we take the value $\lambda = 0.005\ packets/frame$. A service admitted into the system specifies its requirement profile (voice/data) and the bit rate requirement. The generation of bursts for a voice service is Poisson distributed and those of data services ($data_1$ and $data_2$) are Pareto distributed with shape parameters α_1 and α_2, respectively. Remember that for the Pareto distribution, if the shape parameter $\alpha < 2$, then the distribution has infinite variance. And for $\alpha < 1$ the distribution has infinite mean. Therefore, for $data_1$ we use $\alpha_1 = 1.5$, which is a practical value, [11]. For $data_2$ we use different values for the shape parameter in order to have a service with different burstiness. For practical purposes we have $1 < \alpha_2 < 2$ for different scenarios. The session length for the voice service is Poisson distributed and for the two data services is Pareto distributed, respectively. We assume that the average number of channels requested r=6. The parameters used for simulation are given in Table 3.1.

In our work, the number of traffic channels is kept constant, but in actual deployment of CDMA systems they are more likely to be statistical. If a service is admitted in the system, then the service goes from the null state to the control hold state. Once a service is admitted, it goes to the active state and generates its bursts. To obtain the optimal amount of time the dedicated traffic channels need to be held even after the transmission of the burst, we conduct extensive simulation over a range of T_{active}.

3.1 Scenarios and Results

In this section we propose different scenarios in order to study the behavior of the parameters channel utilization (U), saving in signalling overhead (S) and waiting time (W).

Table 3.1: Simulation Parameters

Parameter	Value
Number of Traffic Channels (C)	120
Service Arrival Rate (λ)	0.005/frame
Activity (A)	0.3 - 0.9
Mean Service Length (L)	2 mins
Mean Burst Length (l)	30 frames
Frame Duration (T)	20 ms

Table 3.2: Parameters for Scenario 1

Parameter	Value
V_f	0.7
D_{f1}	0.1
D_{f2}	0.2
α_1	1.5
α_2	1.05, 1.3, 1.7, 1.95
A	0.3, 0.5, 0.7

3.1.1 Scenarios for Saving in Signalling (S)

We first analyze the behavior of saving in signalling overhead (S). In Table 3.2 we can see the parameters used for Scenario 1.

The voice fraction (V_f) is 0.7, this is because we are assuming that there will be more voice users than data users in the network. Thus, the data fraction for the $data_1$ and $data_2$ services is 0.1 and 0.2, respectively. As we said before, we used a shape parameter $\alpha_1 = 1.5$ for $data_1$ and for $data_2$ we study the effect of the shape parameter α_2 for different values. Figure 3.1 shows the performance for an activity A=0.3, 0.5 and 0.7. These different values of activity correspond to low activity, average activity and high activity. From the plots we can see that as T_{active} is increased, the savings in overhead are also increased. This is expected because the new bursts arriving find that the dedicated channels are being held and thus, they do not have to retransmit the usual signalling overhead.

The parameters for Scenario 2 are given in Table 3.3. Now, we are giving for the voice fraction a low value of 0.4 and increasing the $data_2$ fraction to 0.5. Similarly to Scenario 1, Figures 3.2, 3.3 and 3.4 show the performance for A=0.3, 0.5 and 0.7. We observe that the behavior of the curves is similar to that of Scenario 1, i.e., the more activity we have, the more savings in the signalling overhead.

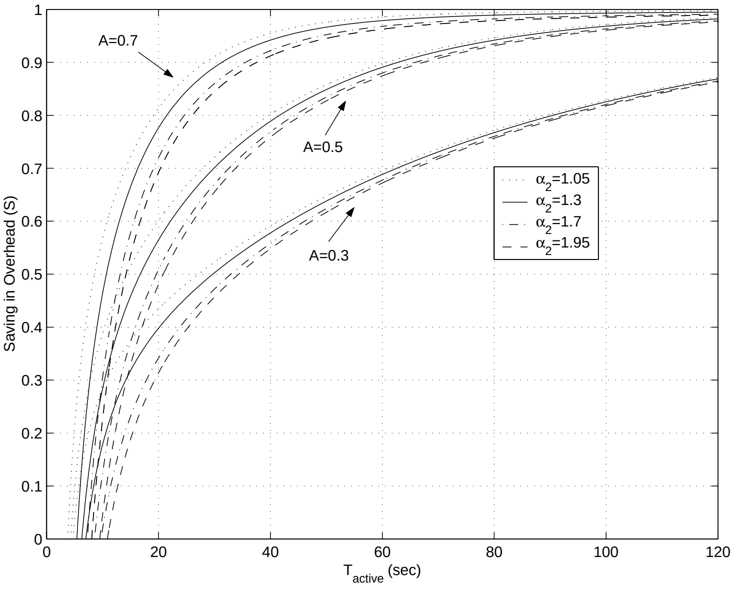

Figure 3.1: Scenario 1 with A=0.3, 0.5 and 0.7

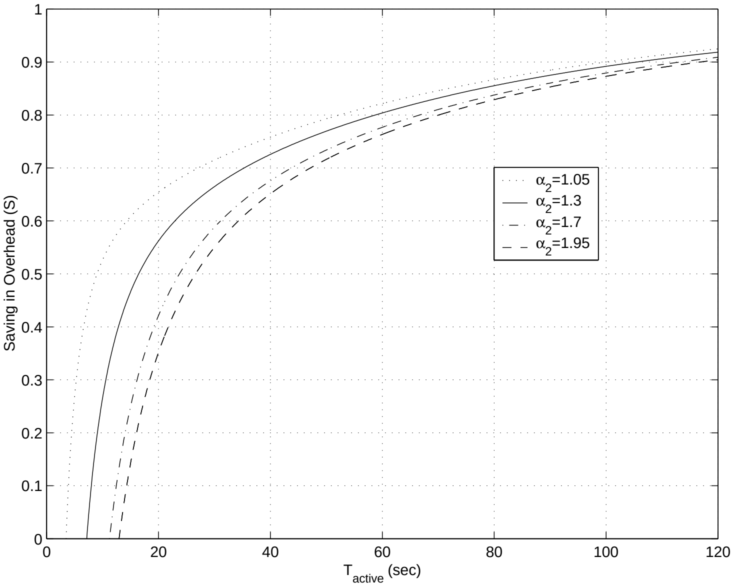

Figure 3.2: Scenario 2 with A=0.3

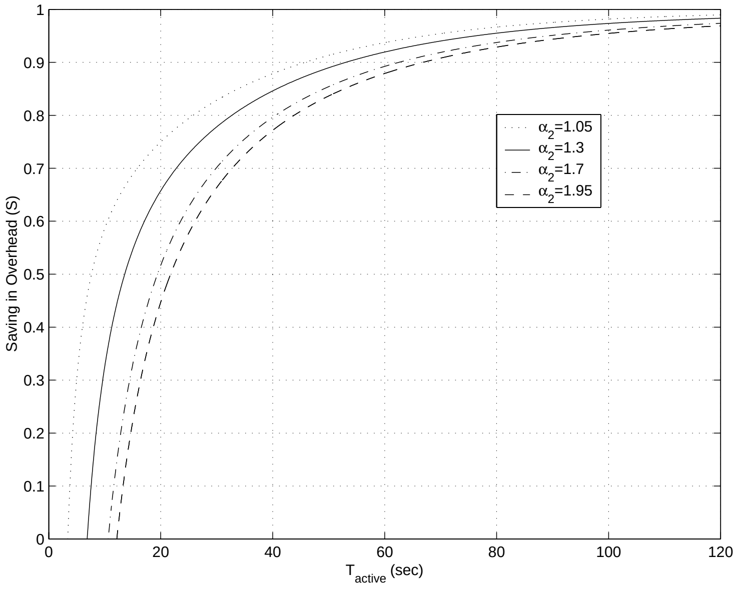

Figure 3.3: Scenario 2 with A=0.5

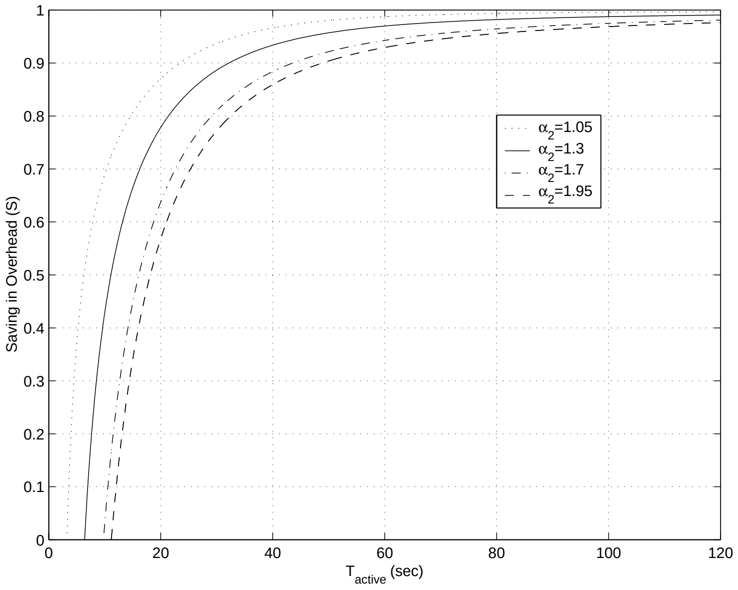

Figure 3.4: Scenario 2 with A=0.7

Table 3.3: Parameters for Scenario 2

Parameter	Value
V_f	0.4
D_{f1}	0.1
D_{f2}	0.5
α_1	1.5
α_2	1.05, 1.3, 1.7, 1.95
A	0.3, 0.5, 0.7

3.1.2 Results for Channel Utilization (U)

Now, we analyze the performance of the parameter channel utilization (U). We take the same values as those given in Table 3.2, with an average channel requirement of r=6 channels. Figure 3.5 shows the behavior for U with A=0.3, 0.5 and 0.7, respectively.

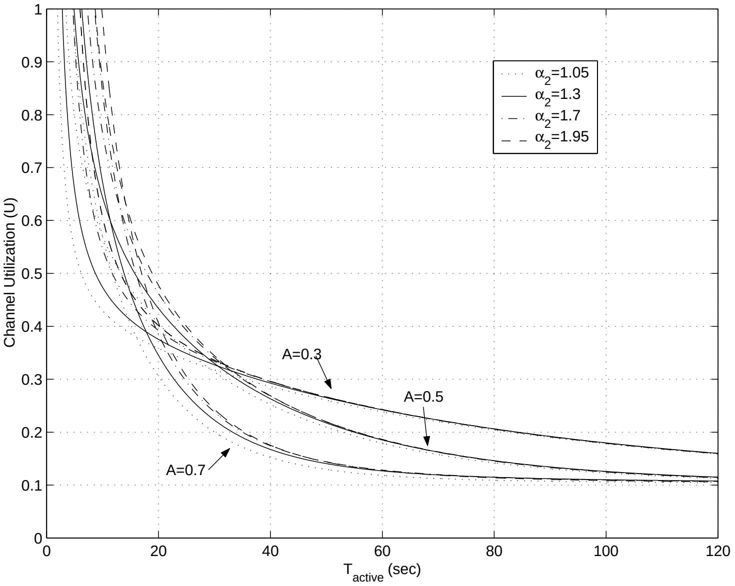

Figure 3.5: Channel Utilization (U) with A=0.3, 0.5 and 0.7

As in Scenario 2, we use the same parameters as those given in Table 3.3. The behavior of the channel utilization when varying the shape parameter α_2 can be seen in Figures 3.6-3.8. From the plots obtained we see that as T_{active} is increased the channel utilization falls. This is an expected result because the dedicated channels are held for longer time and they can not be used for other services, this results in wastage of resources.

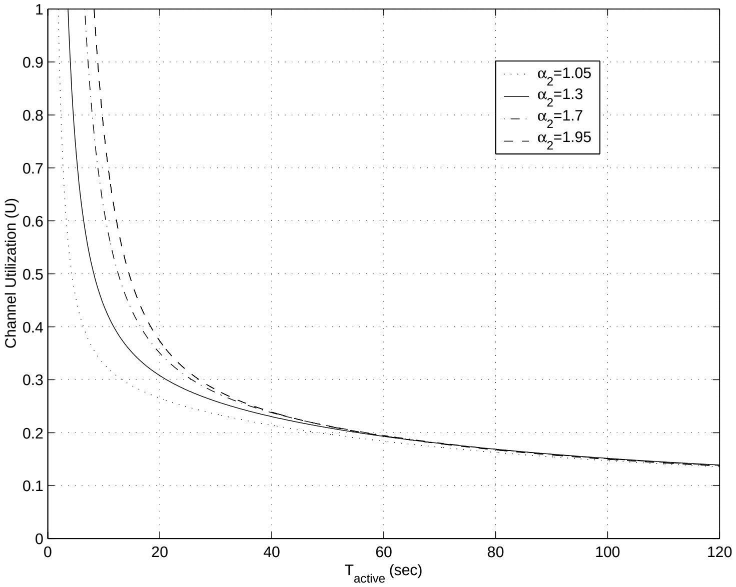

Figure 3.6: Channel Utilization (U) with A=0.3

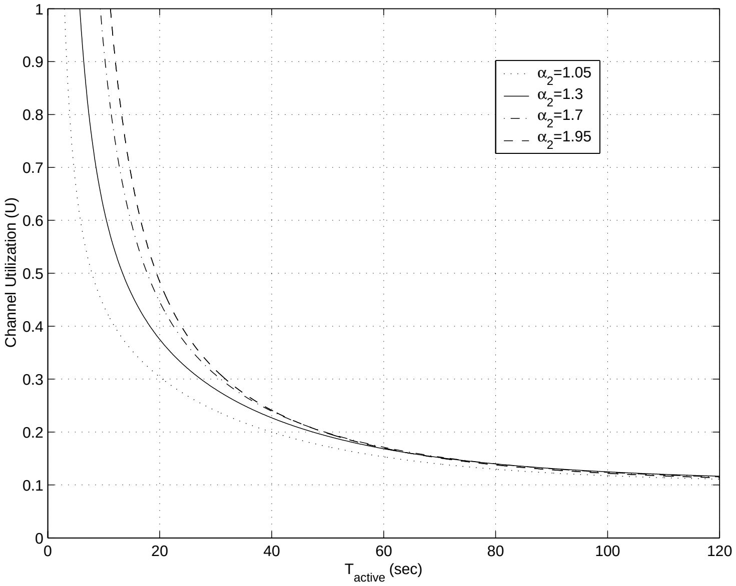

Figure 3.7: Channel Utilization (U) with A=0.5

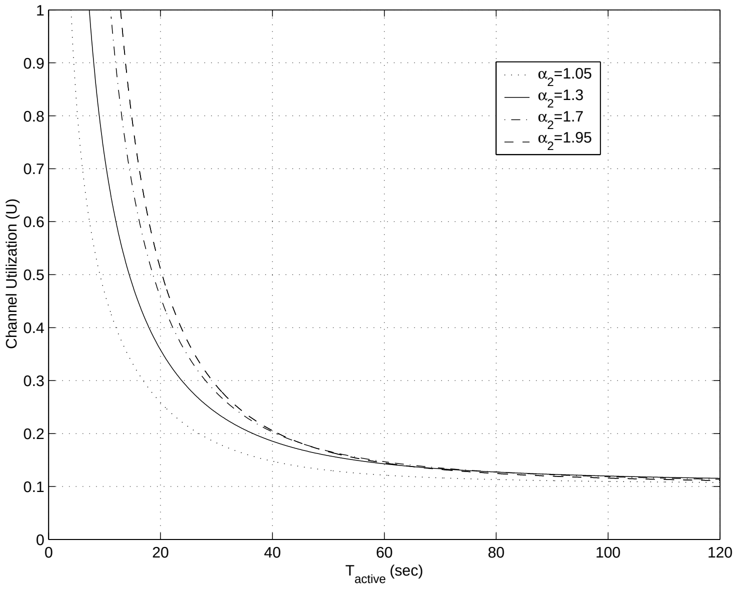

Figure 3.8: Channel Utilization (U) with A=0.7

3.1.3 Results for Waiting Time (W)

As we said before, the waiting time (W) will depend on whether channels are being held for the burst or it has to go through the contention process. In order to study the behavior of this parameter we plot W using the parameters given in Table 3.2, with r=6 channels and a frame duration T=20 ms. In Figure 3.9 we observe the waiting time for different shape parameters α_2 with activities A=0.3, 0.5 and 0.7.

Now we use the parameters of Table 3.3 where the voice fraction is reduced to 0.4 and the $data_2$ fraction increased to 0.5. Again, we observe in Figures 3.10, 3.11 and 3.12 the performance of W for different activities. As we can see, the curves for the waiting time decay with the increase in T_{active}. This mean that the bursts have to wait more time before they acquire the traffic channels.

As expected, the plots show that for low activity the waiting for the new bursts arriving is low. This is because the traffic channels are released in less time due to the average length of the bursts using them is not high. From the different scenarios studied in the plots we can see that there is a trade off between saving in signalling (S), channel utilization (U) and waiting time (W). In the next section we study the behavior of the overall system performance in order to find the optimal value for T_{active}.

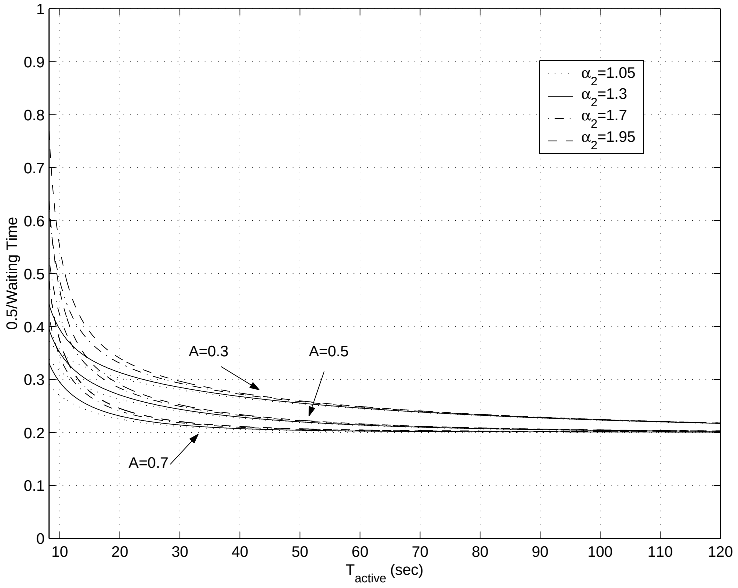

Figure 3.9: Waiting Time (W) with A=0.3, 0.5 and 0.7

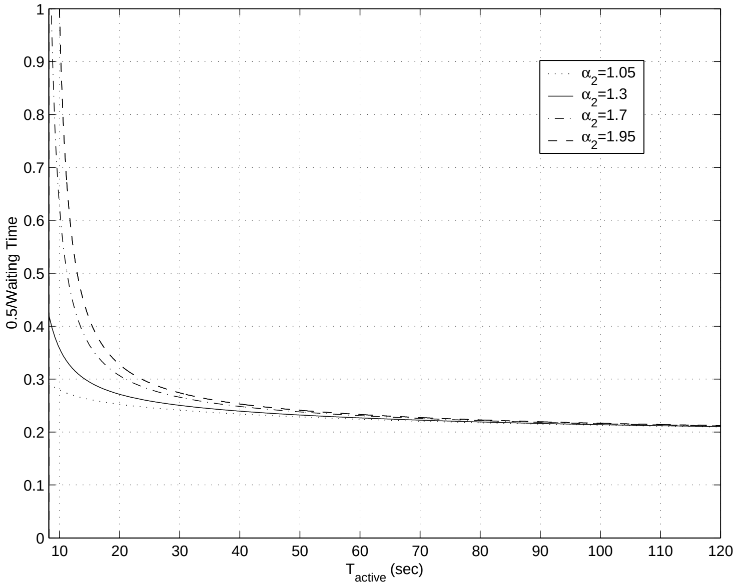

Figure 3.10: Waiting Time (W) with A=0.3

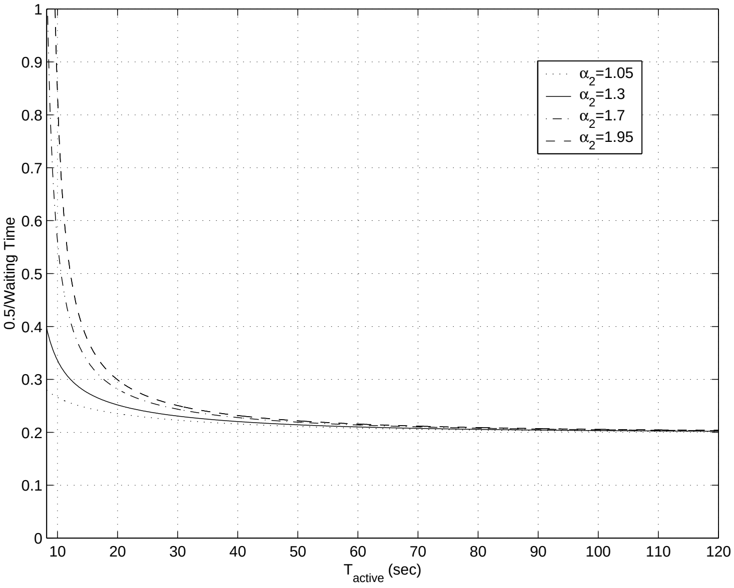

Figure 3.11: Waiting Time (W) with A=0.5

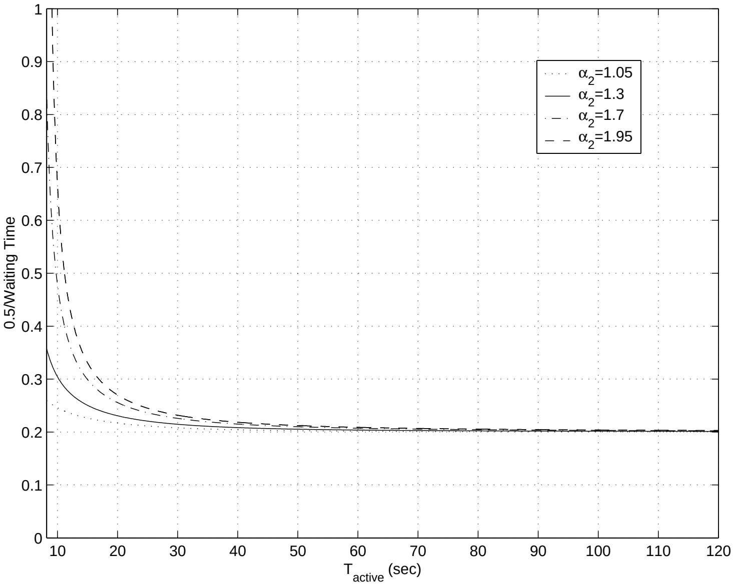

Figure 3.12: Waiting Time (W) with A=0.7

3.1.4 Results for the Overall System Performance

The overall system performance can be found equating to zero the first derivative of the sum of the individual parameters. The weights used for the scaling functions are: $f_u(T_{active}) = 1$, $f_s(T_{active}) = 1$ and $f_w(T_{active}) = 1$. First, the overall system performance is obtained for activities A=0.3, 0.5 and 0.7, (Figures 3.13, 3.14 and 3.15), with a voice fraction V_f=0.7. From the plots we can observe that as T_{active} increases the overall performance increases and attains a maxima. We note that for values of T_{active} >10 the performance falls. The region of T_{active} for which the overall performance is maximized depends on the values of the scaling functions used, but, whatever scaling functions are used, the performance will be the same.

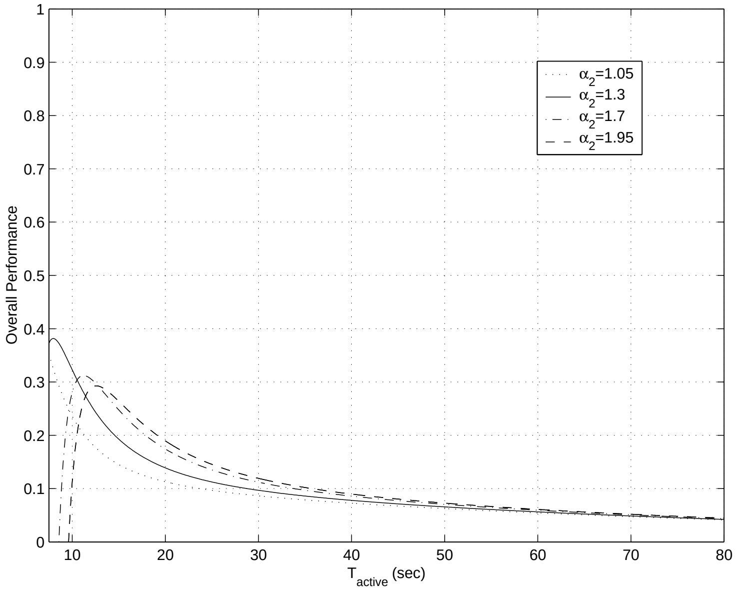

Figure 3.13: Overall Performance with A=0.3 and V_f=0.7

In order to study the effect on the overall performance with low voice fraction we use the values from Table 3.3. In Figures 3.16, 3.17 and 3.18 we vary the shape parameter α_2 with different activities. As expected, the plots show that as T_{active} increases, the overall performance has an optimal operating point, but after that, the performance is degraded.

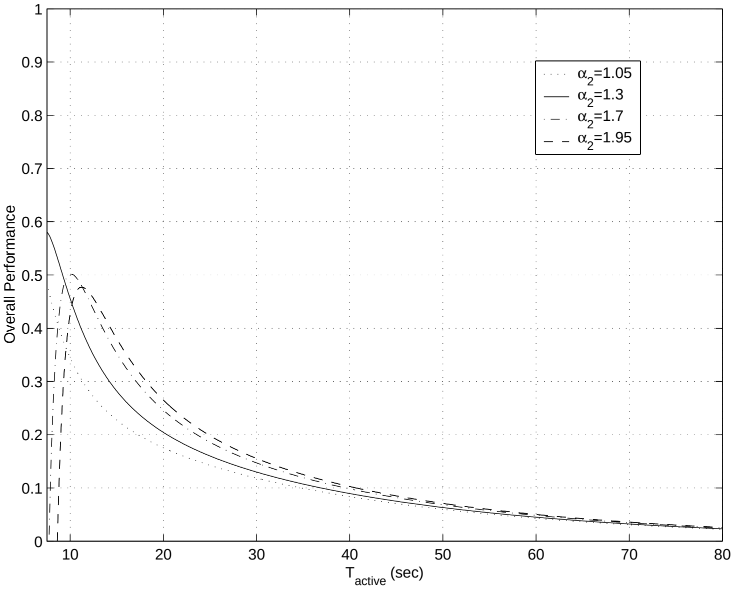

Figure 3.14: Overall Performance with A=0.5 and V_f=0.7

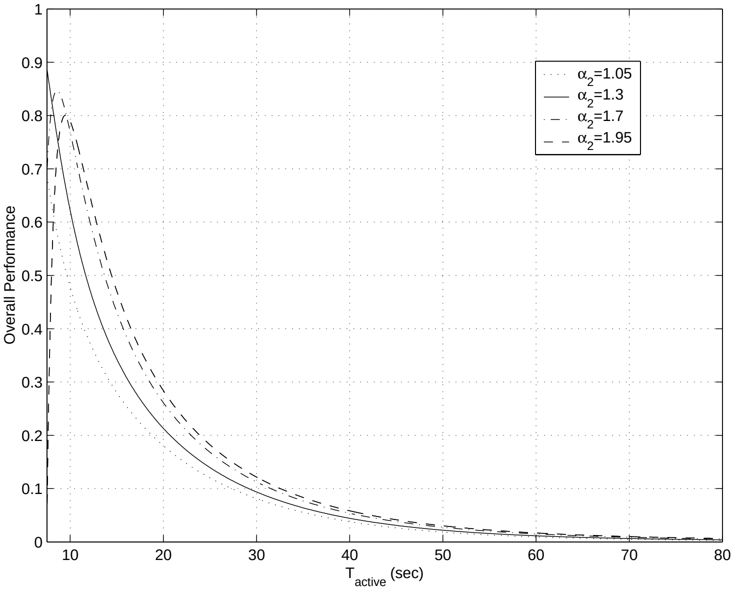

Figure 3.15: Overall Performance with A=0.7 and V_f=0.7

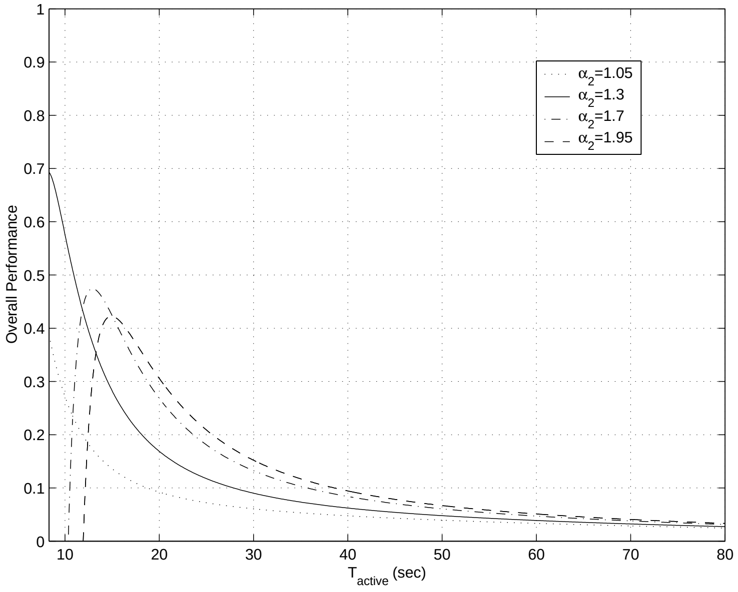

Figure 3.16: Overall Performance with A=0.3 and V_f=0.4

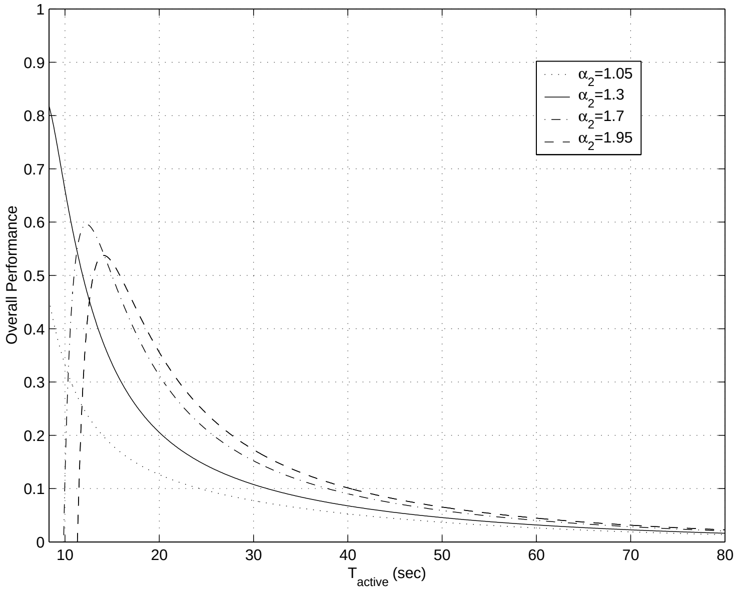

Figure 3.17: Overall Performance with A=0.5 and V_f=0.4

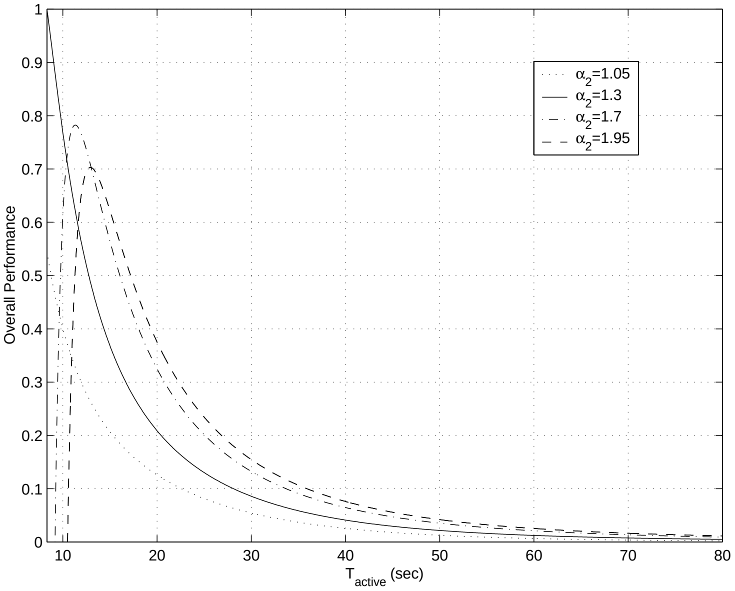

Figure 3.18: Overall Performance with A=0.7 and V_f=0.4

Bibliography

[1] Rappaport, Theodore S., *Wireless Communications. Principles & Practice,* Prentice Hall, 1999.

[2] Garg, Vijay K., *IS-95 CDMA and cdma2000. Cellular/PCS Systems Implementation,* Prentice Hall, 2000.

[3] Prasad, Ramjee, "An Overview of CDMA Evolution Toward Wideband CDMA," *IEEE Communications Surveys*, Vol. 1, No. 1, pp. 1-29, 1998.

[4] Rao, Y.S. and Kripalani, A., "cdma2000 Mobile Radio Access for IMT-2000," *IEEE International Conference on Personal Wireless Communications*, pp. 6-15, 1999.

[5] Knisely Douglas N. and Kumar Sarah, "Evolution of Wireless Data Services: IS-95 to cdma2000," *IEEE Communications Magazine*, Vol. 36, No. 10, pp. 140-149, October 1998.

[6] Garg, Vijay K., *Wireless Network Evolution. 2G to 3G,* Prentice Hall, 2002.

[7] Mainak Chatterjee and Sajal K. Das, "Optimal MAC State Switching for cdma2000 Networks," *IEEE INFOCOM 2002*, pp. 400-406, 2002.

[8] Vern Paxson and Sally Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, Vol. 3, No. 3, pp. 226-244, June 1995.

[9] W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Transactions on Networking*, Vol. 2, No. 1, pp. 1-15, February 1994.

[10] M.E. Corvella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *IEEE/ACM Transactions on Networking*, Vol. 5, No. 6, pp. 835-846, December 1997.

[11] Z. Hadzi-Velkov and L. Gavrilovska, “Performance of the IEEE 802.11 Wireless LANs under Influence of Hidden Terminals and Pareto Distributed Packet Traffic,” *IEEE International Conference on Personal Wireless Communications*, pp. 221-225, 1999.