Congestion Control Techniques In Hybrid ATM/CDMA Network

Ning Zhou

A Thesis

In

the Department

of

Electrical Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Applied Science
at Concordia University
Montreal, Quebec,
Canada H3G 1M8

April, 1996

ISBN 0-612-10917-8

ABSTRACT

Congestion Control Techniques in hybrid ATM/CDMA network

Ning Zhou

In this work we pursue a performance analysis to compute the various performance criteria in a hybrid Time Division/ Asynchronous Transfer Mode/Code Division Multiply Access Network, i.e. TDMA/ATM/CDMA Network. Users accessing this TDMA/ATM/CDMA uplink frame are assumed to belong to one of 4 service classes, namely video, voice, file and interactive data. Each user accesses only a portion of the subframe slots assigned to its class. A variabale frame boundary strategy is used to adjust the subframe boundaries depending on the call load . To alleviate congestion in the ass umed-hubless signalling free satellite network, the satellite measures the uplink traffic of each class and issues pilot congestion control indicators to on going calls of each class. These will be subsequently used by ground users to control their activities and police their calls using modified vers sions of Leaky Bucket and Virtual Leaky Bucket congestion control techniques. The new techniques alleviate many of difficulties of specific slot assignment, on-board call management, superframe-counting and management involved in the state of the art TDMA based systems, and yield a call establishmem free yet a dynamic and very well controlled access technique.

ACKNOWLEDGMENTS

I would like to express my profound gratitude to my supervisor Dr. A.K. Elhakeem, for providing the guidance and support that made this work possible. He has been always available to direct and advise whenever the occasion arose. It was an enjoyable experience working with someone so enthusiastic and insightful.

I would also like to express my deep gratitude to Dr. M.Kadoch for his constant inspiration and support throughout this work.

Many thanks are due to my family and friends for their help, advice and encouragement.

Contents

List of Figures

Chapter 1

Introduction

1.1 Spread Spectrum Communication

Spread Spectrum (SS) communication system can be typically defined as one in which the average energy of the transmitted signal is spread over a bandwidth which is much wider than the information bandwidth. This system has inherent advantages such as robust immunity to narrow band interference and jamming, achieving privacy in communication, low probability of intercept and high accurancy ranging. The system has been mainly used in the military field. However, recently, the other important advantage which is random access capability (spread spectrum multiple access: SSMA) is becoming closed up and several devices such as surface acoustic devices (SAWD), have been adopted. On the ground of these techniques, research and development of SS communication system have increased for commercial applications such as mobile communications, consumer communications and so on.[37]

Spread spectrum techniques are divided into four basic types:

1. Direct Sequence (DS).
2. Frequency Hopping (FH).
3. Time Hopping (TH).
4. Chirp.

Combining two or more of the four techniques results in a technique called hybrid system. One reason for using hybrid techniques is that some of the advantages of types of systems are combined in a single system. Hybrid techniques are widely used in millitary spread spectrum systems and are currently the only practical way of achieving extremely wide spectrum spreading.

1.2 Direct Sequence System (DS)

In a conventional DS system, the spread signal is obtained by multiplying directly the input data (digital data of ± 1, time slot duration T sec) with a pseudo-noise (PN) sequence (elements of ± 1, the code length N, chip duration T_c, period $T = NT_c$) as show in Figure 1.1. Bandwidth of spreading signal is equal to bandwidth of PN sequence.[39]

Figure 1.2 is a functional block diagram of a direct-sequence system with binary phase modulation. This system, which provides message privacy, is the most widely used direct-sequence implementation in practice. Synchronized data symbols, which may be information bits or binary code symbols, are modulo-2 added to chips before the phase modulation. A coherent phase-shift keying (PSK) demodulator may be used in the receiver. Alternatively, if the uncertainty in the carrier frequency is sufficiently small, a differential phase-shift keying demodulator may be used.

The receiver spread-spectrum signal can be represented by

$$r(t) = \sqrt{2E}m(t)p(t)cos(\omega_o t + \theta) \tag{1.1}$$

where E is the signal power, m(t) is the data modulation, p(t) is the spreading waveform, ω_o is the carrier frequency, and θ is the phase angle at t=0. The data modulation is a sequence of nonoverlapping rectangular pulses, each of which has an amplitude equal to +1 or -1. Each pulse of m(t) represents a data symbol and has a duration of T_s. Each pulse of p(t) represents a chip, is usually rectangular with an amplitude equal to +1 or -1,

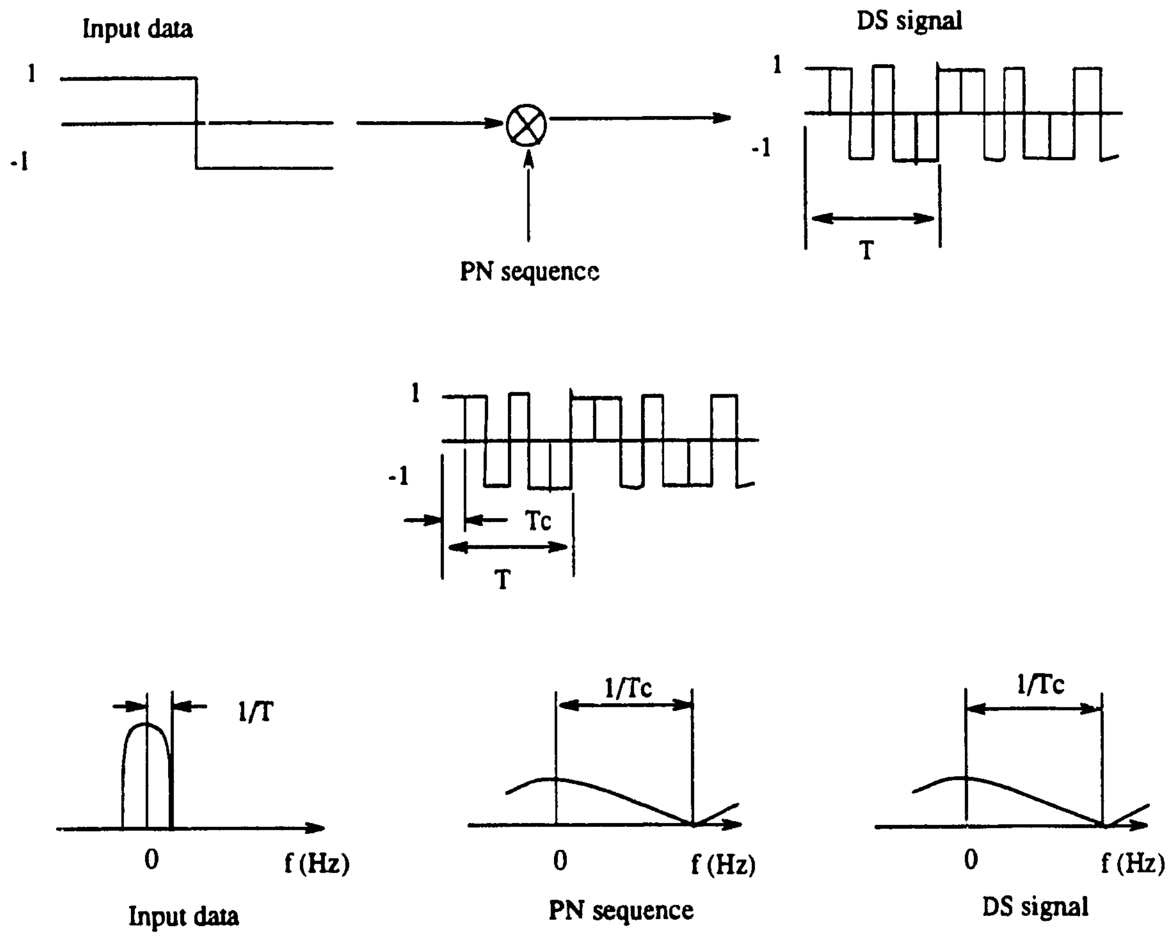


Figure 1.1: DS Signal and its spectrum

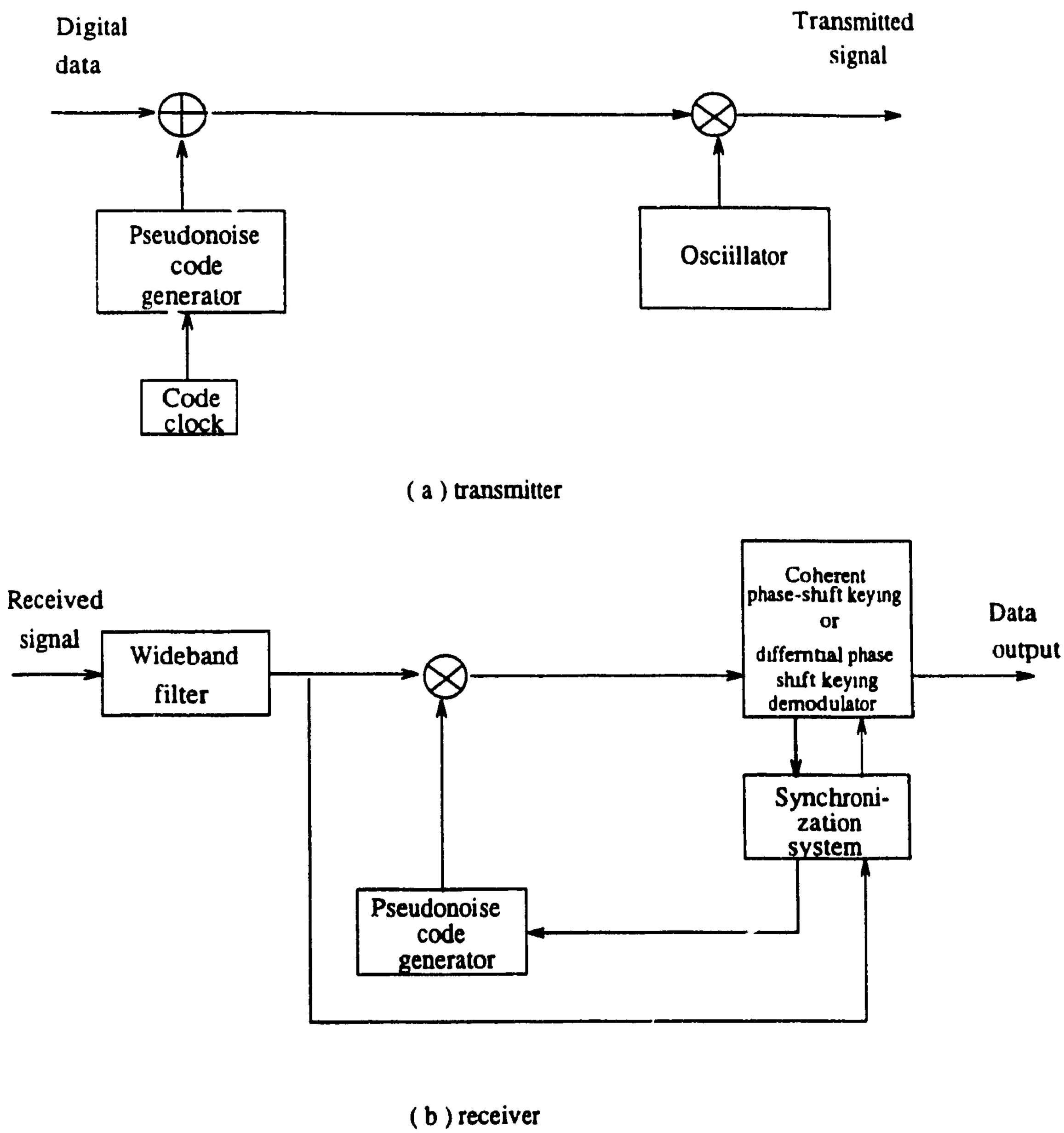


Figure 1.2: Direct-sequence with binary phase modulation

and has a duration of T_c. Since message privacy requires that the transitions of the data symbol and the chips coincide on both sides of a symbol, the ratio of T_s to T_c is an integer. If W is the bandwidth of r(t) and B is the bandwidth of $m(t)cos\omega_0 t$, the spreading due to p(t) gives $W \geq B$.

Assuming that code synchronization has been established, the received signal passes through the wideband filter and is multiplied by a local replica of p(t). If $p(t) = \pm 1$, then $p^2(t) = 1$, and this multiplication yield the despread signal

$$r_1(t) = \sqrt{2E}m(t)cos(\omega_0 t + \theta) \tag{1.2}$$

at the input of the demodulator. Because $r_1(t)$ has the form of a PSK signal, the corresponding demodulation extracts m(t).

The receiver reduces interference as qualitatively illustrated in Figure 1.3; quantitative results are given subsequently. Figure 1.3(a) shows the relative spectra of the desired signal and interference at the output of the wideband filter. Multiplication by the spreading waveform produces the spectra of Figure 1.3(b) at the demodulator input. The signal bandwidth is reduced to B, while the interference enegy is spread over a bandwidth exceeding W. The filtering action of the demodulator removes most of the interference spectrum that does not overlap the signal spectrum. Thus, most of the original interference enegy is eliminated and does not affect the receiver performance. An approximate measure of the interference rejection capability is given by the ratio W/B. The processing gain, defined by

$$G = \frac{T_s}{T_c} \tag{1.3}$$

is equal to the number of chips in a symbol interval. Whatever the precise definition of a bandwidth, W and B are usually proportional to $1/T_c$ and $1/T_s$, respectively.

$$G = \frac{W}{B} \tag{1.4}$$

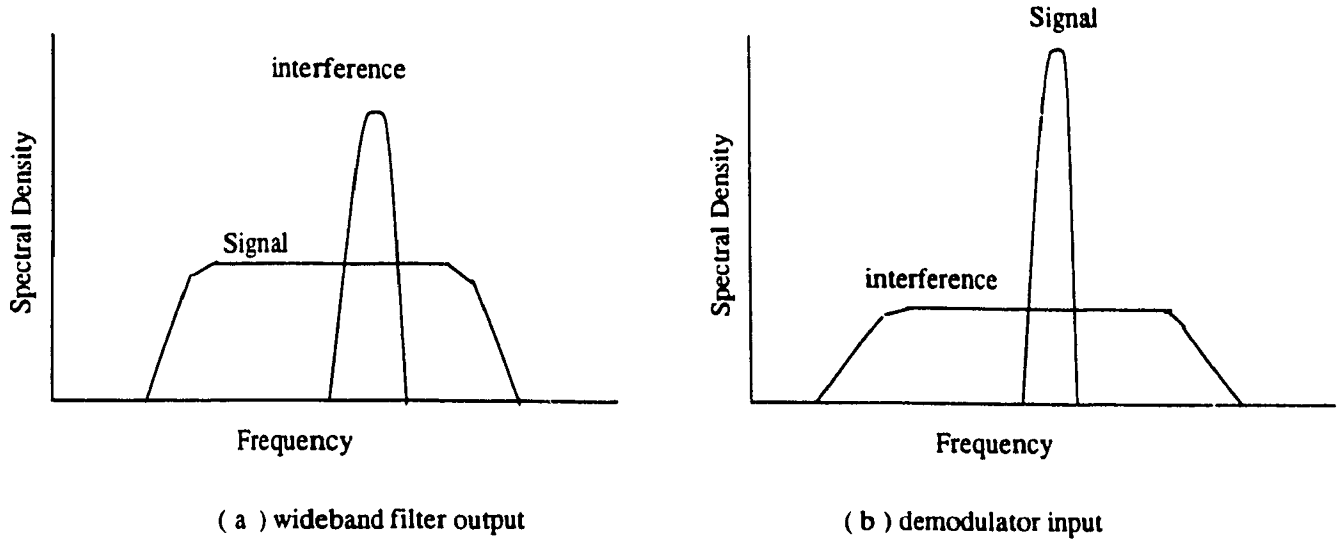


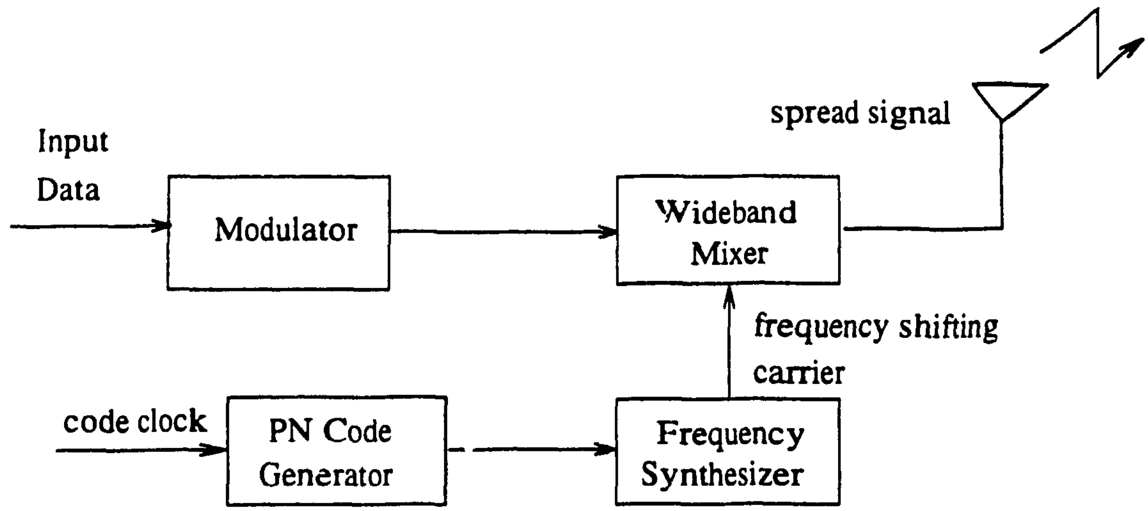
Figure 1.3: Spectrum of desired signal and interference

which links the processing gain with the interference rejection illustrated in the Figure 1.3.

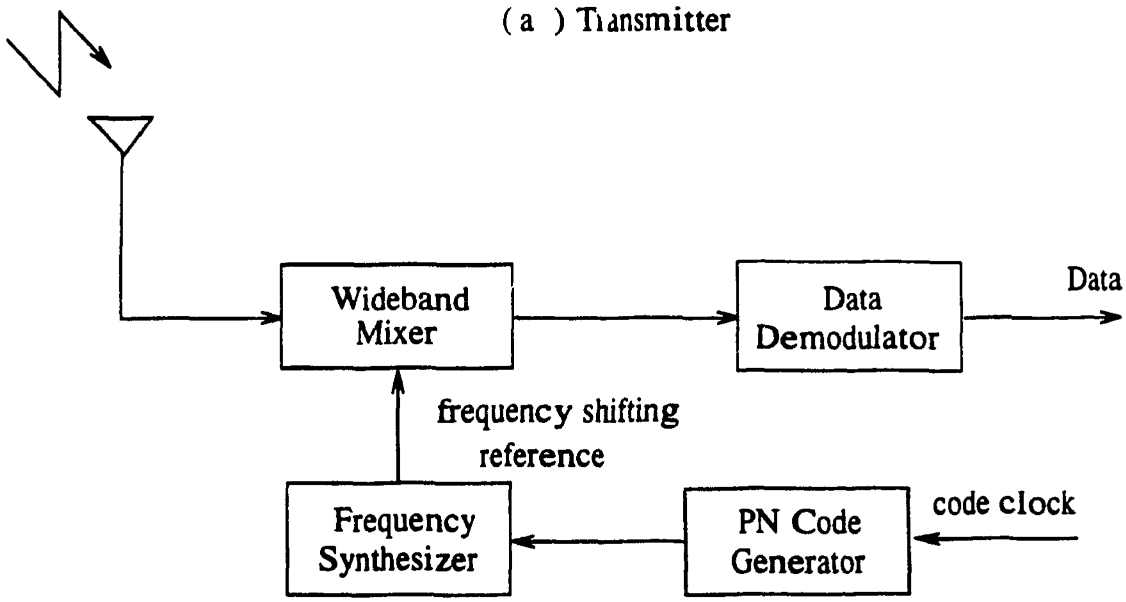
1.3 Frequency Hopping (FH)

Frequency hopping is the periodic changing of the frequency or frequency set associated with a transmission. FH systems are classfied into two groups, i.e. slow frequency hopping (SFH) and fast frequency hopping (FFH), by the relations between the data rate and the hopping rate. The hopped frequency and hopping rate are determined by the Pseudo Random (PN) sequence and code rate, respectively. The code sequences are the same type as those used by the direct sequence system, with the exception that the code clock rate is usually low (where direct sequence code rate are usually in the 1Mbits/s to 100Mbits/s range, frequency hopping codes do not normally exceed a few hundred kilobits per second)[40].

Figure 1.4 depicts the general form of a frequency hopping system. Spreading of the



(a) Tiansmitter



(b) Receiver

Figure 1.4: Block diagram of frequency hopping system

spectrum is obtained by changing the carrier frequency over the whole available band, according to a PN sequence. The bandwith over which the enegy is spread is essentially independent of the code clock rate and can be chosen by a combination of the number and size of frequency hops. In the receiver, the frequency hopping is removed by mixing (down-converting) with a local oscillatior signal which is hopping sychronously with the received signal.

As in direct sequence spread spectrum modulation, a meaningful parameter, with respect to interference spread uniformly across the RF band, is the processing gain, which for an FH system is

$$G = \frac{W_s}{W_d} \tag{1.5}$$

where W_s is the total FH bandwidth and W_d is the data bandwidth. When several frequency-hopping signals occupy a common W_s channel, simultaneous occupancy of a given frequency slot may result in message errors. This is the basic reason why error-correction coding is needed in a frequency-hopping system since the error-correction capability of the code can restore the messages that are destroyed by mutual interference.

Frequency-hopping system advantages:

- Create amount of spreading.
- Can be programmed to avoid portions of the spectrum.
- Relatively short acquisition time.
- Less affected by near-far problem.

Frequency-hopping system disadvantages:

- Complex frequency synthesizer.

- Not useful for range and range-rate measurement.

- Error-correction required.

1.4 Time Hopping (TH)

In time hopping systems spreading of the spectrum is achieved by compressing the information signal in the time domain. That is, the time hopping systems control their transmission time and period with a code sequence in the same way that frequency hoppers control their frequency. In fact, time hopping can be viewed as pulse modulation under code sequence control, and diagram for a time hopper can be generated by adding carrier-on-off control to the direct sequence diagram. A typical sample time hopping system is shown in Figure 1.5. A time-hopping waveform is given in Figure 1.6, where the time axis is divided into intervals known as frames, and each one of these frame is subdivided into M time slots. During each frame one and only one time slot will be modulated with a message by any reasonable modulation method. Time slot is chosen for a give frame is selected by means of a PN code generator.

Time hopping may be used to aid in reducing interference between systems in TDM. Interference among simultaneous users in a time-hopping system can be minimized by coordinating the time at which each user can transmit a signal, which also avoid the problem of very strong signals at a receiver swamping out the effects of weaker signals. In a non-coordinated system, overlapping transmission bursts will result in message errors, and for this it will normally require the use of error-correction coding to restore the proper message bits.

Time Hopping Advantages

- High bandwidth efficiency

- Implementation simplier than FH

Time Hopping Disadvantages

- Long acquisition time
- Error correction needed

1.5 Chirp System

Frequency-modulation pulse compression or chirp is a technique in which a carrier is swept linearly over a wide range of frequencies during a given pulse. This technique of spectrum spreading was developed a number of years ago to improve radar operation by obtaining the resolution of a short pulse, but with the detection capability of a long pulse. A long transmitted pulse is suitably modulated and the receiver is designed to act on the modulation to compress the pulse into a much shorter one. In this technique each transmitted signal element gives a change of frequency with time. Each received signal element is operated by a matched filter, which coherently combines the signal spectral components into a narrow signal of increased amplitude. The advantage of this technique for radar systems is that significant power reduction is possible. Also chirp waveform are less affected by Doppler shifts due to motion of the target relative to the radar and this is often significant when high velocity targets must be catched for. However, the chirp waveforms have the property that the apparent time of arrival shifts as a function of Doppler offset, because this chirp modulation is used in jam-resistant communication and pulse compression radars.

In Communication systems the receiver signal processing is similar to that used in chirp radar. The main difference is that in the communication system, both upward and downward frequency sweeps have to be detected, whereas in radar application only sweeps in one direction are used. Also, in communication SS systems Mark and Space data signals are defined by chirp signals having ascending and descending frequency sweeps, respectively, and are transmitted in the same frequency band. At the receiver, two matched

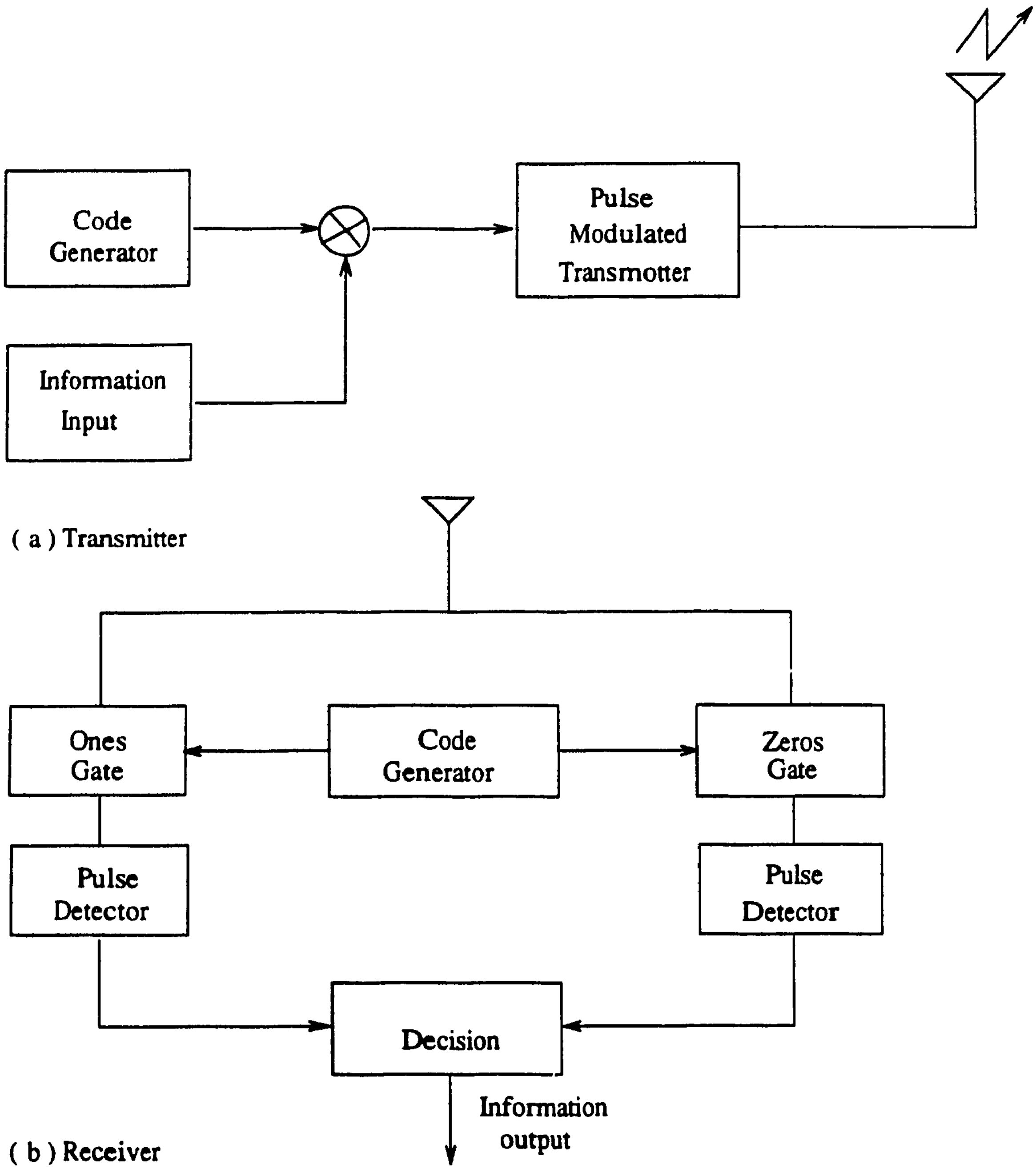


Figure 1.5: Simple time hopping system block diagram

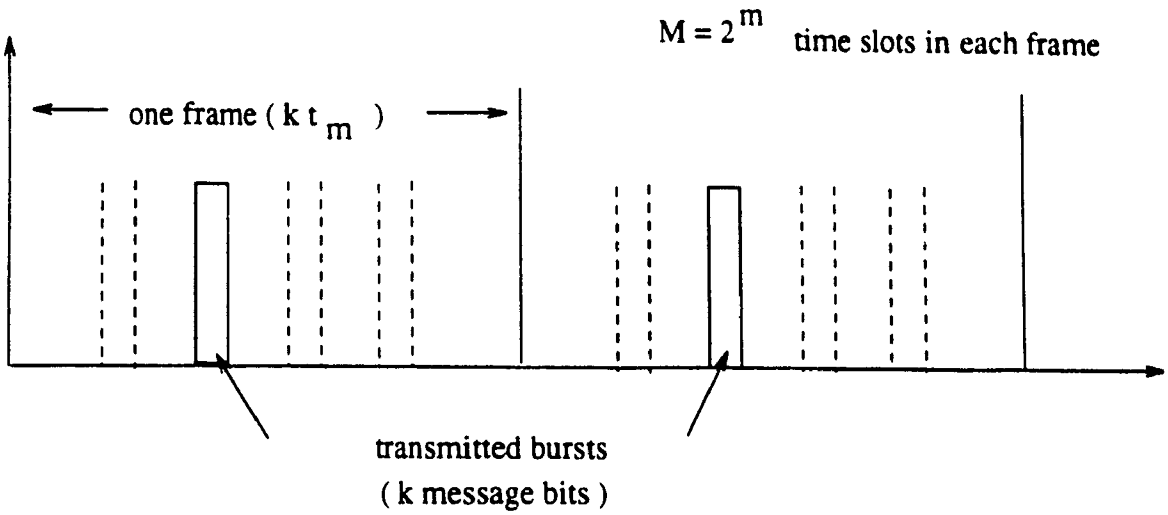


Figure 1.6: Time - hopping waveform

filters are used, one matched to the mark chirp signal and the other to the space chirp signal, where mark and space are used to represent " ones and zeros " in digital communication.

Chirp SS Systems advantages

- Signaficant power reduction is possible
- Coding is not normally used

1.6 Hybrid Spread - Spectrum Systems

Hybrid spread spectrum systems made up by combining two or more of direct-sequence, frequency-hopping, time-hopping, chirp, modulation technique, to offer certain advantage of a particular (method/technique) while avoiding the disadvantages, or to offer very wide-band and/or very high process gain, or to combine some of the advantages of two or three types of systems in a single system, and minimize the disadvantages of those types. Many different hybrid combination are possible. Some of these are:

DS/FH

DS/TH

DS/Chirp Signals

FH/TH

DS/FH/TH

Hybrid systems are widely used in military spread systems and they are now being employed in most of the newest systems at this time, as mobile communication services, indoor communication, ... , etc. Use of hybrid systems meets two basic needs: First, high performance which require wide spectrum expansions; by combining more techniques, wide spectrum expansion is achieved without stressing too much each single one. Second, is high flexibility that actual operational scenario, a hybrid system allows to adjust the mix of (evasion) and (resistancce) to optimize the (protection) against any defined (threat) and, inperspective, to keep pace with the evolution of the threat. Implementation in hybrid systems is not necessarily increased in difficulty by the same factor, however; that is, a combined frequency hopping and direct sequence (DS/FH) transmitter might be built up of readily a straightforward direct sequencce modulator or frequency hopper could not be construnted to do the same job.

1.7 Scope of the thesis

The contributions of this research can be summarized as follows:

Four basic types of spread spectrum techniques are introduced briefly. The rest of the thesis is organized as follows:

Chapter 2. presents the concept of CDMA system. The benefits and features of CDMA in digital cellular system is also given.

Chapter 3. analyse three traffic models for data sources, voice sources and video sources in ATM network, respectively. ATM cell format and two prime cogestion control schemes (reactive and preventive) are discussed in detail.

Chapter 4. discribes a performance analysis to compute the various performance criteria in a hybrid Time Division/ Asynchronous Transfer Mode/Code Division Multiply Access Network, i.e. TDMA/ATM/CDMA Network. Users accessing this TDMA/ATM/CDMA uplink frame are assumed to belong to one of four service classes, namely video, voice, file and interactive data. Each user accesses only a portion of the subframe slots assigned to its class. A variable frame boundary strategy is used to adjust the subframe boundaries depending on the call load. To alleviate congestion in the assumed hubless signalling free satellite network, the satellite measures the uplink traffic of each class and issues pilot congestion control indicators to on going calls of each class.

Chapter 5. summarizes the results of the research work and offers suggestions for further research.

Chapter 2

CDMA System Introduction

2.1 Introduction

Direct-Sequence Code Division Multiple Access (DS-CDMA) networks are gained momentum recently [1] due to such characteristics as multipath fading resistance, some inherent security, low power cost per station (compared, for example, to TDMA networks), inherent voice silence utilization, automatic cellular frequency reuse (in contrast to cellular FDMA networks)[2], etc. CDMA network are potential candidates for indoor communications, personal and consumer communications [2], satellite communications, factory automation, cellular communications, cordless Local Area Networks, etc. to name but a few.

On the performance evaluation side, the literature is rich with the signal-to-noise ratio comparisons between CDMA and non-CDMA networks, [3], with voice applications being the main motivation. On the data transmission side, some CDMA queueing models have been analyzed [4], either without evaluating the exact bit or packet error probabilities in the various fading environments or these probabilities were evaluated without attention to the networking aspects such as traffic characterization, buffering, etc. [5], [6].

Code Division Multiple Access Technique offer the following advantages as compared to their TDMA and FDMA counterparts:[38]

- Inherent security and selective counterparts:
- Fading and multipath resistance (Fact)
- Improved average power requirements (TDMA operates at peak power) (Fact)
- Better adaptation to varying voice and data traffic (TDMA may give similar effects)
- Automatic utilization of idle voice call periods (TDMA may give similar effects)
- Graceful degradation and replacement of the call blocking by the less severe packet loss (fact)
- Overlay on existing TDMA and FDMA with minimum interference and no radio licence requirement (Fact)
- Automatic frequency reuse in cellular systems thus improving the spectrum efficiency by many orders of magnitude. (Fact)
- Soft hand off (Yes, but some TDMA systems may yield similar effects)
- Provides useful hybrids, e.g. Direct Sequence/Slow Frequency Hopping (DS/SFH) TDM/SFH (Fact)
- Flexible, i.e. CDMA/TDMA modems could be designed to allow interoperation and migration of users. (Fact)
- Ease of reconfiguration as compared to cable based networks (Fact)
- Accurate ranging (Fact)

2.2 CDMA System Description

2.2.1 The CDMA Concept

CDMA is a modulation and multiple access scheme based on spread spectrum communication, a well-established technology that has been applied only recently to digital cellular radio communications and advanced wireless technologies. The approach will solve the near-term capacity concerns of major markets and the industry's long-term need for an economic, efficient, and truly portable communications.

Ever since the second pair of wireless telegraphs came into existence, we have been confronted with the problem of multiple access to the frequency spectrum without mutual interference. In the early days of wireless telegraphy, both frequency division in the form of resonant antennas, time division in the form of schedules, and netted oparations were employed. As the number of wireless radios in oparation increased and as the technology allowed, it became necessary to impose some discipline on the complex process we have today for world-wide frequency allocations and licensing by service type.

The multiple access problem can be thought of as a filtering problem. There are as many simultaneous users that want to use the same electromagnetic spectrum and there is a choice of an array of filtering and processing techniques which allow the different signals to be separately received and demodulated without excessive mutual interference. The techniques that have long been used include: propagation mode selection, spatial filtering with directive antennas, frequency filtering, and time sharing. Over the last 40 years, techniques involving spread spectrum modulation have evolved in which more complex waveforms and filtering processes are employed.

Propagation mode selection involves a proper choice of operating frequency and antenna so that signals propagate between the intended communicators but not between (very many) other communicators. Frequency reuse in cellular mobile telephone systems is an example of this technique carried to a great degree of sophistication.

Spatial filtering uses the properties of directive antenna arrays to maximize response in the direction of desired signals and to minimize response in the direction of interfering signals. The current analog cellular system uses sectorization to a good advantage to reduce interference from co-channel users in nearby cells.

With CDMA , each signal consists of a different pseudorandom binary sequence that modulates the carrier, spreading the spectrum of the waveform. A large number of CDMA signals share the same frequency spectrum. If CDMA is viewed in either the frequency or time domain, the multiple access signals appear to be on the top of each other. The signals are separated in the receivers by using a corralator which accepts only signal enegy from the selected binary sequence and despread in bandwidth and as a result, contribute only to the noise and represent a self-interference generated by the system.

The increased signal-to-noise ratio for the desired signal is shown in Figure 2.1. The signal-to-interference ratio is determined by the ratio of desired signal power to the sum of the power of all the other signals, and is enhanced by the system processing gain or the ratio of spread bandwidth to baseband data rate.

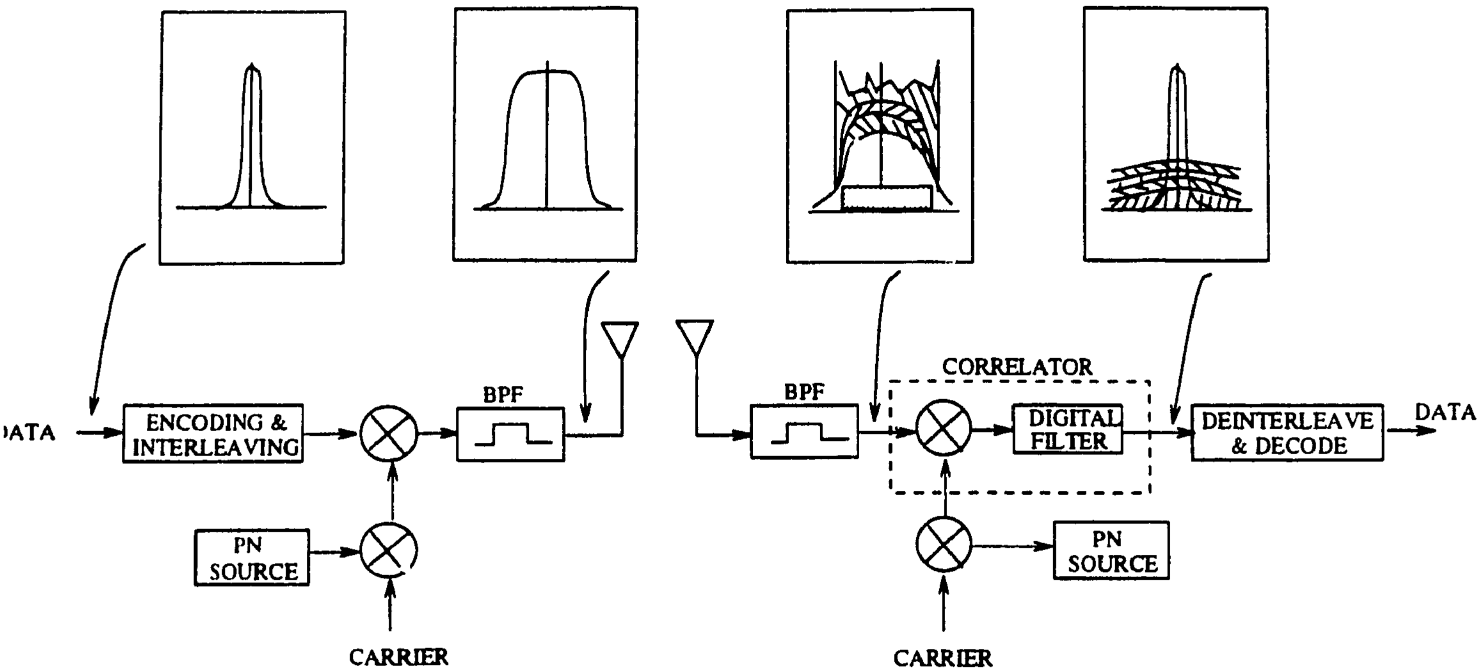


Figure 2.1: A view of the CDMA concept

2.2.2 Synchronous CDMA System Model

In Direct-Sequence CDMA (DS-CDMA) each user is assigned a specific code, called a "spreading code", used to spread the spectrum of the linearly modulated transmitted signal.

Both the source and spreading sequences are assumed to be bipolar: s_k's and b_k's are vectors with elements ± 1. The duration of one source symbol is T (also called the baud period), and the duration of one spreading sequence symbol (also called a chip) is T_c, $T = NT_c$. We also assume that each spreading signal $a_k(t)$, $1 \leq k \leq K$ and the carrier waveform g(t) are zero outside the interval [0, T]. Finally, we denote the power user k by E_k

Consider a synchronous CDMA system with K users and a set of pre-assigned signature waveform $s_k(t), k = 1, 2, ...K$. The common symbol duration is assumed to be T sec., where without loss of generality

$$\int_0^T s_k(t)^2 dt = 1, \tag{2.1}$$

The signature waveform is composed of a spreading sequene of L chips, i.e.,

$$s_k(t) = \sum_{m=1}^{L} a_m^k P_{Tc}[t - (m-1)T_c] \tag{2.2}$$

where $P_{Tc}(t)$ is the spreading pulse of duration $T_c = \frac{T}{L}$ and the chip symbols $a_m^k \in (-1, 1)$. Fig.2.2 shows the Synchronous CDMA System Model

The receiver signal in additive white Gaussian noise (AWGN) $w(t)$ is then given by (refer to 2)

$$r(t) = \sum_i \sum_{k=1}^{K} \sqrt{E_k} b_k(i) s_k(t - iT) + w(t) \tag{2.3}$$

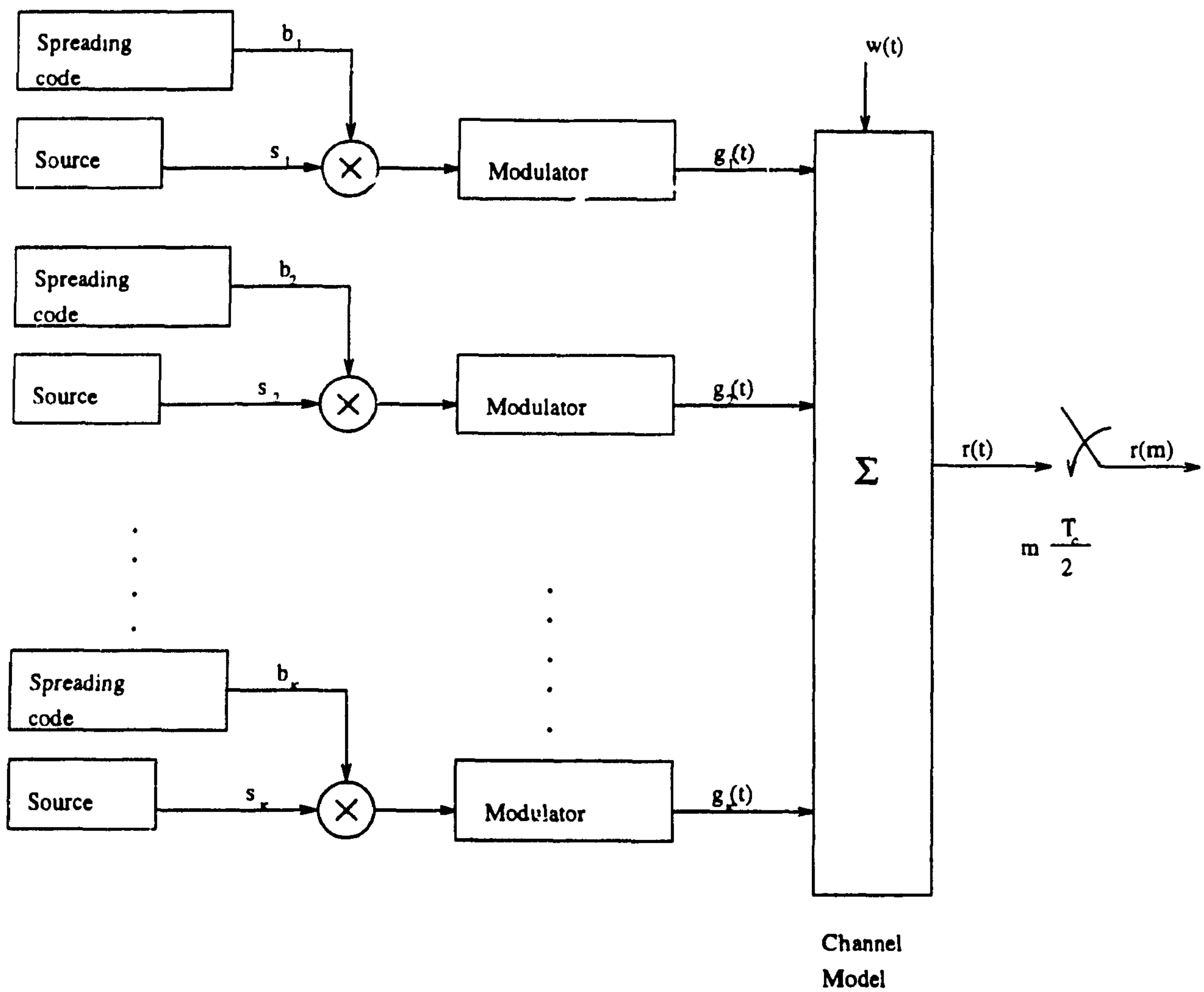


Figure 2.2: Synchronous CDMA system model

where $b_k i \in (-1, 1)$ represent the k^{th} user's bit during the i^{th} interval. Since we consider synchronous transmission, we may set $i = 0$ without loss of generality to obtain (dropping the superscript representing the bit interval for simplicity)

$$r(t) = \sum_{k=1}^{K} \sqrt{E_k} b_k s_k(t) + w(t), \tag{2.4}$$

E_k is the received energy of the k^{th} user's signal. The received signal is sampled twice per spreading chip at instant $t = m\frac{T_c}{2}$ to yield the discrete-time model

2.3 CDMA System Benefits

This section highlights some of the major attributes of the digital cellular system developed by QUALCOMM, that provide these benefits.

2.3.1 Multiple Forms of Diversity

In relatively narrowband modulation systems such as analog FM modulation employed by the first generation cellular phone system, the existence of multiple paths causes severe fading. With wideband CDMA modulations, however, the different paths may be independently received greatly reducing the severity of the multipath fading. Multipath fading is not completely eliminated because multipaths which cannot be independently processed by the demodulator occasionally occur. This will result in some fading behavior.

Diversity is the favored approach to mitigate fading. There are three major types of diversity: time, frequency, and space. Time diversity can best obtained by the use of interleaving and error correction coding. Wideband CDMA offers a form of frequency diversity by spreading the signal energy over a wide bandwidth. Space or path diversity is obtained three different ways by providing the following:

- Multiple signal paths through simultaneous links from the mobile station to two or more cell sites (soft handoff)
- Exploitation of the multipath environment through spread spectrum processing (rake receiver), allowing signals arriving with different propagation delays to be received separately and combined.
- Multiple antennas at the cell site.

2.3.2 High Capacity

In the frequency reuse concept, interference is accepted but controlled with the goal of increasing system capacity. CDMA does this effictively because it is inherently a better anti-interference (narrow) waveform than FDMA or TDMA. Indeed, its genesis was in military anti-jamming systems. Narrowband modulations are limited in frequency by the requirement to achieve a Carrier-to -Interference (C/I) ratio of about 18 dB. This requires that a channel used in one area is not reused in a nearby area. In CDMA, the wideband channel is reused in every area. [38]

In CDMA frequency reuse efficiency is determined by the signal-to-interference ratio that results from all system uses within range, instead of the users in any given area. Since the total capacity becomes quite large, the statistics of all users are more important than those of a single user. The "Law of large numbers" can be said to apply. This means that the net interference to any signal is the average of all the users' received power times the number of users. As long as the ratio of received signal power to the average noise power density is greater than a threshold value, the channel will provide an acceptable signal quality.

2.3.3 Low Transmit Power

Besides directly improving capacity, one of the more important results of reducing the required E_b/N_o (signal-to-interference level) is the reduction of transmitter power required to overcome noise and interference. This reduction means that station have reduced transmitter output requirements which reduces cost and allows lower power units to operate at larger ranges than the similarly powered analog or TDMA units. Furthermore, a reduced transmitter output requirement increases coverage and penetration and may also allow a reduction in cells required for coverage.

An even greater gain is the reduction of average (rather than peak) transmitted power that is realized because of the power control used in CDMA. Narrow band systems must always transmit with enough power to override the occasional fades. CDMA uses power control to provide only the power required at the time, and thus reduces the average power by transmitting at high levels only during fades.

2.3.4 Low E_b/N_o (or C/I) and Error Protection

E_b/N_o is the ratio of energy per bit to the noise power spectral density and is the standard figure-of-merit by which digital modulation and coding schemes are compared. It is directly analogous to the C/N (Carrier-to-Noise ratio) for analog FM modulation. Due to the wide channel bandwidth employed in the CDMA system, it is possible to use extremely powerful, low redundancy error correction codes must be used to conserve channel bandwidth. The CDMA system employs a powerful combination of forward error correction coding together with an extremely efficient digital demodulator in its implementation of the CDMA digital cellular system. The lower E_B/N_o increases capacity and decreases transmitter output power requirements.

2.4 CDMA System Features

The CDMA system has many other feature that do not relate directly to capacity, but rather to ease of system operation and to improved link quality. The following sections discuss three of these features: acquisition, mobile assisted handoff, and variable data rate.

2.4.1 System Pilot Acquisition

In the CDMA cellular telephone system, each cell site transmits a pilot carrier signal. This pilot carrier is used by the mobile station to obtain initial system synchronization and to provide robust time, frequency, and phase tracking of the signals from the cell site. This signal is tracked continuously by each mobile station. Variations in the transmitted power level of the pilot signal control the coverage area of the cell as is explained in the following section.

The pilot carriers are transmitted by each cell site using the same code but with different spread spectrum code phase offsets, which allow them to be distinguished. The fact that the pilots all use the same code allows the moble station to find system timing synchronization by a single search through all code phases. The strongest signal found corresponds to the code phase of the best cell site.

Each cell also transmits a setup or sync channel. This channel uses the same PN sequence and phase offset as the pilot channel and can be demodulated whenever the pilot channel is being tracked. This sync channel carries cell site identification, pilot transmit power, and the cell site pilot PN carrier System Time and knows the proper transmit power to initiate calls.

2.4.2 Mobile Station Assisted Soft Handoff

In a cellular telephone system, a handoff mechanism is provided to allow a call to continue when a mobile station crosses the boundary between two cells. In analog cellular telephone systems, the cell site receiver handing a call notice that the received signal strength from a mobile station has fallen below a predetermined threshold value. The cell site receiver then assumes that the mobile station must be near the cell border. When this occurs, the cell site asks the system controller to determine whether a neighboring cell site can receive the mobile station with a better signal strength. The system controller, in turn, sends message to the neighboring cell sites with the handoff request. These cell sites, employing special scanning receivers, look for the signal from the mobile station on the specified channel. Handoff is attempted if one of the neighboring cell site is selected and a control message to switch to this new channel is sent to commanding mobile station. At the same time, the system controller switches the call from the first cell site to the appropriate radio at the second cell site.

In the analog system, handoff can fail if an idle channel is not available to accept the call in the neighboring cell. It can also fail if another cell site reports hearing the mobile station in question when it actually hears a different mobile station using the same channel in a different cell. This results in the call being switched to the wrong cell. Also, the handoff fails if the mobile station fails to hear the command to switch channels. Actual operating experience indicates that handoffs fail frequently and that improvements are necessary. Another common problem is that when the mobile station is near the border between two cells, the signal levels tend to fluctuate at both cell sites and results in a ping-ponging effect. This overloads the system controller and increases the likelihood of a dropped call.

Since TDMA uses the same control structure as the analog system, it suffers from the same problems. In CDMA, soft handoff greatly reduces the link outage in the transition

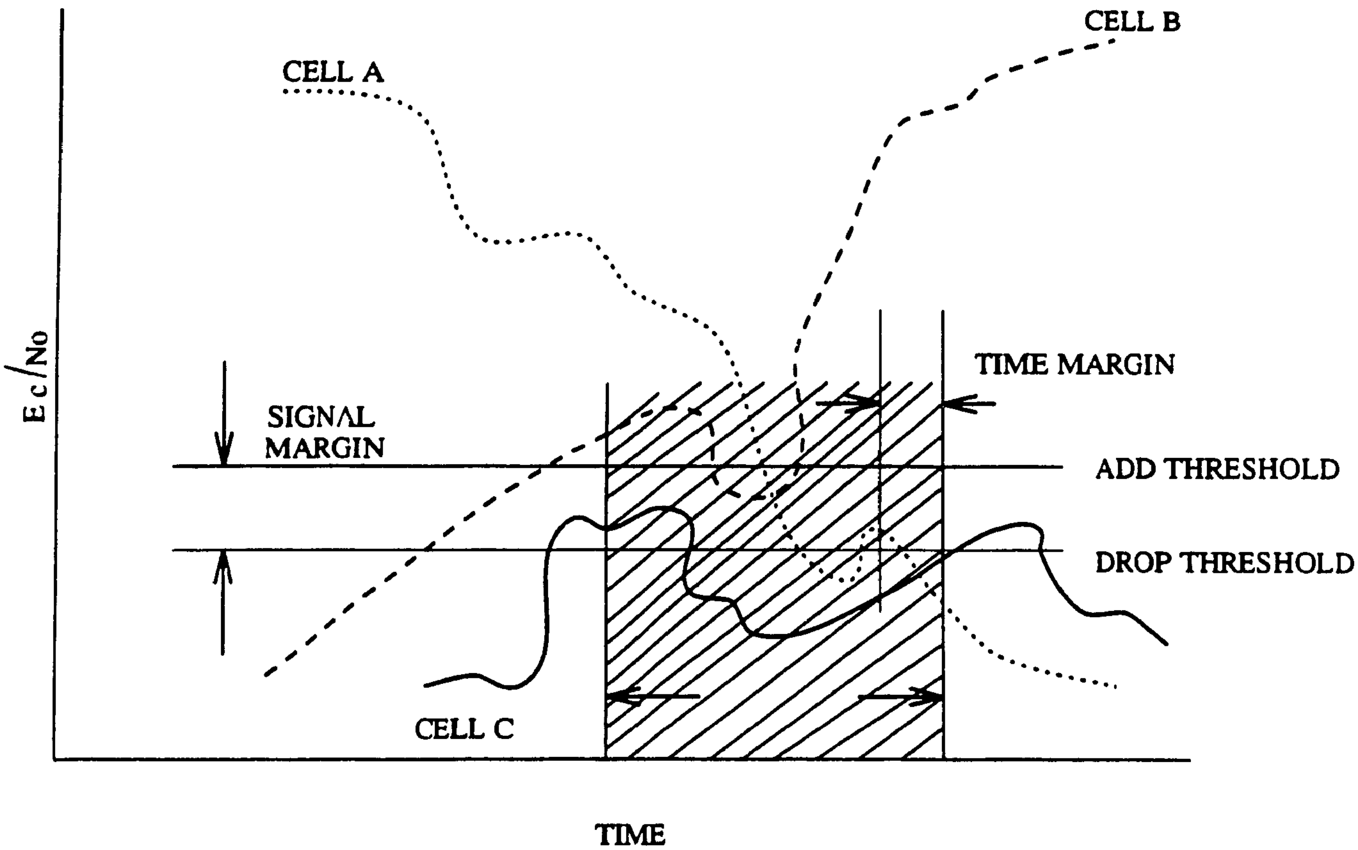


Figure 2.3: Mobile senses handoff requirements

region and in handoff using a technique that allows similtaneous transmission to and from the subscriber through two cells.

At call initiation, the subscriber is supplied a tailored set of handoff thresholds and a list of cells that are most likely to be candidatas for handoff. While tracking the signal from the original cell, the subscriber searches for all the possible pilots (with an emphasis on the candidate) and maintains a list of all pilots whose signals are above a threshold (as shown in figure 2.3) established in the initial setup. This list is transmitted to a center whenever it is requested, whenever the list changes by having a new pilot appear on the list, or when an existing pilot falls below a level that is useful to support the traffic.

Upon command from the center, via the initial cell, the subscriber unit commences tracking the second cell and uses diversity combining of the two signals (identical data is on each cell's transmission) to enhance the overall received signal. Power control information is received from both cells; both cells have to request a power increase for the subscriber to increase its power. Data from the subscriber unit received by both cells and is forwarded to the center where the best source is selected on a frame-by-frame (20 ms 0 basis and is used to represent the data transmitted by the subscriber.

This linkage can be terminated by returning to the original cell, dropping of the original cell, or initiation of tracking another cell prior to the completion of the handoff. The criteria for these actions are the E_c/N_o of the pilots; robustness is added to the process by requiring new pilots to exceed old ones by a selectable margin, and by requiring signals crossing thresholds to remain in the new zone for a certain length of time. The combined system of subscriber determination of E_c/N_o. The center commanded setup from subscriber identified cell information, continuous monitoring of the pilot signal strengths, and the level and time margins in the decisions, creates a very robust handoff process which field tests have shown to be very reliable.

2.4.3 Variable Data Rate

The appearence of the modulated signal does not change if the baseband data rate is changed. This allows considerable data rate flexibility to accommodate differing grades of voice quality which can be priced accordingly without impact on system operation or costly redesign of the system.

The cellular operator maintains a bank that operate at different rates, as a shared resource at the MTSO. A given vocoder's channel data rate is determined by a software parameter changeable from call to call. A wide range of new non-voice services (e.g., data, facsimile, imaging, ISDN, etc.) can be accommodates simultaneously.

Chapter 3

ATM Congestion Control

3.1 Introduction

ATM provides a non-hierarchial structure in which cells from different virtual connections (VCs) are multiplexed or switched using a common fabric, independent of the VC's bit rates or burstiness. VC's can be allocated resources based on deterministic multiplexing or statistical multiplexing. Since an ATM network will support mainly bursty or variable-bit rate sources, statistical multiplexing of sources would allow many VCs to implicitly share resources on the assumption that each VC only requires these resources for a small fraction of its time. Whenever statistical multiplexing is used there is a finite probability of congestion in the network. Congestion control are necessary to engineer the probability of occurrence of congestion and possibly to minimize the impact of congestion on the end users.

3.2 Modelling of Traffic Sources

ATM networks must support various communications services, such as data, voice and video, each having different traffic characteristics. To evaluate the performance of such networks, accurate source modeling is required. The purpose of this section is to examine several traffic models proposed for data, voice and video sources.

3.2.1 Input Traffic Models for Data Sources

It is well known that generation of data from a single data source is well charactered by a Poisson arrival process (continuous time case) or by a geometric interarrival process (discrete time case). For interactive data transmission, a single cell may be generated at a time. For a bulk data transmission, such as a file transfer, a large number of cells may be generated at a time (batch arrivals).

In existing packet networks, packets could be either of variable or constant length. In ATM networks, however, the cell size is fixed. Furthermore, because the size of a cell is relative short compared to the length of a packet in existing networks, multiple cells may be created from one data packet.

3.2.2 Input Traffic Models for Voice Sources

An arrival process of cells from voice source (and a video source) is fairly complex due to the strong correlation among arrivals. In this subsection, input traffic models proposed for a voice source are examined.

The arrival process of new voice calls and the distribution of their durations can be characterized by a Poisson process and by an exponential distribution, respectively. Whithin a call, talkspurts and silent periods alternate. During talkspurts, voice cells are generated periodically; during silent periods, no cells are generated. The generation of correlated voice cells within a call can be modeled by an Interrupted Poisson Process (IPP) [9]-[13]. In an IPP model, each voice source is characterized by ON (corresponding to talkpurt) and OFF (corresponding to silence duration) periods , which appear in turn. The transition from ON to OFF occurs with the probability β, and transition

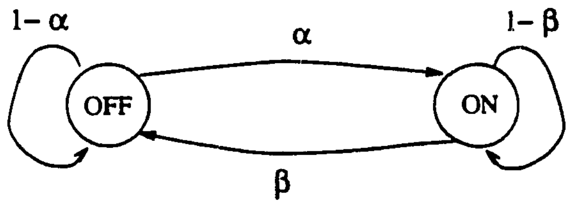


Figure 3.1: IPP model

from ON to OFF occurs with the probability α. In a discrete time case, ON and OFF periods are geometrically distributed with the mean $1/\beta$ and $1/\alpha$, respectively. Cells are generated during the ON period according to a Bernoulli distribution with the rate λ; no cell is generated during the OFF period (FIG. 3.1) (The continuous time analog is an exponential distribution using a Poisson process.)

When N independent voice sources are multiplexed, aggregated cell arrivals are governed by the number of voice sources in the ON state. Assuming a discrete time system, the probability P_n that n out of N voice sources are in the ON state (n voice cell arrivals in a slot) is given by

$$P_n = \binom{N}{n} \left(\frac{\alpha}{\alpha+\beta}\right)^n \left(\frac{\beta}{\alpha+\beta}\right)^{N-n}, \qquad for \quad 0 \leq n \leq N \tag{3.1}$$

The continuous time analog represent the number of voice sources in the ON state as a birth-death process with birth rate $\lambda(n)$ and death rate $\mu(n)$, where

$$\lambda(n) = (N-n)\alpha, \qquad \mu(n) = n, \qquad for \quad 0 \leq n \leq N \tag{3.2}$$

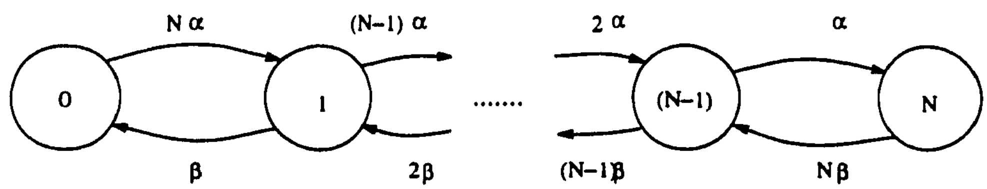


Figure 3.2: Birth-death model for the number of active sources

Figure 3.2 shows the birth-death model. For the continuous time case, the probability P_n that n out of N voice sources are in the ON state is also given by (3.1) [11].

Another common approach for modeling aggregate arrivals from N voice sources is to use a two-state Markov Modulated Poisson Process (MMPP) [14], [15]. The MMPP is a doubly stochastic Poisson process where the rate process is determined by the state of a continuous-time Markov chain [14]. In the two-state MMPP model, an aggragate arrival process is characterized by two alternating states. It is usually assumed that the duration of each state follow a Bernoulli (or a Poisson) distribution with different rate in each state. Note that an IPP, a process used to describe a single voice source, is a special case of the MMPP in which no cell arrives during an OFF period.

To determine the value of these four parameters, the following MMPP statistical characteristics are matched with the measured data[14]:

1. the mean arrival rate;

2. the variance-to-mean ratio of the number of arrivals in a time interval (0, t_1);

3. the long term variance-to-mean ratio of the number of arrivals;

4. the third moment of number of arrivals in (0, t_2).

Note that the analytical models described in Section 3.2.1 and 3.2.2 can model only constant bit rate traffic. Analytical models which can adequately model variable bit rate traffic are not yet available.

3.2.3 Input Traffic Models for Video Sources

Video traffic requires large bandwidth. For instance, in TV applications a frame of 512x512 resolution is transmitted every 1/30 second, generating 512x512x8x30 bits per second (approximately 63 Mbits/s), if a simple PCM coding scheme is used. Therefore, video sources are usually compress by using an interframe variable-rate coding scheme which encodes only significant differences between successive frames. This introduces a strong correlation among cell arrivals from successive frames.

Like a voice source, a video source generates correlated cell arrivals; however, its statistical nature is quite different from a voice source. Two types of correlations are evident in the call generation process of a video source; short-term corresponds to uniform activity levels (i.e., small fluctuations in bit rates), and its effects last for a very short period of time (on the order of a few seconds)[16]. In Section 3.2.3-1), models which consider only short-term correalation (i.e., models for video sources without scene changes) are examined. In Section 3.2.3-2), models which consider both short-term and long-term correlation (i.e., models for video sources with scene changes) are examined.

1) Models for Video Sources Without Scene Changes: These models are applicable to

video scenes with relatively uniform activity levels such as videotelephone scenes showing a person talking. Two models have been proposed. The first model approximates a video source by an autoregressive (AR) process. This model describes the cell generation process of a video source quite accurately. However, because of its complexity, queueing analysis based on this model is very complicated and may not be tractable in queueing analysis than the first model, and yet describe the cell generation process of a video source (or video sources) well.

2) Models for Video Sources with Scene Changes: These models capture both short-term and logn-term correlations explained at the beginning of Section 3.2.3 and thus these models are suitable to describe a cell generation process from video scenes with sudden changes, such as videotelephone scenes showing changes between listener and talker modes, or scene changes in broadcast TV [16]. Two models have been proposed: the first model is an extension of Model B explained above; the second model approximates a video source by this discrete state continuous-time Markov process with batch arrivals.

3.3 ATM Cell Format

In the CCITT Recommendation I.121, a guideline for future B-ISDN standardization, ATM has been accepted as the final transfer mode for B-ISDN[36]. According to this recommendation, information flow in ATM is organized into fixed-size cells, each consisting of a header and an information field. Fixed-size cells are chosen over variable-size units because, based on the state of the existing experimental fast packet switching technology, it is believed that fixed-size cells can be switched more efficiently[37]. Theses cells are transmitted over a virtual circuit and cells belonging to same virtual circuit are identified by the header. ATM is , by definition, a connection-oriented technique. This connection-oriented mode minimizes delay variation since cells belonging to the same call follow the same route. It also minimizes the processing required to make routing decisions.

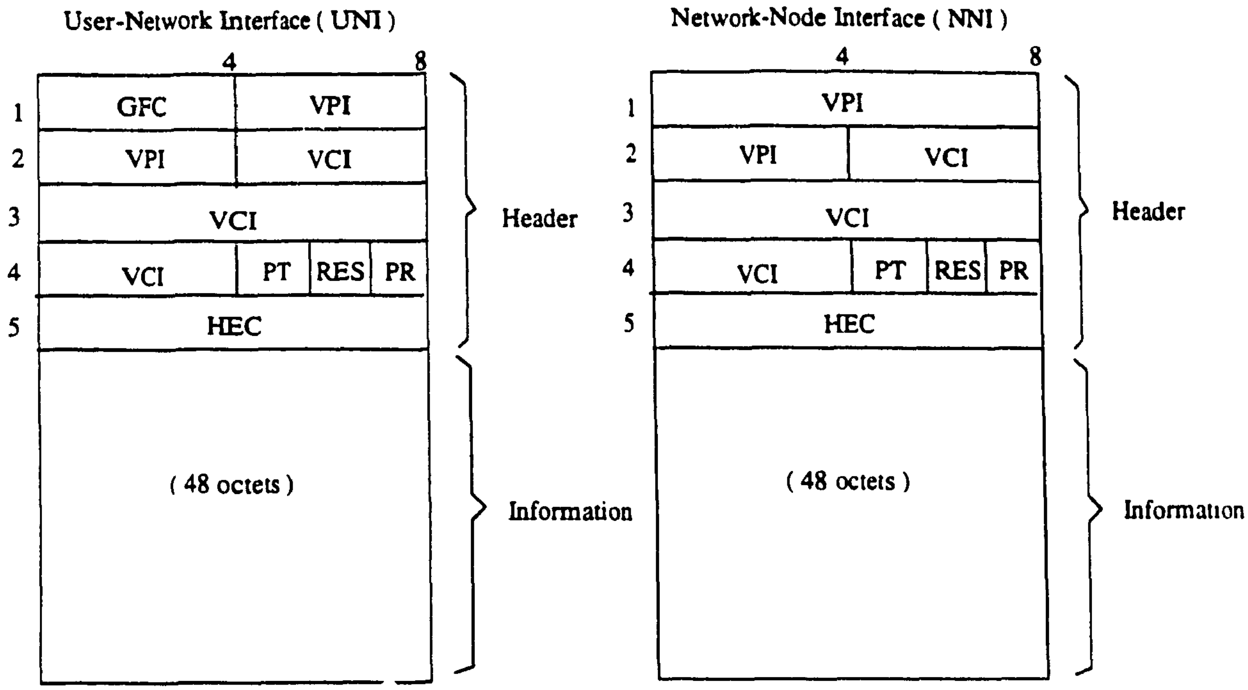


Figure 3.3: ATM cell structure

The size of an ATM cell should be small in order to reduce the degrading effect of the packetization delay at the source. For instance, considerable delay could be introduced during creation of a voice cell if the size of a cell is large. The CCITT header formats which will be used at User-Network Interface (UNI) and Network-Node Interface (NNI) are shown in Figure 3.3. For UNI, the header contains a 4-bit " generic flow control" (GFC) field, a 24-bit label field containing Virtual Path Identifier (VPI) and Virtual Circuit Identifier (VCI) subfields (8 bits for the VPI and 16 bits for the VCI), a 2-bit payload type (PT)field, a 1-bit reserved field, a 1-bit priority (PR) field, and an 8-bit header error check (HEC) field. For NNI, the header does not contain a GFC field, and the extra 4 bits are used for a VPI field.

The GFC field is used to assist the customer premises in controlling the flow of traffic for different qualities of service. The exact procedures for how to use this field are not agreed upon as yet. One candidate for the use of this field is a multiple priority level indicator to control the flow of information in a service dependent manner. The GFC field appears only at the UNI.

The " Virtual Path " concept is adopted in a label field. The VPI provides an explicit path identification for a cell, while VCI provides an explicit circuit identification for a cell. Basically, a virtual path is a bundle of virtual circuits which is switched as a unit by defining one additional layer of multiplexing on a per-cell basis underneath the VCI. A predefined route is provided with each virtual path; thus it is not necessary to rewrite the routing table at call setup. Therefore, call-by-call processing at switching nodes is reduced and call set-up delay is decreased. Although the transmission efficiency may decrease because of the label overhead, this effect is negligible since large bandwidth will be available as hugh capacity optical fibers become more widely used. The PT field can be used for maintenance purposes, and it indicates whether the call contains user information or network maintenance information. This field allows for the insertion of cells on to a virtual channel without impacting the user's data.

The PR field indicates cell loss priority and is used to selectively discard cells when congestion occurs. One possible implementation is to set this field to zero for the cells which are droppable. When congestion occurs, cell whose PR field set to one are dropped first.

3.4 Congestion Control Schemes

3.4.1 Introduction

In an ATM network, most traffic sources are bursty. A bursty source may generate cells at a near-peak rate for a very short period of time and immediately afterwards it may become inactive, generating no cells. Such a bursty traffic source will not require continuous allocation of bandwidth at its peak rate. Since an ATM network supports a large number of such bursty traffic sources, statistical multiplexing can be used to gain bandwidth efficiency, allowing more traffic sources to share the bandwidth. But it should be notice that if a large number of traffic sources become active simultaneously, severe network congestion can result.

Due to the effects of high-speed channels, congestion control is a challenge for an ATM network. High-speed channels significantly limit the congestion control schemes applicable. As an example, consider two adjacent switching nodes, A and B , linked by a 100-km cable. Assume 500-bit long cells and the typical propagation delay time of 5 μs per 1 km of a cable. (The exact cell size is 424 bits (53 octets) according to the CCITT standards . However, the cell size of 500 bits is used in this example for simplicity.) Consider the following scenario. Assume a 1 Mbits/s channel speed. One cell transmission time becomes (500 bits)/(1 Mbits/s) =0.5 ms. Node A starts transmitting a cell. It takes 500 μsec for the electric signal to propagation to node B. Thus when the first bit of the cell reaches B, A is transmitting the last bit of the same cell. Let's replace the channel with a Gbits/s fiber optic cable. The cell transmission time reduces to (500 bits)/(1 Gbits) = 0.5 μsec, while the propagation delay time remains the same. Again, A starts transmitting a packet. This time, when the first bit of the packet arrives at B, A is transmitting the 1000th cell. 1000 cells are already on the channel propagating towards B. This example shows that in high-speed networds such as ATM networks, overhead due to propagation delay time becomes significant. Thus control schemes, such as those which adjust A's input rate based on feedback from B, may not work in ATM networks. As clearly shown in this example, having high-speed channels changes the network situation dramatically; thus some of the congestion schemes developed for existing networks may no longer be applicable in such high-speed networks.

Another factor which makes congestion control in ATM challenging is the simplicity of the protocols used in high-speed networks. Simple, possibly hardwired protocols are preferred in ATM networks in order to match the high speed of the network channels. As shown in the above example, replacing a Mbits/s from 0.5 ms to 0.5 μsec. On the other hand, the time required to process a protocol remains the same. As a result, in a high-speed network environment protocol processing time can be a bottleneck. In order to avoid such a bottleneck, ATM networks use simplified protocols, pushing most of the link-by-link layer protocols to higher edge-to edge layers. This makes it difficult to implement link-by-link congestion control schemes.

For these reasons, many of the congestion schemes developed for exisiting networks may not be applicable to ATM networks. Many of the congestion control schemes developed for existing networks fall in the class of reactive control. Reactive control reacts to congestion after it happens and tries to bring the degree of network congestion to an acceptable level. However, reactive control is not suitable for use in ATM networks.

A new concept is therefore required for congestion control in an ATM environment. Various congestion control approaches have been proposed for ATM network, most of which fall in the class of preventive control. Preventive control tries to prevent congestion before it happens. The objective of preventive control is to ensure a priori that network traffic will not reach the level which causes unacceptable congestion.

Therefore, congestion control schemes are required to ensure that sufficient bandwidth and satisfactory performance are provides to all accepted VCs at the same time remaining operationally simple, easy to implement and robust to evolving ATM traffic characteristics.

Proposed CCITT recommendation I.311 indicacates the following four objectives for ATM layer congestion control:

- ATM-layer congestion controls should not rely on ATM adaptation layer or higher layer protocols.

- ATM-layer congestion control should be optimal considering the tradeoff between system complexity and network efficiency.
- ATM-layer congestion controls should support a set of ATM-layer QOS classes sufficient for all foreseeable BISDN services.
- ATM-Layer congestion controls should attempt to maintain the VC's QOS even under congestion conditions.

3.4.2 Reactive Control Schemes

At the onset of congestion reactive control instructs the source nodes to throttle their traffic flow by giving feedback to them. A major problem with reactive control in high-speed netwotks is slow feedback. There is a possible improvement technique to overcome the difficulty caused by slow feedback. If reactive control is performed between network users and the edge of the network , the effect of propagation delay may not be significant since the distance feedback information propagates is short. However, this limits the reactive control to the edge of the network.

Reactive flow control, in general, may not be effective in an ATM environment because of the previously discussed problem. Preventive control, however, tries to overcome this problem with reactive control and controls congestion more effectively in ATM networks.

3.4.3 Preventive Control Schemes

Unilike reactive control where control is invoked upon the detection of congestion, preventive control does not wait until congestion actually occurs, but rather tries to prevent the network from reading an unacceptable level of congestion. The most common and effective approach is to control traffic flow at entry point to the network (i.e., at the access nodes). This approach is especially effective in ATM networks because of its connection-oriented transport. With connection-oriented transport, a decision to admit new traffic can be made based on knowledge of the state of the route which the traffic would follow.

Preventive control for ATM can be performed in two ways: admission control and bandwidth enforcement. Admission control determines whether to accept or reject a new connection at the time of call setup. This decision is based on traffic characteristics of the new connection and the current network load. The bandwidth enforcement monitors individual connections to ensure that the actual traffic flow conforms with that report at call establishment.

Admission control: Admission control decides whether to accept or reject a new connection based on whether the required performance can be maintained. When a new connection is requested, the network examines its service requirements (e.g. acceptable cell transmission delay and loss probability) and traffic characteristics (e.g. peak rate, average rate, etc.) The network then examines the current load and decides whether or not to accept the new connection.

Three major research issues in admission control are listed as follows.

- What traffic parameters (traffic descriptors) are required to accurately predict network performance?
- What criteria should the network use to decide whether or not to accept a new connection?
- How does network performance depend on various traffic paremeters ?

Bandwidth enforcement: since users may deliberately exceed the traffic volume de-

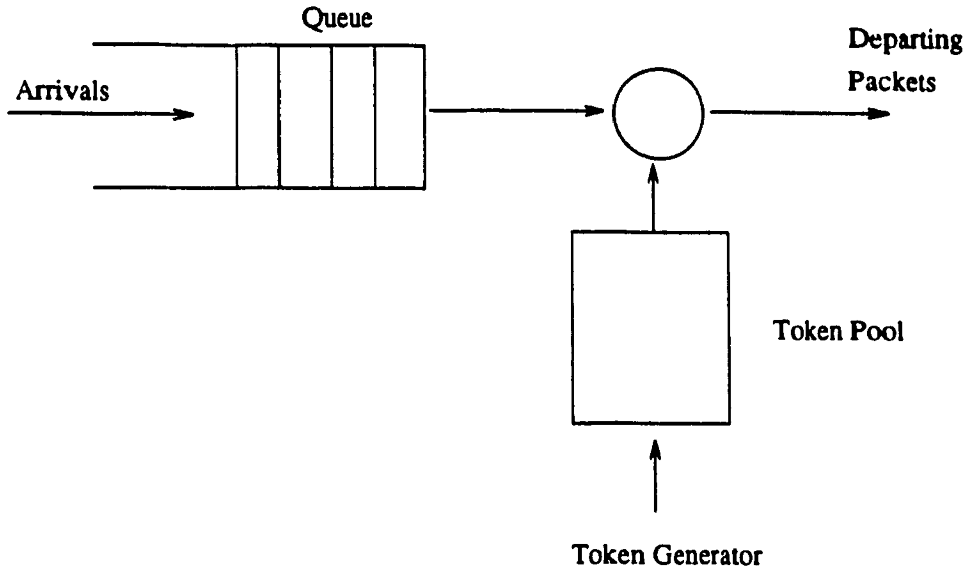


Figure 3.4: A queueing model for a leaky bucket method

clared at the call setup (i.e., values of their traffic descriptors), and thus easily overload the network, admission control alone is not sufficient. After a connection is accepted, traffic flow of the connection must be monitored to ensure that the actual traffic flow conforms with that specified at call establishmented at the edges of the network. Once a violation is detected, the traffic flow is enforced by discarding and /or buffering violating cells.

A Leaky Bucket method [20]-[24] is one of the typical bancwidth enforcement mechanisms used for ATM networks; this method cau enforce the average bandwidth and the burst factor of a traffic source. One possible implementation of a Leaky Bucket method is to control the traffic flow by means of tokens. A queueing model for the Leaky Bucket method is illustrated in Figure 3.4.

An arriving cell first enters a queue. If the queue is full, cells are simply discarded. To enter the network, a cell must first obtain a token from a token-pool; if there is no token, a cell must wait in the queue until a new token is generated. Tokens are generated at a fixed rate corresponding to the average rate of the connection. If the number of tokens in the token poll exceeds some predefined thershold value, the process of token generation stops. This threshold value corresponds to the burstiness of the transmission; the larger the threshold value, the bigger the burstiness. This method enforces the average input rate while allowing for a certain degree of burstiness. The Leaky Bucket method can also enforce the peak bandwidth by generating tokens at the rate corresponding to the peak rate.

In the original Leaky Bucket methed proposed in[34], the input buffer is not provided. In [36], the input buffer is suggested to provide better control of the trade-off between the cell waiting times and the cell loss probabilities. In an extreme case where no input buffer is provided, incoming cells do not have to wait in the buffer, but a large number of cells may be lost since all the violating cells are discarded. In the other extreme case (where an infinite input buffer is provied), no incoming cell will be lost, but cells may suffer a long waiting tme. By choosing an approprate input queue size, the trade-off between these two extremes can be controlled.

In the Leaky Bucket method, violating cells are either discarded or stored in a buffer even when the network load is light, and thus network resources are wasted. The total network throughput can be improved by using the marking method presented in [18], [19], [25]. In this scheme, violating cells, rather than being discarded, are permitted to enter the network with violation tags in their cell headers. These violating cells are discarded only when they arrive at a congested node. If there are no congested nodes along the routes, the violating cells are transmitted without being discarded. This marking method can easily be implemented using the Leaky Bucket method described above. When the queue length exceeds a threshold, cells are marked as 'droppable' instead of being discarded. One possibe disadvantage of this marking scheme is that processing time in each node is increased slightly because each node has to distinguish tagged cells from nonviolating cells when the node is in a congested state. Each node must also monitor its state to determine if it is in congestion. (For instance, each node may check its queue length ot detect the congested state. However, this extra processing can be done quickly and easily and the overall merites of the marking method far exceed its slight disadvantages.

Chapter 4

Congestion Control In Signalling Free Hybrid ATM/CDMA Satellite Network

4.1 Introduction

The objective of this research is to develop and analyze signalling techniques that will minimize the access time to the satellite during call setup as well as on board implementation costs by simplifying the overall access scheme. Conventionally in TDMA based systems the satellite isochronous communication user (voice or video) requests, at call setup time, a slot allocation for the duration of the call. It is only after allocating the slot (or slots) that the user starts transmitting information. This delay is substantial and is equal to the time the user sends the request for slot allocation to the satellite (one way propagation delay) plus the satellite processing and servicing time to other users as well as responding to this request (a second propagation delay) which could be prohibitive for data applications. In the proposed call setup scheme, the user does not have to waste any time waiting for a slot; rather it utilizes randomly selected slot from within a range of slots called a subframe according to a specific algorithm that is developed in [30] In this new approach, the satellite mainly acts as a traffic measuring, and congestion and policing control entity for the new accepted calls . Complicated slot assignment and on board management schemes common with TDMA systems are no longer necessary. The satellite and ground users firmware that was in place to track and manage these slots are now saved . Moreover the frame capacity that was used to inform users about their assigned slot is not relevant.

The new uplink or downlink frames do not have reservation and information field as found in conventional techniques. In this new approach, only the information field is required. Since various classes of service are typically accommodated, the frame is subdivided into subframes one for each class. The size and boundary of each subframe is defined by the satellite and both measures are updated cyclically by the satellite based on each class usage of its assigned bandwith. An algorithm will be devised for the refreshing of the classes subframes boundaries by considering the congestion that may occur because of higher demands by one class and lower demands by another class.

The frame is apportioned among the classes according to moving boundary TDMA protocol[30]. However users of each class access the coresponding subframe according to a Code Division Multi Accessing scheme (CDMA scheme). Each subframe has a number of slots, and users randomly select the slots (within their subframe) they wish to use for their transmission. Different users may overlap in the same slot in a particular subframe but if the number of users is within the spread spectrum interference that guarentees an acceptable bit error performance, users transmission quality will not be much affected; however if error occurs, users could change randomly the slot location in the subframes and/or use sophisticated forward error correcting codes.

4.2 Description Of The New CDMA/TDMA Satellite Network

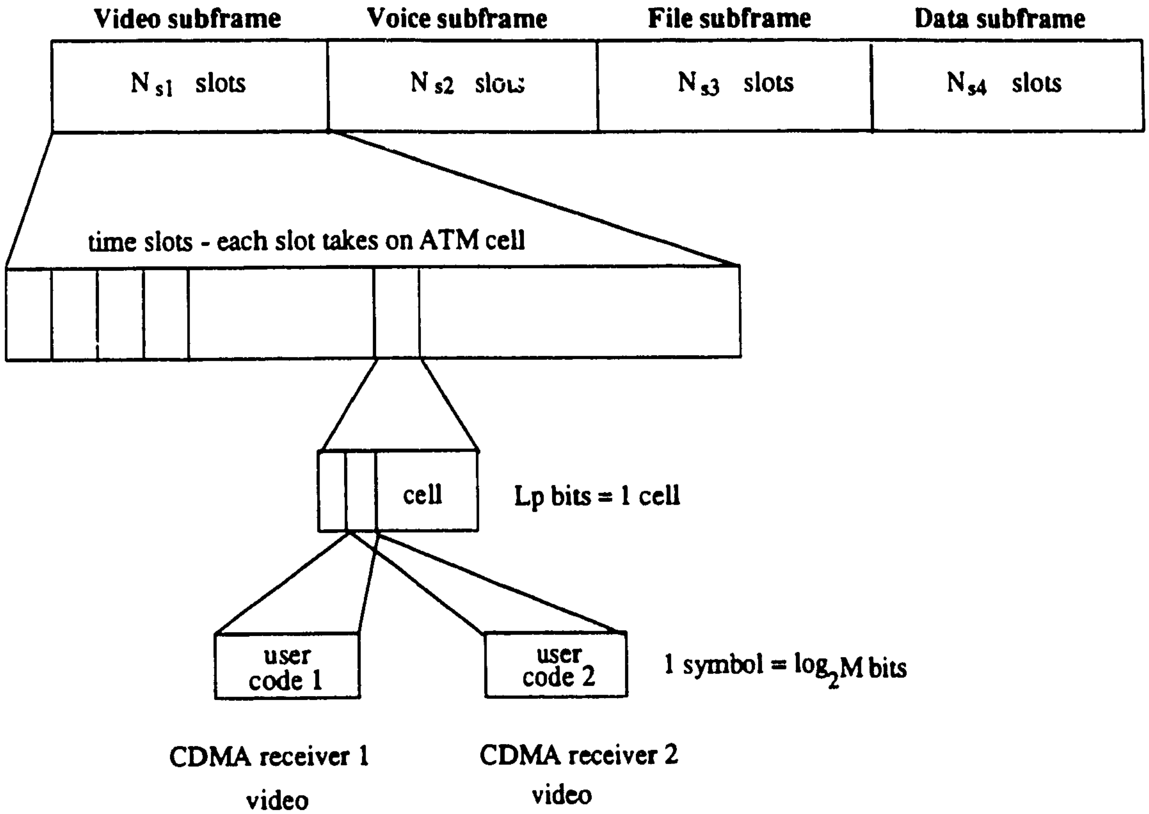
As outlined previously users of the 4 classes of services (video, voice, file and interactive users) access their individual subframes (Fig 4.1.); but users of the group overlap within the same subframe of length N_{si} slots. For all four classes, call connection control and overhead is minimal. All that new calls have to do is to monitor the so called congestion indicators within the satellite downlink frame. If congestion conditions permit, then users start transmitting their CDMA signals within their subframe. Transmission of various user signals may overlap and that is the reason for selecting a CDMA technique.

We assumed constant bit rate for the all four classes, and

A specific subclass of CDMA system is the one presented by Enge [27] and used here. In this system (Fig 4.2) the k^{th} user employs a long code and a set of M orthogonal spreading code, i.e. { V_1^k, V_2^k, ..., V_M^k } where k=1, 2, ..., K_i', where K_i' is the maximum allowable number of users in class i (i=1, 2, 3, 4). During a symbol interval $T_s = (log_2 M)T_b$(a bit interval), the user groups each $(log_2 M)$ information bits and selects one of the M complex orthogonal spread spectrum waveforms above accordingly, where V_m^k={ $V_{m,1}^k$, $V_{m,2}^k$, ..., $V_{m,N-1}^k$ } where N is the length of the short direct sequence complex code signal that is transmitted based on the $(log_2 M)$ information bits. Each of the $V_{m,i}^k$ components can be expressed as e^z, (where $z = j\theta_{mr}^k$) which is the r^{th} root of unity (in a manner similar to MPSK modulation).

As an example if M=2 (the binary case) and N=4, the orthegonal codes assigned to user 3 for example are given by $V_1^3 = \{1\ 1 - 1 - 1\}$, $V_2^3 = \{-1 - 1\ 1\ 1\}$, which are orthogonal over a period $T_s = 1 * T_b = T_b$.

M-ary equally likely data symbols are transmitted at a rate of one every T_s seconds. The signal transmitted by the k^{th} user to send the m^{th} symbols during the interval $[0, T_s]$



Showing an example where 2 users overlap in the same CDMA slot

Figure 4.1: Frame structure

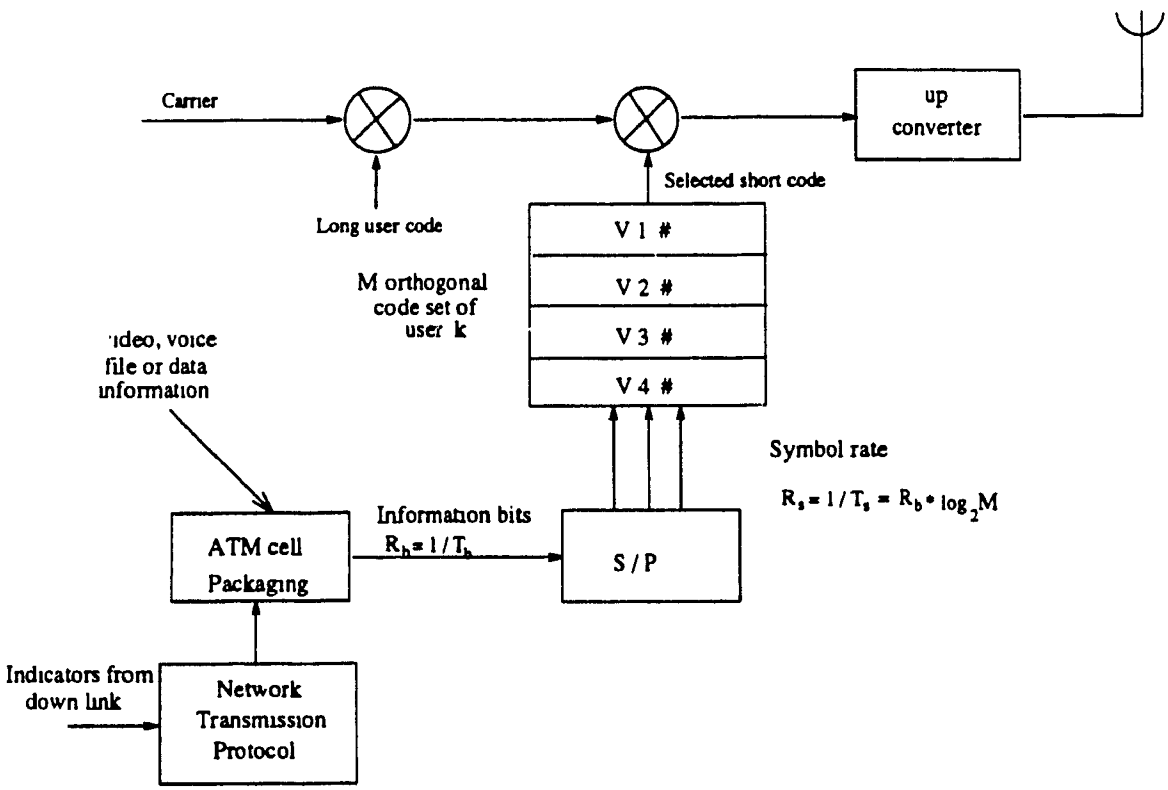


Figure 4.2: Generic transmitter of the orthogonal CDMA system used by all users of four classses

is [1]

$$S(V_m^k, t) = Re\left[\sqrt{2P_s}\sum_{n=0}^{N_1}[V_{m,n}^k]^*\Gamma(t - nT_c)exp(j\omega_c t + \theta_k)\right] \qquad (4.1)$$

where P_s is the average signal power. The chip duration T_c is T_s/N, $\Gamma(t)$ is the chip waveform, θ_k the carrier phase, ω_c is the carrier frequency common to all users and $\omega_c T_c = 2\pi n$, n an integer. The chip waveform $\Gamma(t - nT)$ defined for $0 \leq t < T_c$, is zero outside the range, and is normalized so that the energy per chip is equal to T_c. Therefore, the energy per symbol is $E_s = P_s T_s$ and the energy per bit is $E_b = E_s/log_2 M$ since $log_2 M$ bits are conveyed by each symbol.

At any time slot (whose length is packet length $= T_p$ sec) the signal received at the satellite (or mobile base station BS) is the sum of the possible k_i' received signals of the users that generated packets at that time slot (Fig.3). As will follow shortly we assume K_i' to be the number of active calls of class i, k_i' is the number of active bursts generated by all of the K_i' calls and m the actual number of packets (alternately cells in ATM environments) generated and overlapping in certain slot from the active K_i' calls.

Fig 4.3 shows how the receiver signal is obtained at a certain time slot (cell time) in the satellite uplink. While figure 4.4 shows the optimal matched filter receiver for this orthogonal set detection. Following the despreading by the long user code and at the end of each $(T_b * log_2 M) = T_s$ seconds $= NT_c$ all matched filters (MF) are sampled and binary expansion of the identity of the MF that peaked yields the data bits values (e.g. MF #5 means 101 detected as data bits). Repeating the process yields all bits of each cell. Needless to say that the satellite or mobile base station will have to have K_i' CDMA receivers like the one in figure 4 to be able to track and detect simultaneously and in parallel the maximum of k_i' cells possible from K_i' calls in a certain time slot (as in Figure

[1]This excludes the long code multiplicative term which is decorrelated at the receiver

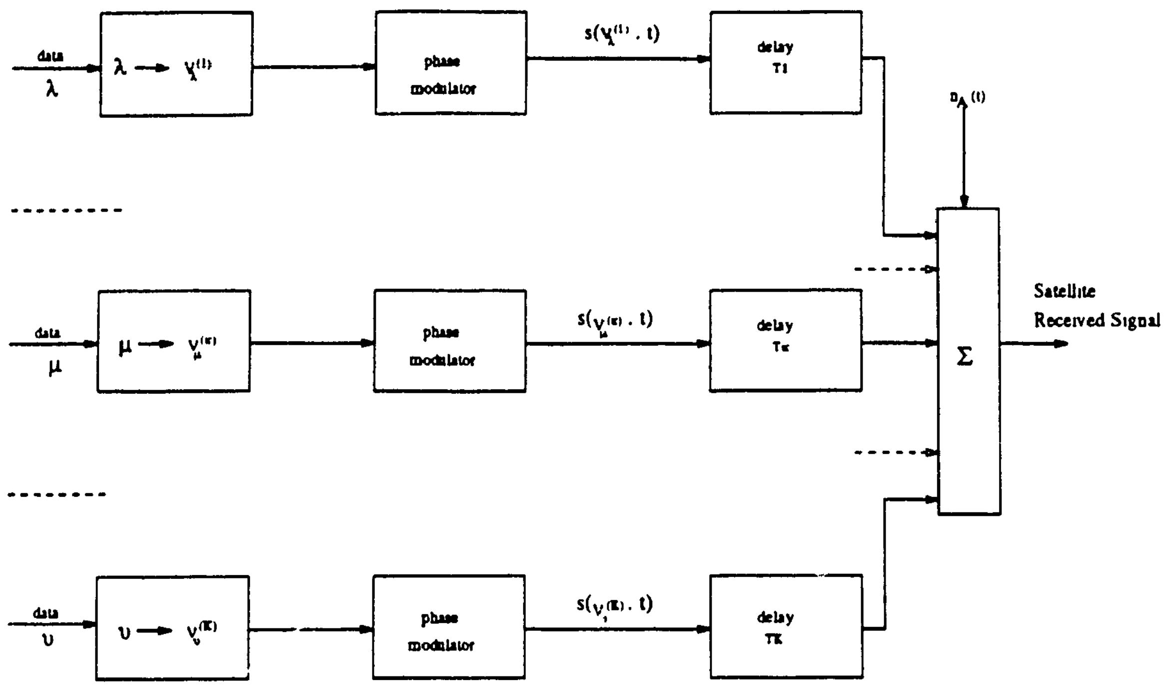


Figure 4.3: A direct sequence, spread-spectrum, multiple-access communication system

4.5).

For further analysis the capability of the CDMA to reject multiaccess interference depends on the processing gain, i.e.

$$PG = \frac{R_c}{R_s} = \frac{T_s}{T_c} = \frac{R_c}{R_b} * log_2 M \tag{4.2}$$

This is higher than that of a classic CDMA system which has the value

$$PG = \frac{R_c}{R_b} = \frac{T_b}{T_c} \tag{4.3}$$

Needless to say that with the classic system the bandwidth expansion factor due to the DS system is $\frac{T_b}{T_c} = \frac{W}{R_P} = N$, while for the orthogonal CDMA system, this is given by $\frac{T_b}{T_c} = \frac{W}{R_P} = \frac{N}{log_2 M}$, this provides better efficiency, i.e.

$$\bar{\eta} = \sum_{i=1}^{4} \frac{K_i' R_{P_i}}{W} = \sum_{i=1}^{4} \frac{K_i' log_2 M}{N} \tag{4.4}$$

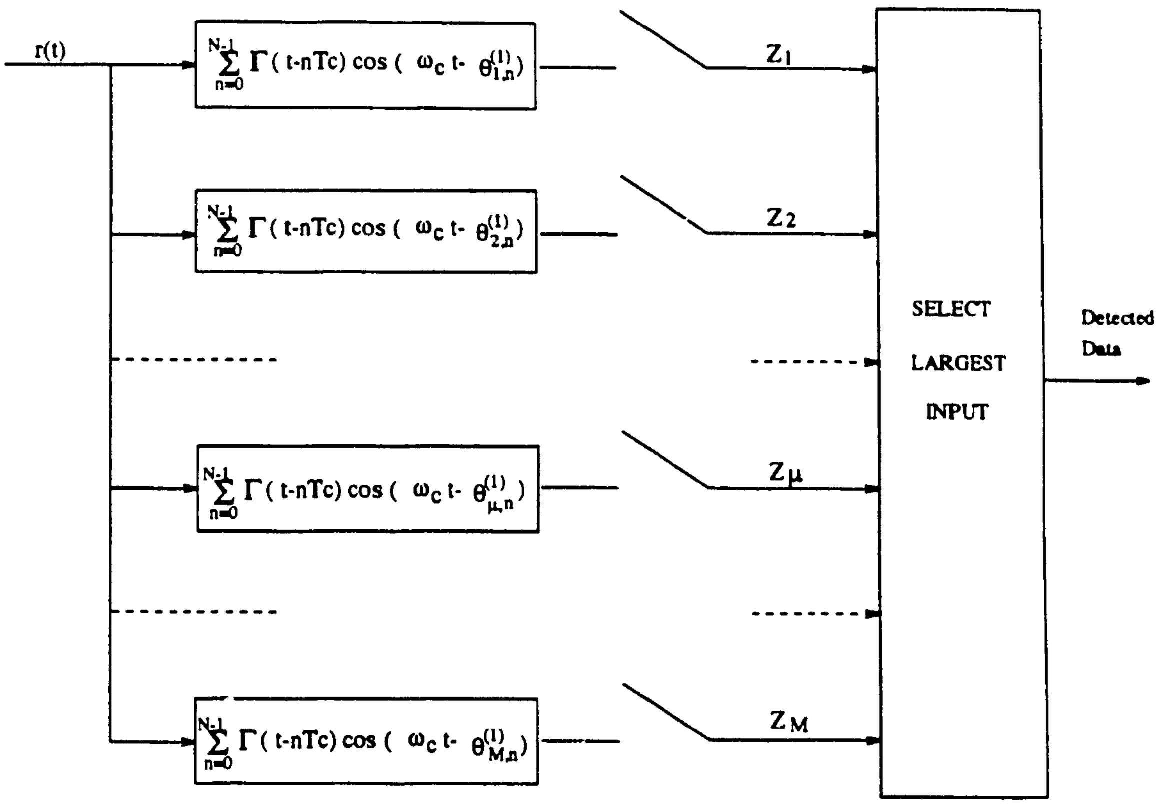


Figure 4.4: Matched filter receiver for a typtical despreaing by long user code is assumed to yield r(t)

As apposed to

$$\bar{\eta} = \sum_{i=1}^{4} \frac{K_i'}{N} \tag{4.5}$$

for the classic CDMA (no orthogonal) systems. Where the summation in equations (4.4) and (4.5) are due to the presence of 4 classes of services in the 4 subframe, constituting the total frame.

Because congestion control is the main theme of this thesis, we don't elaborate further on the modulation, multipath, fading, diversity, and bit error probabilities aspects of the CDMA access scheme used within the same class. However, the reader is urged to consult [Pahlevan][26] and [Enge][27] from which we obtain the probabilities of bit errors and hence the probability of cell success on the uplink frame as will follow shortly. Worthnothing too that our admission and congestion control procedures could be used with other Direct Sequence systems besides the specific orthogonal set CDMA technique that was presented here.

4.3 Flow Control Techniques At The Cell Level

In order to control the uplink congestion, (i.e. improve cell success probability) the downlink frame has a control overhead where the satellite sets up indicators to dictate whether new calls are accepted or blocked for each class. To arrive at this decision, the satellite will have to average out the number of correctly received packets over all CDMA receiver banks as well as over the slots of the subframe of the class.

In the case of isochronous calls (within their specific subframes) the congestion indicators bits are looked upon only by the new calls. If the indicator is set to "blocked" for such a class, the user will not start the new call to the satellite on this subframe, it will wait until the indicator shows "new calls accepted" and then transmits the call to the satellite. Ongoing isochronous calls do not have to monitor those indicators in the downlink frame. This monitoring process is considered a part of the dialing process (call establishment).

The congestion indicators from the satellite radiated on the downlinks on different links make up a safety mechanism that prevents new calls from congesting their uplink and/or the destination downlink (for example, in multibeam and hot spot traffic situation many uplink calls may want to go to the same destination). However the initial analysis in this paper considers only uplink congestion of a single beam situations.

In the case of asynchronous class such as interactive data and file transfer, each and every transmission is permitted only after the user has identified that congestion indicator is set to "data accepted". Whenever the indicator for these classes show "blocked" the ongoing user call must stop the transmission in the uplin k subframe untill congestion indicators permit again. In the asynchronous classes, the indicator is always consulted by the user even when the call is in progress.

It is assumed that the satellite bases the evaluation of the indicator bit status (Fig 4.5) on both source and destination uplinks which is not a problem in single link coverage. For multibeam switched satellites, the onboard processor will have to couple the congestion status of the zones of the communicating ends.

In the latter case, the full indicators for all the classes and all the beams are transmitted on each and every downlink frame in each beam. This will enable the user trying to establish a call to immediately recognize whether he can proceed with the call or not even if the source and destination are not on the same beam with very little involvement on the part of the satellite. The satellite bases its indicators on the congestion status of all zones. In this paper we do not address these switched satellite congestion control schemes,

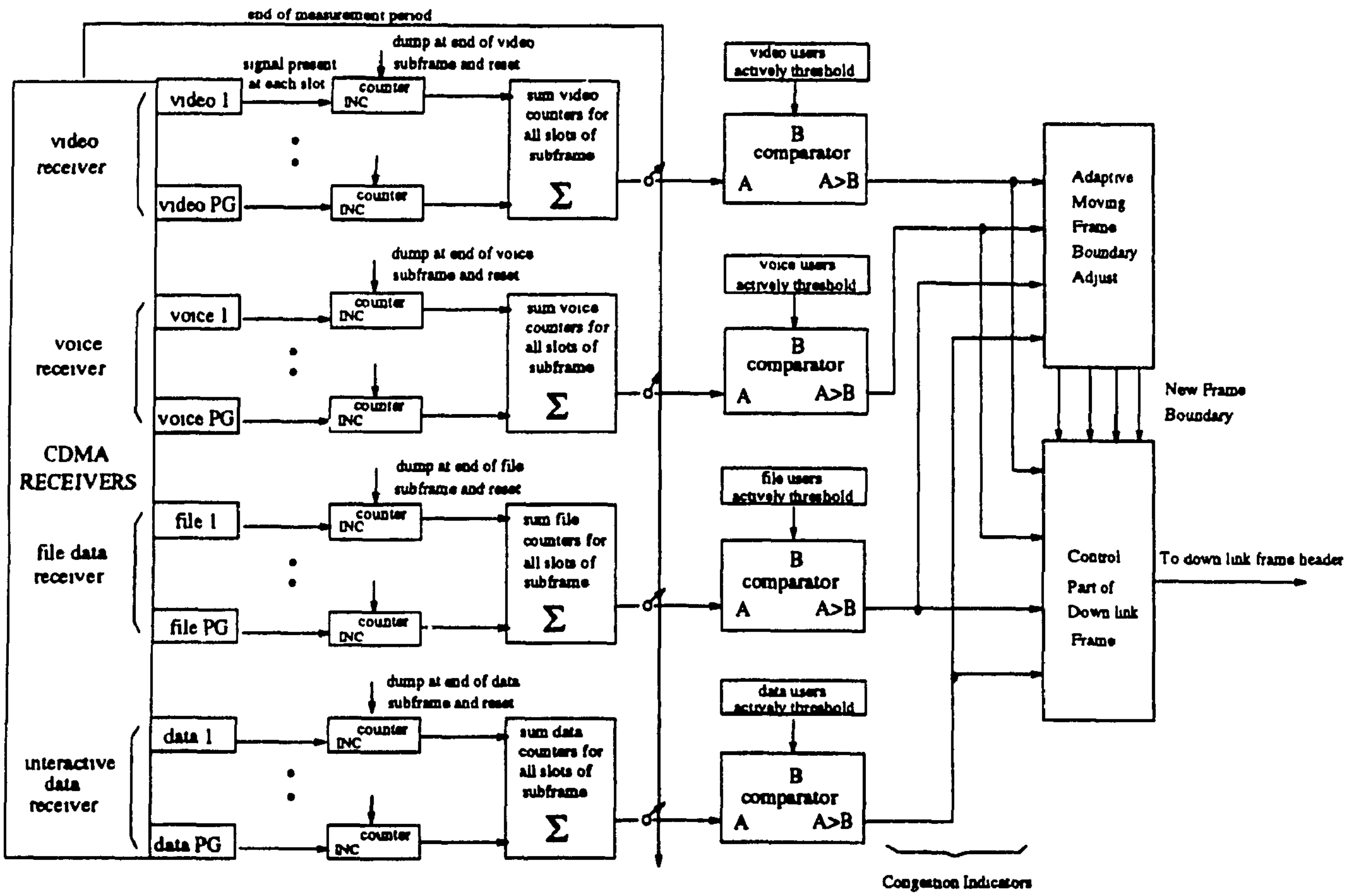
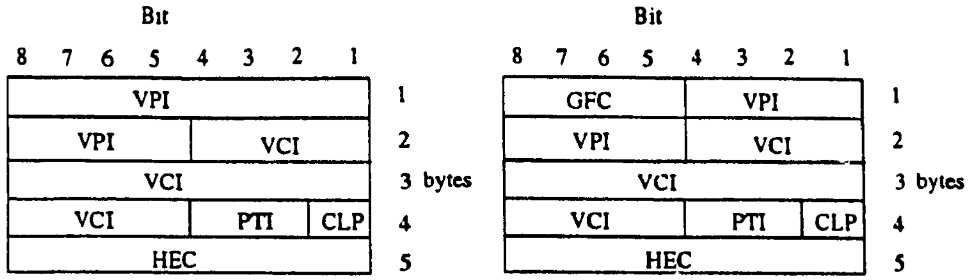


Figure 4.5: Schematic diagram for generating the congestion indicators



ATM cell header format across the NNI   ATM cell header format across the UNI

Figure 4.6: ATM cell header

labelfig 4.6

we restrict the analysis to single zone (one beam) satellite systems and discuss only the uplink frames issues.

4.4 Cell Structure Of The ATM/CDMA Oriented Network

Various service classes share the uplink frame satellite channel. Each class transmission is within specific subframes and it is done according to the ATM approach whereby fixed size cells (Fig 4.6) are generated and transmitted . Since each cell is transmitted over a slot in the class subframe and the slot loaction is not fixed for a particular communication of that class, the satellite has to be able to route uplink incomming cells to appropriate downlink outgoing cells. The satellite may have to convert the Virtual Path Indentifier (VPI) , and Virtual Circuit Identifier (VCI) in the comming cell to another VPI, VCI in the forwarded cell downlink. The satellite is thus performing some simple translation functions.

The allocation of the VPI, VCI could be either fixed to users or it could be variable from call to call. This latter approach would require satellite involvement and negotiation for the channel assignment. In our analysis we will consider fixed allocation of the values of VPI, VCI identified respectively by 9 and 28 bits fields of cell, used by the satellite. Also, in this thesis we consider the case where each user is identified by one long spread spectrum code as in Fig 4.2 and fixed VPI, VCI.

The 28 bits of the combined VPI, VCI fiels give 2^{28} possible combinations. Each user is assigned one or more of these combinations and the identities of available unused spread codes are announced in the downlink for new calls to grab them. As soon as the satellite recognizes the user code and VPI, VCI uniquely, it will recongnize that the user became active without having to go through the classic call establishment procedure. However this does not preclude the necessity of translation to yield the downlink with the VPI, VCI values assigned to a destination for the specific call.

4.5 Subframe Assignment And Moving Boundary Policies

The bandwidth necessary for an isochronous call (video or voice) changes during the call. However in order to simplify the subframe size determination process to follow, we assume the maximum band necessary for each call. Also in the process of evaluating the instantaneous performance, we allow more calls to exist in the video or voice frame portion for better utilization of the channel in face of video or voice call rate variable (call burstiness). To avoid deadlock, video or voice users may randomly select slots in each frame for their transmission and change those from frame to frame; this has the throughput advantage of statistical multiplexing but on the expense of creating delay jitter. However this delay jitter is minimal since video calls are restricted to a certain portion of the overall frame.

If the video user slots location is selected as fixed in the frame for the duration of the call (whether contiguous or not). We will have isochronous service (no delay jitter) on the expense of a higher bit error probability due to a possible higher number of users transimitting on the same slot. In this case, one could further assign non overlapping slots for video or voice users , where each call uses different slots. However in the latter case, call blocking probability could become higher thus requiring more involved control for scheduling the calls (not followed in this work).

One approach to this call scheduling as envisioned here is to allow more than one call from different users calls (up to an acceptable limit) to coexist in the same subframe assigned to a call, but not specific slots. The satellite indicates to earth station the section of the frame to use for a specific service. It does not allocate specific slots to specific calls. The access technique proposed is the CDMA/TH (Time Hopping) which is a method that has the adventage of requiring less control since the satellite has only to relay the service subframe assignments to the earth station, i.e. the total number of slots assigned to each class of service rather than assign specific slots for each call of each class. This assignment is reconfigured periodically and the size of the subframe may change according to usage and priority. Users will be notified by special control field in the satellite downlink in regard to changes in uplink subframe size as well as short range average cell congestion.

Each user selects sufficient slots of each subframe, randomly at the start of the call but keeps trying to transmit only in these slots throughout the call (so as to minimize dalay jitter). The randomness of call arrival and the resulting statistical multiplexing are very similar to the case where the user changes his slots from frame to frame.

Four classes of service share the satellite uplink frame. Their transmissions follow an ATM approach whereby fixed size cells are generated and transmitted over the uplink frame. The services users have different priorities. The video service has the highest priority and is followed by the voice service, the file data service and the interactive data service in that order. The uplink frame is divided into subframes for each of these services. The initial subframes size are N_{si} slots, i=1,2,3,4.

The maximum size for the video service is

$$N_{s1} = \frac{K_1 n\mu_1}{L_1} \tag{4.6}$$

where K_1 is the design maximum allowable number of video calls in the frame.(corresponds to full video load presence)

N_{s1} is the design number of slots in each frame used for the video service.

L_1 is the maximum allowable number of cells in each slot for the video service. As determined, by the required video quality (delay, cell blocking, ... etc) which depends on the probability of bit and cell transmission errors as will follow shortly.

$n\mu_1$ is the peak number of cells per call in the frame necessary to accommodate a single video call.

If the video subframe is not fully utilized, i.e. the number of active cells used is $K_1' < K_1$, then the subframe size N_{s1}' and the limit L_1' follows the simple rules

$$N_{s1}' = N_{s1}\sqrt{\frac{K_1'}{K_1}} \tag{4.7}$$

and

$$L_1' = L_1\sqrt{\frac{K_1'}{K'}} \tag{4.8}$$

This active number of calls K_1' is estimated at the satellite depending on the number of active CDMA satellite receivers (each handling one call) as in Fig.4.5 and a similar process is conducted to estimate K_2', K_3', K_4' and the necessary N_{s1}', N_{s2}', N_{s3}', N_{s4}'. These measures and the inherent moving frame boundary policy are executed on call basis, i.e. the congestion indicators of Fig.4.5 and the frame reconfiguration commands to accommodate new mixes of traffic are updated and dumped for each call (on slow call basis). It is possible that while this policy is in effect, the instantanous traffic per slot (# of overlapping cells) is still high[2], thus requiring an additional and faster congestion control technique that operate at the cell level as will follow shortly[3].

The reduction of N'_{s1} in equation (4.7) follows (among other possibilities) a square root law. Noting That,

$$N'_{s1}L'_1 = N_{s1}L_1\frac{K'_1}{K_1} = K'_1 n\mu_1 \tag{4.9}$$

we see

$$K'_1 = \frac{N'_{s1}L'_1}{n\mu_1} \tag{4.10}$$

Which implies a barely adequate capacity (N'_{s1},L'_1) to accommadate the current number of calls K'_1. Voice users, in addition to their own assigned subframe, could be given slots from the video subframe if available. File data users, likewise could be assigned additional slots from the video and voice subframes (if available). This service has however less priority over the video subframe than voice service. Similarly, the interactive data service could be offered available slots from the video, voice and file data subframes.

Voice slots, when only restricted to the design voice subframe size has the minimum capacity N_{s2}. However when surplus slots from the video subframe is also available then

$$N'_{s2} = N^*_{s2}\sqrt{\frac{K'_2}{K_2}} \tag{4.11}$$

and

$$L'_{s2} = L_2\sqrt{\frac{K'_2}{K_2}} \tag{4.12}$$

[2]e.g. due to imperfection of the traffic estimation process in Fig. 5 and/or changes in the instantaneous burst activities of the users .

[3]Users will monitor certain control field in the satellite downlink to recognize the instaneous congestion and adopt accordingly.

From the above discussion, it follows that,

$$N^*_{s2} = N_{s2} + (N_{s1} - N'_{s1}) \tag{4.13}$$

For the i^{th} class i=2,3,4 it is easy to conclude that

$$N'_{si} = N^*_{si}\sqrt{\frac{K'_i}{K_i}} \tag{4.14}$$

$$L'_{si} = L_i\sqrt{\frac{K'_i}{K_i}} \tag{4.15}$$

$$N^*_{si} = N_{si} + (N_{si-1} - N'_{si-1}) \tag{4.16}$$

and

$$K'_i = \frac{N'_{si}L'_i}{n\mu_i} \tag{4.17}$$

It is classical exercise to show the effect of this moving frame policy on the distribution of active calls and blocking probability.[30]

4.6 Congestion Control Policies For The Uplink Frame

The objective of this section is to analyze the steady state performance of the various service classes under different congestion control policies. According to Fig.5, the satellite computes the number of busy receivers in each class, average this over all time slots in certain measurement window of W_i frames, i.e. $W_i * T_F$/sec and set the congestion indicators of each class accordingly. Users on the ground monitor these (on the down link frame) and decide whether to transmit their head of line cell in their buffer or deffer it to next subframe and so on.

The congestion policies that we adopt to limit the instantaneous number of overlapping cells in each slot and hence obtain acceptable probabilities of bit error and packet success are applied at each ground user interface to the satellite uplink and are modified versions of the Leaky bucket (LB) and Virtual Leaky Bucket (VLB) techniques [36]. The traffic of each class i is modelled as an Markov Modulated Poisson Process (MMPP). Various bursty active users are assumed to go from active (generate a cell) to idle (no cell generation) each cell time (slot) as in Fig3.1. (in chapter 3)

Solving this state diagram one gets the probability of one active user call being in an active and idle bursts, i.e. $P_A = [\alpha_i/(\alpha_i + \beta_i)]$ and $P_B = [\beta_i/(\alpha_i + \beta_i)]$. It follows that the probability of generating k_i' active bursts from the K_i' independant active calls is

$$P(k_i') = \begin{pmatrix} K_i' \\ k_i' \end{pmatrix} \left(\frac{\alpha_i}{\alpha_i + \beta_i}\right)^{k_i'} \left(1 - \frac{\alpha_i}{\alpha_i + \beta_i}\right)^{K_i' - k_i'} \qquad (4.18)$$

where $1/\beta_i$ is the mean active time for the call and $1/\alpha_i$ is the mean inactive (idle) time which can easily be derived. For the leaky bucket (LB) congestion control the violating cells are rejected; however for the modified virtual leaky bucket (MVLB), non violating cells always enter the buffer as with LB while violating cells enter only if there is no congestion on the ATM channel (The virtual leaky bucket always admits cells in the buffer even during congestion periods). However for MVLB and as will follow shortly, violating cells are not transmitted (will be dropped) once they become at the head of their buffer and there is congestion on the uplink frame. For LB, no violating cells are allowed in the user buffer to start with. It should be noted that the MVLB is actually the only congestion control mechanism that tries to follow the ATM cell loss priority mechanism (in sequence cell delivery).

In the case of the MVLB, the probability of a typical user of class i finally transmitting a cell on one of the subframe slots is given by

$$\xi_i|_{k_i'} = \left[(1 - P_{v_i}) + P_{v_i}(1 - P_{cg_i}(K_i'|_{k_i'}))\right] \frac{\rho_i}{K_i'} \qquad (4.19)$$

Equation (4.19) states that a user generating a cell with probability ρ_i/K'_i is allowed to let it go on the uplink if it is a non violating (with probability $(1 - P_{v_i})$) or violating but there is no congestion on the channel with probability $(1 - P_{cg_i}(K'_i|_{k'_i}))$.

For the LB technique this becomes

$$\xi_i|_{k'_i} = (1 - P_{v_i})\frac{\rho_i}{K'_i} \tag{4.20}$$

since violating cells are discarded up front before entering the user buffer.

In the case of no policing, no congestion control, we have

$$\xi_i|_{k'_i} = \frac{\rho_i}{K'_i} \tag{4.21}$$

where

$$\rho_i = \left(\frac{\alpha_i}{\alpha_i + \beta_i}\right)\frac{K'_i n\mu_i}{N'_{si}} \tag{4.22}$$

is the total per slot traffic utilization of all users.

The probability of cell violation by a typical user of class i and quality of service q_i is given by,

$$P_{v_i} = \sum_{l=\eta_i+1}^{n\mu_i W_i} \binom{n\mu_i W_i}{l} \left(\frac{\rho_i}{K'_i}\right)^l \left(1 - \frac{\rho_i}{K'_i}\right)^{n\mu_i W_i - l} \tag{4.23}$$

where W_i in frames is the congestion measurement window size at the ATM layer which is equivalent to $W_i(N_{si} + N_{s2} + N_{s3} + N_{s4})slots$.

And

$$\eta_i = n\mu_i * W_i * q_i \tag{4.24}$$

is the number of cells a typical user generating $n\mu_i$ cells per frame at quality rate $q_i \leq 1$ yields over the total measurement window of W_i frames. Beyond this number, i.e. when $l \geq \eta_1$, the user is actually violating the connection agreement.

Note that P_{v_i} is the probability of cell violation and $P_{cg_i}(K_i')/k_i'$ is the probability of cell congestion on the uplink CDMA/ATM satellite channel given k_i' active bursts from K_i' active calls. The distribution of the number of overlapping cells per slot m that is generated and transmitted by the k_i' active users bursts (which is the cell bulk arrival in the state diagram to follow) on a typical subframe of class i,

$$\theta_i(m)|k_i' = \frac{(nn(\xi_i|_{k_i'}))^m/m!}{\sum_{j=0}^{nn}(nn(\xi_i|_{k_i'}))^j/j!} \tag{4.25}$$

Where nn is the maximum potential number of cells that can be generated from all the active k_i' bursts during a subframe of size N_{s_i}' slots.

$$nn = n\mu_i k_i' \tag{4.26}$$

The state diagram in Fig 4.7 represents the total number of cells m_i in steady state in all users buffers of class i (the distributed buffer). The steady state distribution is given by $\Psi_i(m)$ to be found by solving the staedy state equations of Fig 4.7 as shown in following paragraph.

In the Fig 4.7, the probability that the total distributed buffers contents of all users add to j cells is found from the state transition equation (based on flow correservation principle),

$$\Psi_j = \sum_{0}^{LB} b_{ij}\Psi_i \tag{4.27}$$

where the transition probabilities are given by

$$b_{ij} \quad = \quad S_i\theta_j + (1 - S_i)\theta_{j-i} \quad where \quad j > i, j \leq nn, (j - i) \leq nn, forward \tag{4.28}$$

Here, it is easily seen that the total distributed buffer moves buffer users from state i to state j if there are j arrivals (with probability θ_j) and i services (with probability S_i

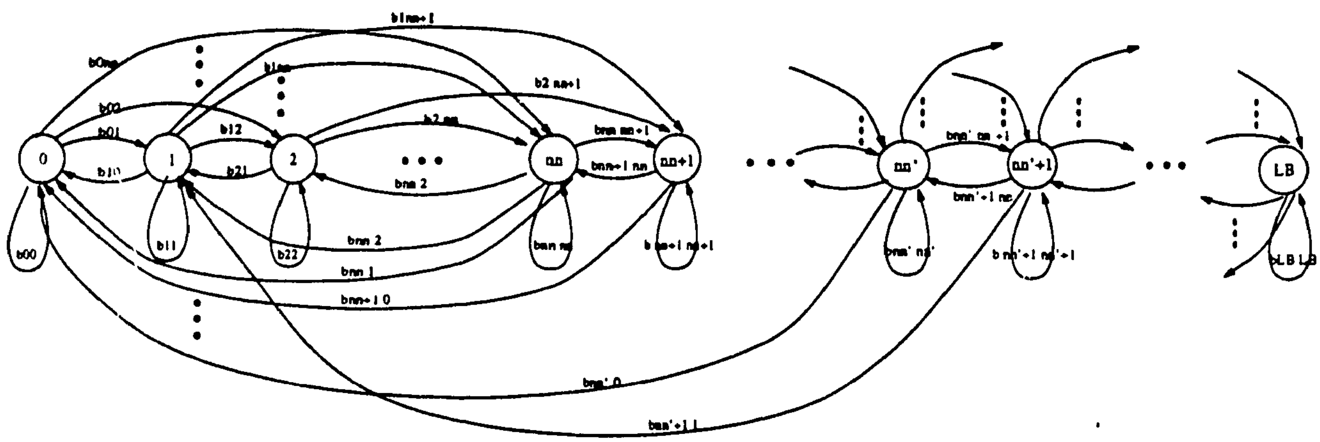


Figure 4.7: State diagram of the total number m_i in the buffers of all active users of class i (distributed buffer)

), or there could be no service (due to CDMA interference with probability (1-S_i)) and (j-i) arrivals (with probability θ_{j-i}) . Thus giving rise to the second term in b_{ij} .

Similarly,

$$\begin{aligned} b_{ij} &= (1-S_i)\theta_{j-i} \qquad where \;\; j>i, j>nn, i\leq nn, (j-i)\leq nn, forward \\ &= (1-S_i)\theta_{j-i} \qquad\qquad\quad j>i, nn\leq i\leq nn', (j-i)\leq nn, forward \\ &= (1-S_{nn'})\theta j-i \qquad\qquad j>i, i>nn', (j-i)\leq nn, forward \end{aligned} \tag{4.29}$$

where $nn' = (N_{si} * PG)/\sigma_i$ and we note the last transition that there could be no service (with probability (1- $S_{nn'}$)) and (j-i) arrivals (with probability θ_{j-i}). Here, we assume that any state beyond state nn' which could serve only the maximum number of cells. Since the maximum number of arriving cells is nn which is smaller than nn', only no service could prevail in this case.

$$b_{ij} = S_i\theta_j \quad where \quad j<i, i\leq nn+1, (i-j)\leq nn+1, backward \tag{4.30}$$

Here we have a backward transition probability that the total distributed buffers contents of all users add to j cells is found from the state transition equation (based on flow correservation principle),

$$\Psi_j = \sum_{0}^{LB} b_{ij}\Psi_i \tag{4.31}$$

where the transition probabilities are given by

$$b_{ij} \;=\; S_i\theta_j + (1 - S_i)\theta_{j-i} \quad where \quad j > i, j \leq nn, (j-i) \leq nn, forward \tag{4.32}$$

Here, it is easily seen that the total distributed buffer moves buffer users from state i to state j if there are j arrivals (with probability θ_j) and i services (with probability S_i), or there could be no service (due to CDMA interference with probability (1-S_i)) and (j-i) arrivals (with probability θ_{j-i}) . Thus giving rise to the second term in b_{ij} .

Similarly,

$$\begin{aligned} b_{ij} &= (1 - S_i)\theta_{j-i} && where \quad j > i, j > nn, i \leq nn, (j-i) \leq nn, forward \\ &= (1 - S_i)\theta_{j-i} && j > i, nn \leq i \leq nn', (j-i) \leq nn, forward \\ &= (1 - S_{nn'})\theta j - i && j > i, i > nn', (j-i) \leq nn, forward \end{aligned} \tag{4.33}$$

where $nn' = (N_{si} * PG)/\sigma_i$ and we note the last transition that there could be no service

(with probability (1- $S_{nn'}$)) and (j-i) arrivals (with probability θ_{j-i}). Here, we assume that any state beyond state nn' which could serve only the maximum number of cells. Since the maximum number of arriving cells is nn which is smaller than nn', only no service could prevail in this case.

$$b_{ij} \;=\; S_i\theta_j \quad where \quad j < i, i \leq nn+1, (i-j) \leq nn+1, backward \tag{4.34}$$

Here we have a backward transition, i.e. from high state to low state. There could be only one possibility which is i services (with probability S_i) and j arrivals (with probability θ_j).

Also,

$$\begin{aligned}
b_{ij} &= S_i\theta_j & \textit{where} \quad & j < i, i \leq nn', j \leq nn, (i-j) \leq nn', backward \\
&= S_{nn'}\theta_{nn'-i+j} & & j < i, i > nn', nn' - nn \leq i - j \leq nn', backward \\
b_{ii} &= S_i\theta_i + \theta_0(1 - S_i) & & 0 \leq i \leq nn, self - loop
\end{aligned} \tag{4.35}$$

where the last equation represents a self-loop, from state i back to state i. There are i arrivals (with probability θ_i) and i services (with probability S_i), or there could be no services (with probability (1-S_i))and no arrivals (with probability θ_0).

Finally,

$$\begin{aligned}
b_{ii} &= \theta_0(1 - S_i) & \textit{where} \quad & nn < i \leq nn', self - loop \\
&= \theta_0(1 - S_{nn'}) & & nn' < i \leq LB, self - loop
\end{aligned} \tag{4.36}$$

Once the all elements of coefficient matrix $[b_{ij}]$ are computed, the following state transition matrix equations will be solved to obtain the requires steady state vector of the distributed buffer average content in cells, i.e. $[\bar{\Psi}]$.

$$[A][\Psi] = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 0 \\ 1 \end{bmatrix} \tag{4.37}$$

where

$$\begin{aligned}
A &= [a_{ij}], & & \\
a_{ij} &= b_{ij} & \textit{when} \quad & i \neq j, i \neq LB, \\
&= b_{ij} - 1 & \textit{when} \quad & i = j, i \neq LB, \\
&= 1 & \textit{when} \quad & i = LB, \\
\Psi &= [\Psi_0, \Psi_1, \Psi_2, \Psi_3, ...\Psi_{LB}] & &
\end{aligned} \tag{4.38}$$

It is essential to determine the bit error probability occuring when a certain number of cells c of a certain class shares the same slot. The same curve for the probability of bit error can be used for all priority classes [26]. For example to obtain P_{b_i} for class i, we determine the number of active cells per slot and apply the result to the probability of bit error curve.

To find S_m, i.e. the probability of c cells per slot corresponding to state m over the CDMA uplink, we notice that the probability of correct cell detection and reception (this is the bulk cell service in the state diagram) for state m is given by,

$$P_{d_i}(m)|_{k'_i} = \left[1 - P_b(c)|_{k'_i}\right]^{L_p} \tag{4.39}$$

Where m=1,2,... nn/σ_i, and $c = m\sigma_i/N'_{s_i}$ is the actual number of active cells from all active bursts trying transmissions in a certain slot, and where L_p is the cell length in bits. σ_i expresses the probability that the user tries to transmit the head of the line cell of the buffer to be found shortly. $P_b(c)|k'_i$ is the probability of bit error detection on the channel given a certain number of overlapping cells c. This is obtained from the probability of error curve in [26]. To find the average number of interfering cells per slot c in the subframe of class i we divide the existing number m in the buffer by the number of available slots N'_{s_i} and use this number to find the probability of bit error $P_b(c)|_{k'_i}$ in equation (4.39) and finally the service probability $S_i(m)$.

$$S_i(m)|_{k'_i} = P_{d_i}|_{k'_i}\left[\frac{\sigma_i m}{N'_{s_i}} + 1\right] \tag{4.40}$$

where m=1,2,..., $(N'_{si} * PG)/\sigma_i$,and [*] denotes the integer part of x.

$S_i(m)/k'_i$ is the per slot correct cell detection probability at a typical satellite receiver which is also the service probability of the state diagram of figure4.7

We turn our attention now to the evaluation of σ_i. For leaky bucket $\sigma_i = 1$. since all cells stored in the user buffer for the leaky bucket case will try transmission to the channel with probability 1. Violating cells are readily discarded before entering the buffer.

However, for the modified virtual leaky bucket (MVLB) we have

$$\sigma_i = (1 - P_{v_i}) + P_{v_i}(1 - P_{cg_i}(K'_i|_{k'_i})) \tag{4.41}$$

where it is seen that the stored head of line in the user buffer tries transmission in a certain slot of the link subframe if it is a non violating cell, or if it is a violating cell but congestion condition on the subframe (as given by the satellite indicators in the downlink permit).

MVLB does not discard the cell upon buffer entry but saves it for latter processing. For MVLB, because of σ_i, we can go to states higher than nn and still serve $S_i(m)/k'_i$ cells for values of m up to nn/σ_i. For MVLB the service is not limited to nn; we can go to states higher than nn and insert cell drop (discard) probability at the source; moreover once congestion develops, virtual leaky bucket will start to drop (discard) cells from the buffer thus leading to less on the channel and consequently less congestion. MVLB uses σ_i which limits the cells that are transmitted.

Note the difference between ξ_i and σ_i. ξ_i is basically the probability of entry into the user buffer and depends, among other things, on ρ_i/K'_i; however, σ_i is the probability of transmitting or leaving the buffer which does not depend on ρ_i/K'_i and thus controls the service process.

All equations presented so far depended on the distributed buffer content of m cells and number of instantaneous users bursts k'_i. To find unconditional entities of interest, one has to average over the distributions of m (i.e. $\Psi(m|_{k'_i})$ and $P(k'_i|_{K'i})$).

The condition value of P_{di} on k'_i is found as

$$P_{d_i}|_{k'_i} = \sum_{m=0}^{LB} P_{d_i}(m)|_{k'i} * \Psi(m|_{k'_i}) \tag{4.42}$$

$P_{d_i}(m)/k'_i$ is the conditional value of the correct cell detection probability over state m and burst k_i'. The unconditional averaged value of P_{d_i} is

$$P_{d_i} = \sum_{k_i'=0}^{K_i'} P_{d_i}|_{k_i'} * P(k_i') \tag{4.43}$$

This is computed after convergence at all values of k_i' at one value of K_i' (number of active calls). Similarly the congestion probability averaged over (m), i.e. the probability that the congestion indicators (Fig.5) are set, is

$$P_{cg_i}(K_i')|_{k_i'} = \sum_{m=Th_i}^{L_B} \Psi(m|_{k_i'}) \tag{4.44}$$

where $L_B = K_i' B_i$ and

$$Th_i = \frac{L_B}{2} \tag{4.45}$$

Averaging this over k_i' we obtain,

$$P_{cg_i}(K_i') = \sum_{k_i'=0}^{K_i'} P_{cg_i}(K_i')|_{k'i} * P(k_i') \tag{4.46}$$

Another performance criterion is the probability that a cell is discarded upon generation.

The probability of discard, in the case of MVLB is

$$P_{discard}|_{k_i'} = \left(P_{v_i} * Pcgi(K_i')|_{k_i'}\right) \frac{\rho_i}{K_i'} \tag{4.47}$$

Which shows that a cell is discarded if it was generated and violated its quality agreement, and congestion indicators were set (indicating a congested uplink CDMA channel).

Obviously, the probability of discard in the case of leaky bucket is

$$P_{discard}|_{k_i'} = P_{v_i} \frac{\rho_i}{K_i'} \tag{4.48}$$

and is 0 in the case of no policing.

Averaging this over k_i' we obtain,

$$P_{discard} = \sum_{k=0}^{K_i'} P_{discard}|_{k_i'} * P(k_i') \tag{4.49}$$

With $L_B = K_i' B_i$ as stated before, we can compute the average statistics of buffer distributions. So the mean buffer content is

$$\overline{\Psi|_{k_i'}} = \sum_{i=0}^{LB} i\Psi(i)|_{k_i'} \tag{4.50}$$

The mean distribution buffer content is

$$\overline{\Psi} = \sum_{k_i'=0}^{K_i'} \overline{\Psi|_{k_i'}} * P(k_i') \tag{4.51}$$

The user average buffe content is

$$\overline{\Phi(r)|_{k_i'}} = \sum_{l=(r-1)k_i'}^{rK_i'-1} \Psi(l)|_{k_i'} \tag{4.52}$$

where $r = 1, 2, 3..., B_i$

The user mean buffer content is

$$\overline{\Phi|_{k_i'}} = \sum_{r=1}^{B_i} r\Phi(r)|_{k_i'} \tag{4.53}$$

The overall mean buffer distribution is

$$\overline{\Phi} = \sum_{k_i'=0}^{K_i'} \Phi|_{k_i'} * P(k_i') \tag{4.54}$$

The probability of the user buffer overflow is

$$P_{overflow}|_{k_i'} = \sum_{r=0.9B_i}^{B_i} \Phi|_{k_i'} \tag{4.55}$$

Finally, the mean probability of overflow is

$$\overline{P_{overflow}|K_i'} = \sum_{k_i'=0}^{K_i'} P_{overflow}|_{k_i'} * P(k_i') \tag{4.56}$$

It is easily seen that all the above equations are interrelated and at each k_i' one has to break the chain of computation by assuming a certain value for $P_{cg_i}(K_i')|_{k_i'}$. By knowledge of this we evaluate all the quantities that are conditional on k_i', solve the state diagram equations, obtain a new value for $P_{cg_i}(K_i')|_{k_i'}$, repeat the procedure (at same k_i') untill two consecutive value of $P_{cg_i}(K_i')|_{k_i'}$ are very close to each other. This computational algorithm is then repeated at different k_i' (at same ρ_i) and all the result are averaged over k_i' distribution (equation 4.18)... etc.

4.7 Results

Based on the preceeding new techniques and analysis theory, we have developed a computer program that calculates the main performance criterion. The performances of the four classes are expressed by the seven parameters, which are probability of cell violation, probability of discarding a cell, probability of congestion, mean distribution buffer content, single user mean buffer content mean probability of overflow and probability of correct cell detection and reception.

The system of equations in (4.18)-(4.56) is used to compute the results of this work. Once the system parameters are specific (K_i', ρ_i, $n\mu_i$, W_i, q_i), we can assume a starting value of $P_{cg_i}(K_i')|_{k_i'}$ at a certain ρ_i and k_i', evaluate at this ρ_i and k_i' the values of $\xi_i|_{k_i'}$ and P_{v_i} from equation (4.19) and (4.23), respectively. Subsequently $\theta_i(m)|_{k_i'}$ is evaluated from equation (4.25), and $S_i(m)|_{k_i'}$ can be obtained from equation (4.40) based on the probability error curve in [26]. Then the state diagram for the total ATM multiplexing mode is solved and $\Psi(m|_{k_i'})$ is obtained. A new value for the conditonal Probability $P_{cg_i}(K_i')|_{k_i'}$ is obtained from equation (4.44). With this value we repeat the procedure above (evaluate new $\xi_i|_{k_i'}$, $\theta_i(m)|_{k_i'}$, etc.) till convergence is reached in the value of $P_{cg_i}(K_i')|_{k_i'}$. Repeating the above for different k_i and average all criteria over the distribution of $P(k_i')$ (equation(4.18)), while the mean value $P_{cg}(K_i')$, $P_{discard}$, $\bar{\Psi}$, $\bar{\Phi}$, $\overline{P_{overflow}}$, P_d from equations(4.46, 4.49, 4.51, 4.54, 4.56) are obtained at the specific ρ_i.

Our results in Figs(A.1 - A.28) (refere to appendix for the details) deal only with congestion control issues. Fig A.1 shows the probability of cell violation P_v vs normalized traffic intensity ρ for interactive data. Since it only depends on the source, it is all the same for three different control schemes. P_v is becoming larger when the traffic is high. Moreover, P_v for the interactive data is largest because there is only one slot for this kind of users. Maximum number of cells per slot can be as high as the processing gain (PG). In contrast P_v for video users is much lower (Fig A.22) due to the large size of video subframe.

Fig A.2 shows the probability of discarding a cell $P_{discard}$ vs ρ for interactive data. There is no discarding for the no policing control, and more discaring for MVLB. Because of low P_v of video users, $P_{discard}$ (Fig A.23) becomes much lower than that for other services.

Fig A.3 shows the probability of congestion P_{cg} vs ρ for interactive data. For the no congestion control case, all cells enter the buffer. For the LB control, all violating cells are discarded upon buffer entry, and the congestion situation improves; For the last scheme–MVLB case, we don't discard cells upon entrying buffers, but as soon as congestion develops, those marked cells are discarded, which makes the congestion better in Fig A.3 Our results also show that the differences among the three schemes is largest for interactive data, while smallest for video.

Fig A.4 and Fig A.5 respectively show the Mean buffer content in cells and the single user mean buffer content vs ρ for interactive data. They all increase as ρ increases contrary to one's expectations MVLB provides the best results for this class and other class as well (Fig A.11, A.12, A.18, A.19, A.25, A.26).

Fig A.6 shows the mean probability of overflow vs ρ for interactive data. Again MVLB provides less overflow.

Fig A.7 shows the Probability of correct cell detection and reception vs the normalized traffic intensity ρ. It is almost 1 when the traffic is very low and decreases with higher traffic. But for the video users, the differences in the results due to the various policies are very small, because of availability of more slots in the video subframe (Fig A.28). For voice and wideband data (Fig A.14, A.21) MVLB yields better results.

Chapter 5

Conclusion

In this thesis, we have developed a novel minimum signalling call admission and congestion control techniques for hybrid ATM/CDMA satellite networks. The satellite works as a congestion and policing entity that ensure that congestion will not develop by minimum processing. We have provided the steady state performance analysis of four service classes under different control policies (Modified Leaky Bucket, Modified Virtual Leaky Bucket and No Policing).

We have applied the three of congestion control policies to the ATM/CDMA satellite network. We have find that MVLB policy provide better results. However, there is extra cost onboad satellite which has to measure uplink traffic and issues congestion control indicators in the down link which are used by users to policing their transmission.

Compared with the general case of one frame for all services, the advantage is that our subframe case reduces overhead bit, increase efficiency; the disadvantage is that it increase traffic.

Appendix

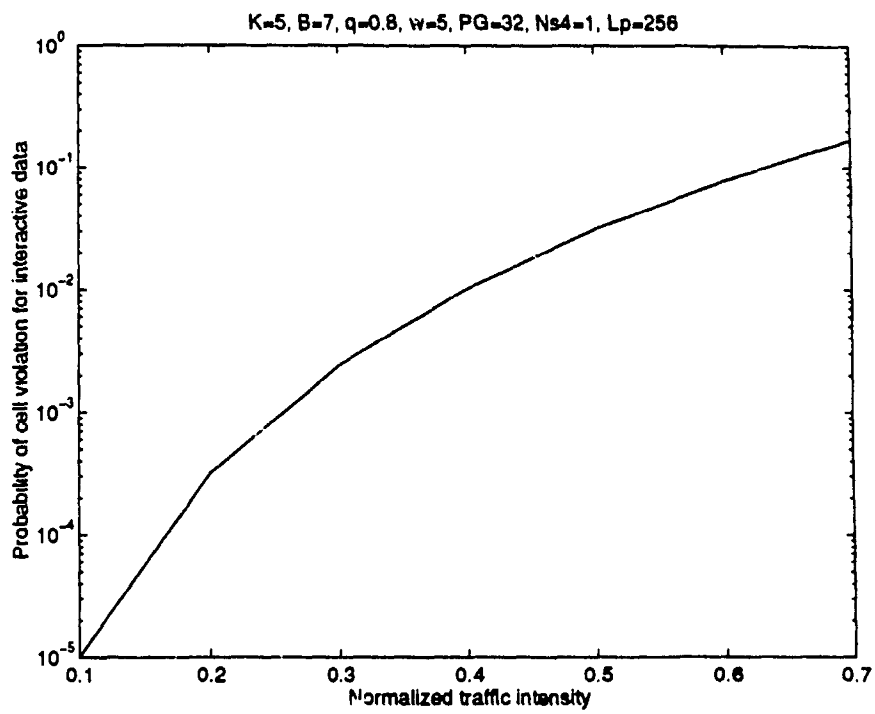


Figure A.1: Data: Prob. of cell violation P_v vs normalized traffic intensity ρ

Note: Probability of cell violation for the three schemes is all the same.

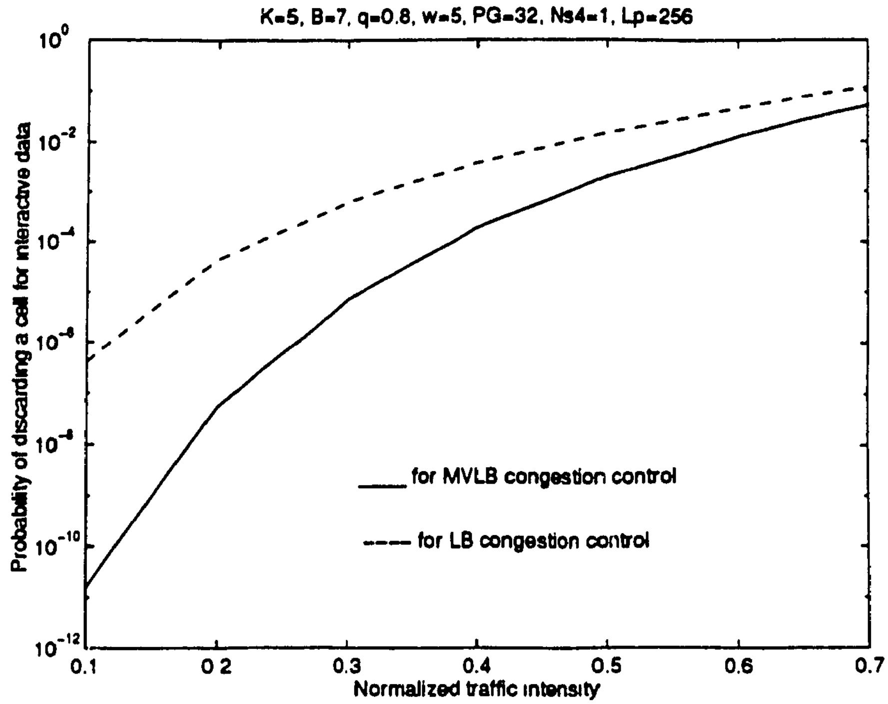


Figure A.2: Data: Prob. of discarding a cell $P_{discard}$ vs normalized traffic intensity ρ

Note: There is no discarding for the no policing control

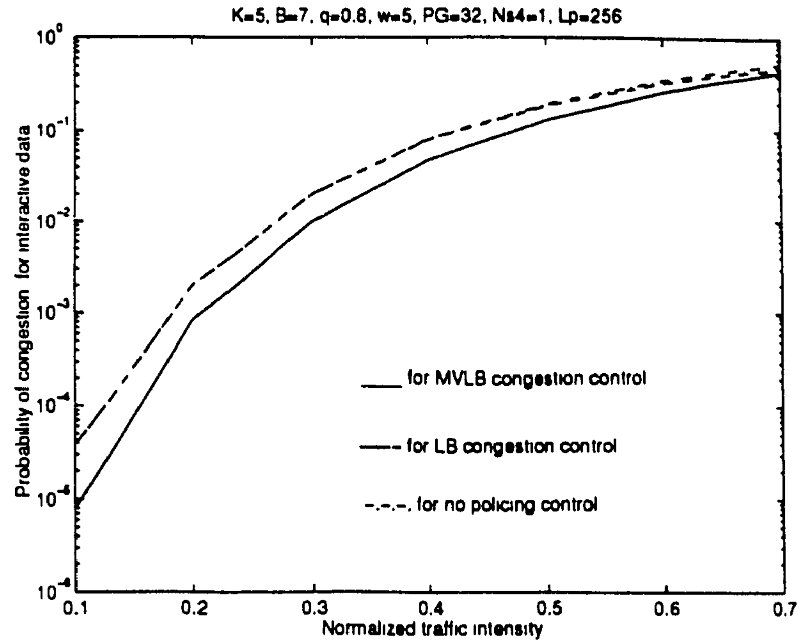


Figure A.3: Data: Prob. of congestion P_{cg} vs normalized traffic intensity ρ

Note: The results for no policing and LB congestion control are too close to distinguish.

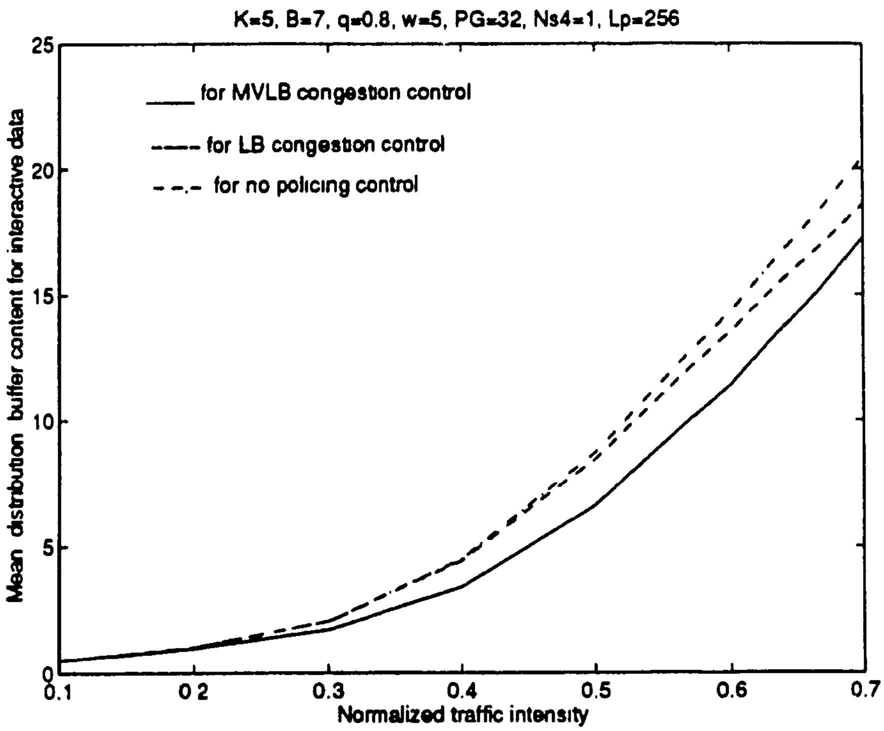


Figure A.4: Data: Mean distribution buffer content $\bar{\Psi}$ vs normalized traffic intensity ρ

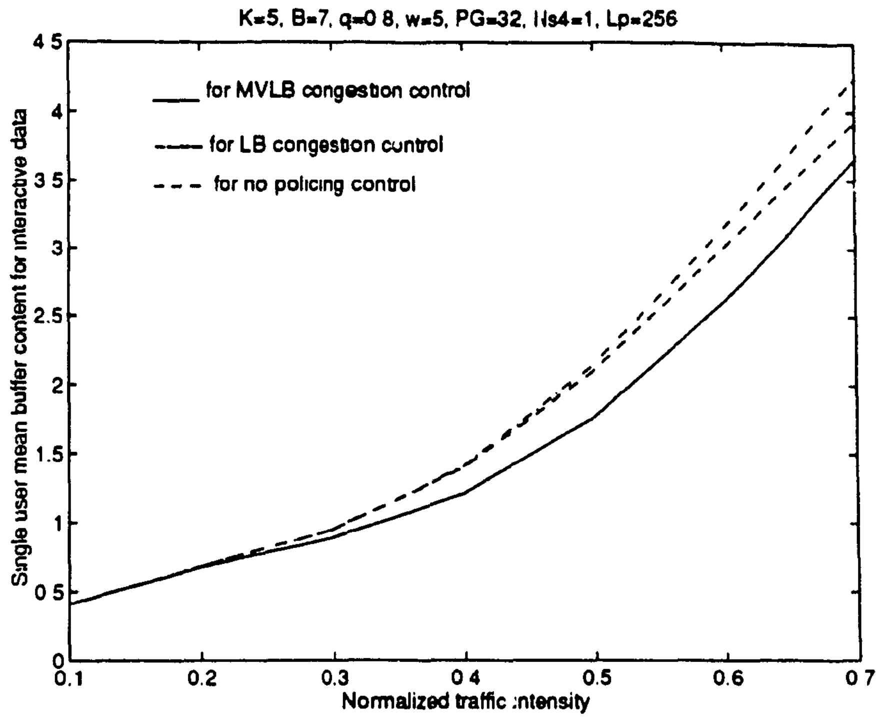


Figure A.5: Data: Single user mean buffer content $\bar{\Phi}$ vs normalized traffic intersity ρ

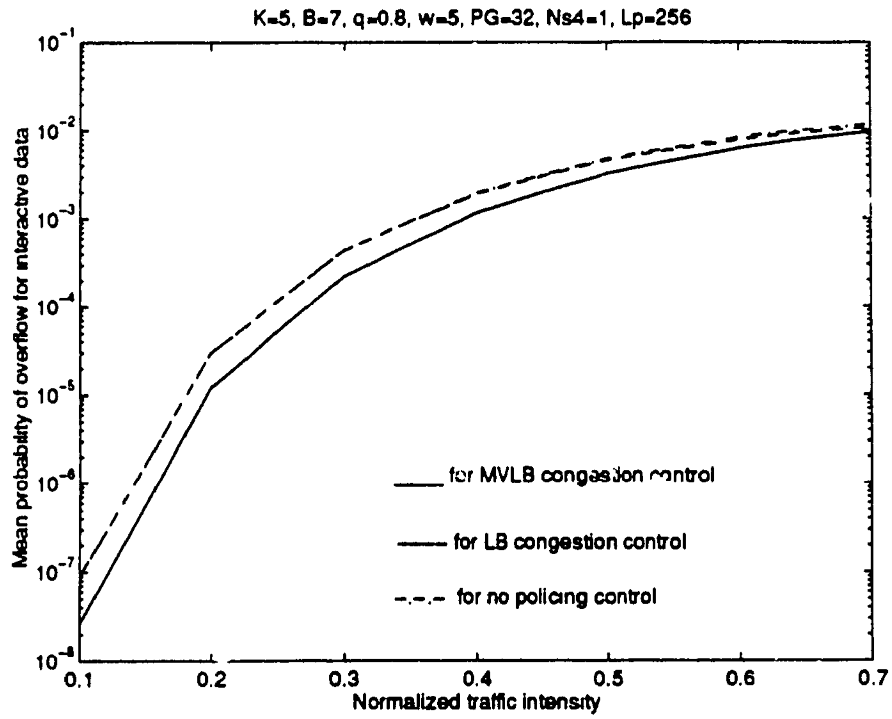


Figure A.6: Data: Mean prob. of overflow $\overline{P_{overflow}}$ vs normalized traffic intensity ρ

Note: The results for no policing and LB congestion control are too close to distinguish.

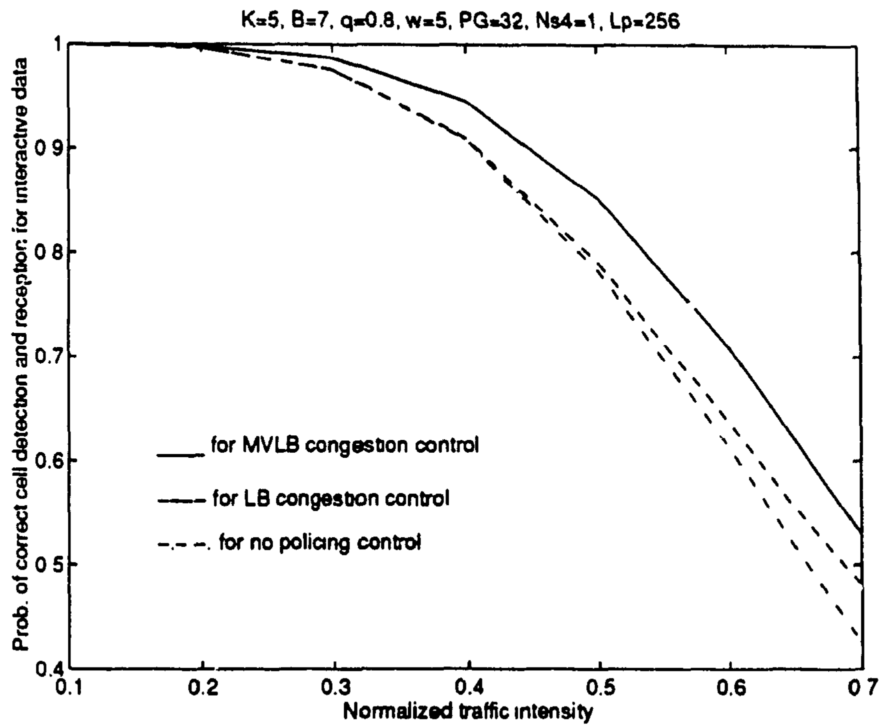


Figure A.7: Data: Prob. of correct cell detection and reception P_d vs normalized traffic intensity ρ

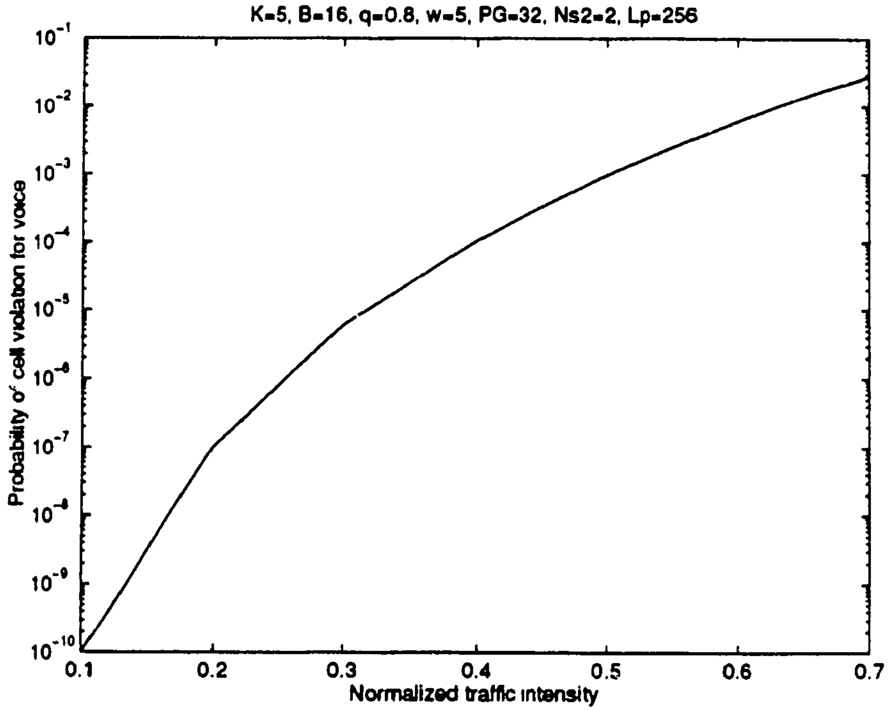


Figure A.8: Voice: Prob. of cell violation P_v vs normalized traffic intensity ρ

Note: Probability of cell violation for three schemes is all the same.

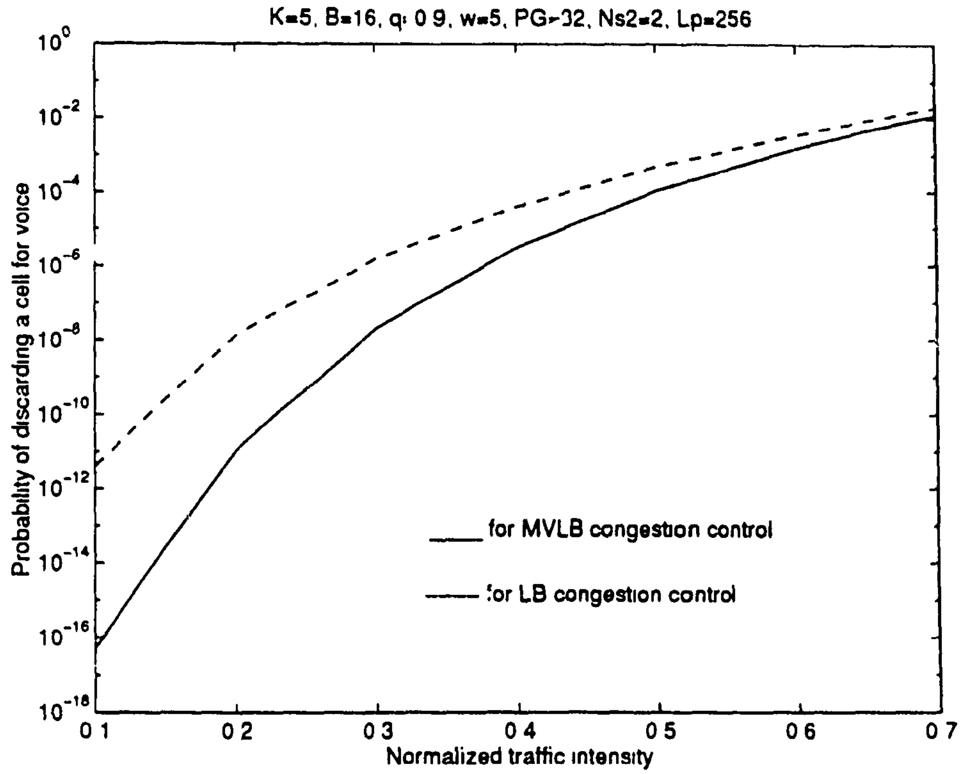


Figure A.9: Voice: Prob. of discarding a cell $P_{discard}$ vs normalized traffic intensity ρ

Note: There is no discarding for the no policing control.

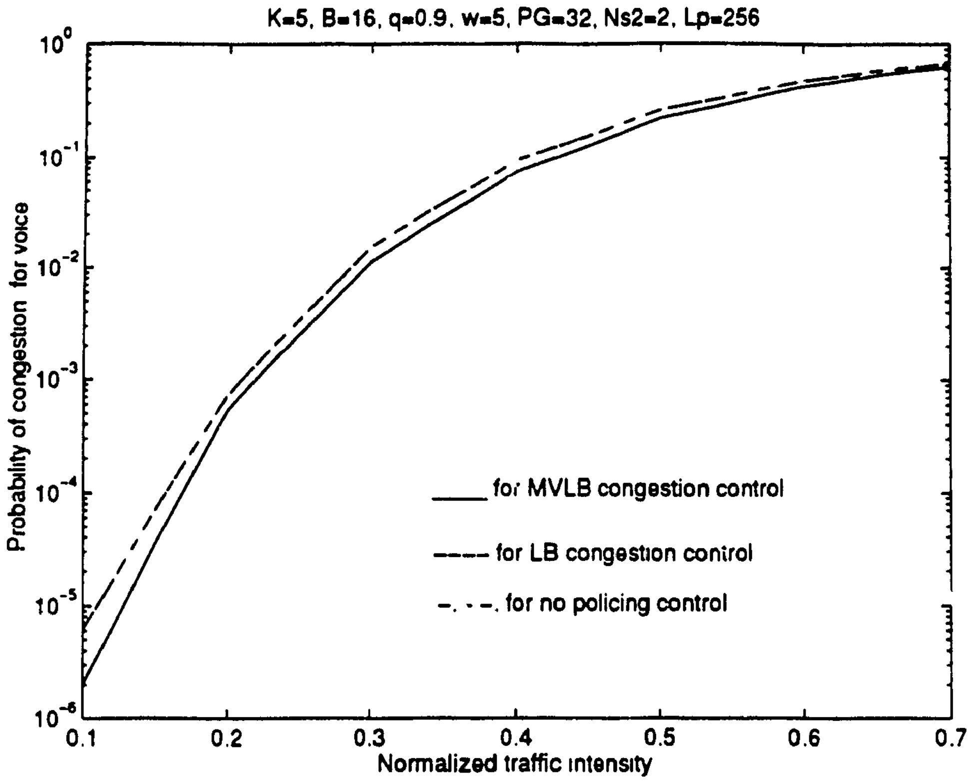


Figure A.10: Voice: Prob. of congestion P_{cg} vs normalized traffic intensity ρ

Note: The results for no policing and LB congestion control are too close to distinguish.

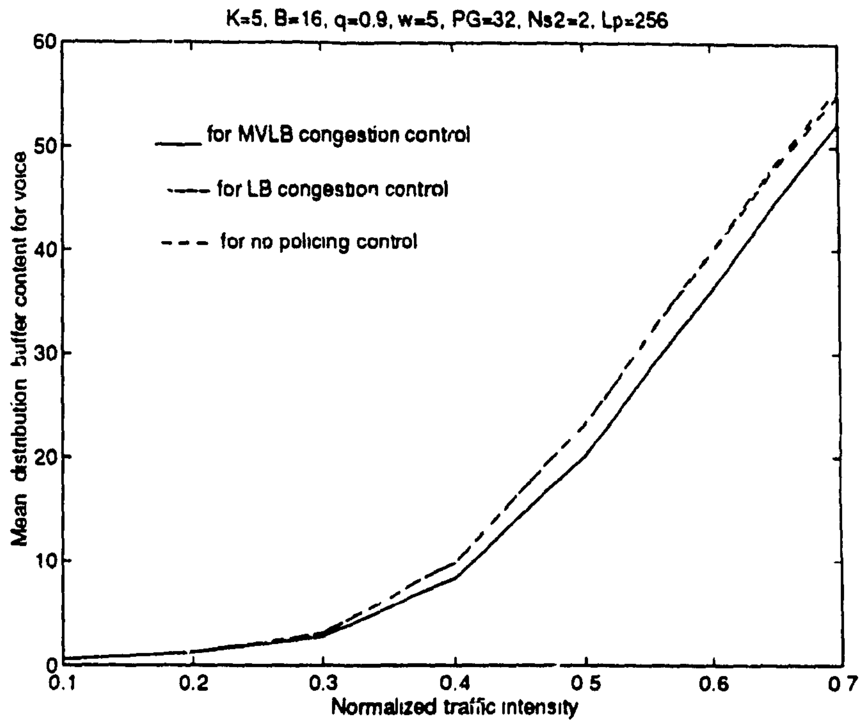


Figure A.11: Voice: Mean distribution buffer content $\bar{\Psi}$ vs normalized traffic intensity ρ

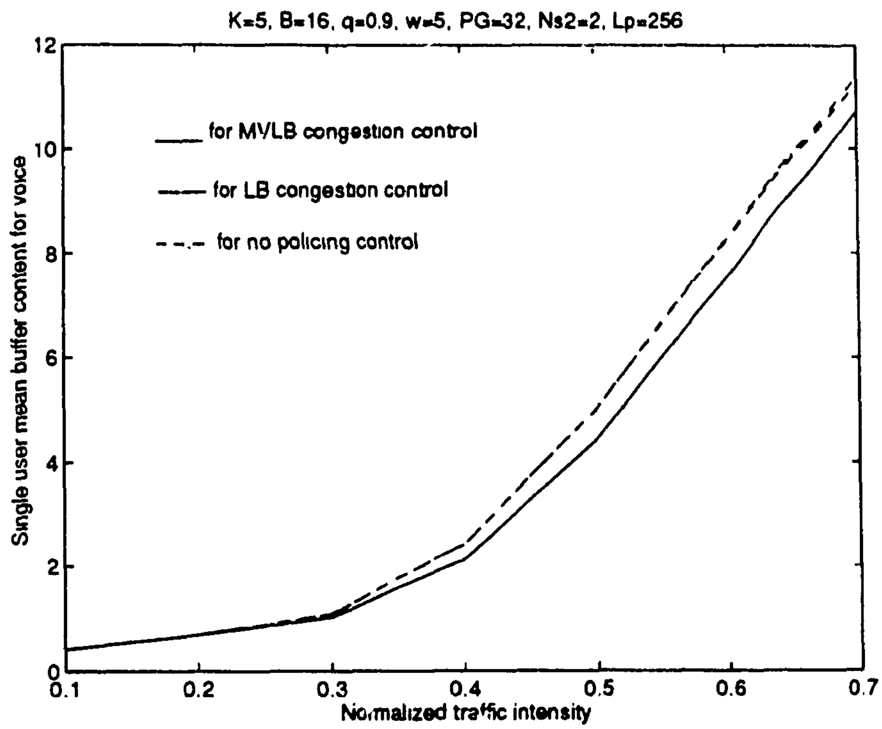


Figure A.12: Voice: Single user mean buffer content $\bar{\Phi}$ vs normalized traffic intersity ρ

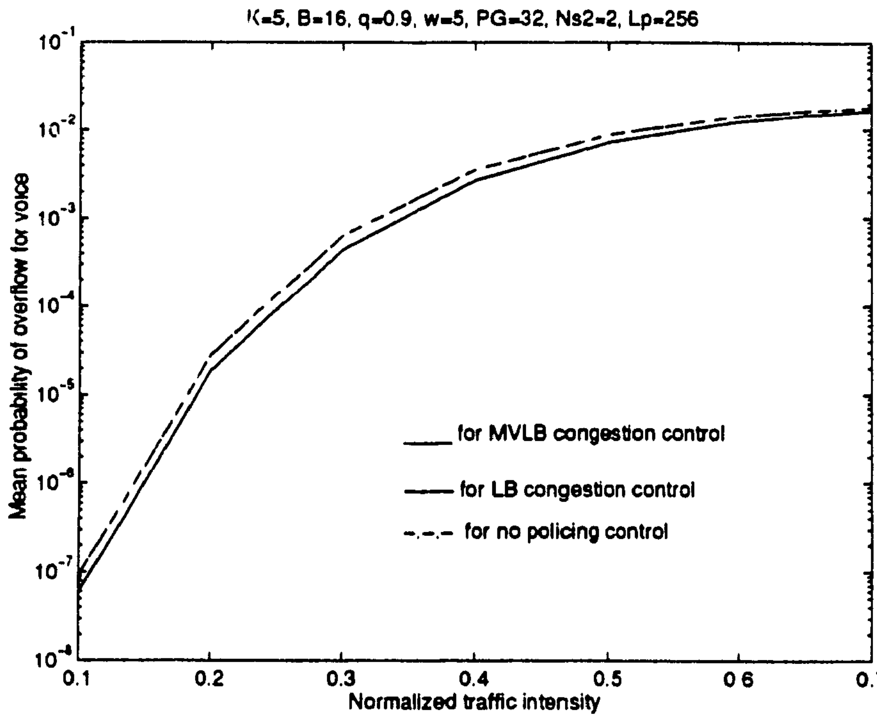


Figure A.13: Voice: Mean prob. of overflow $\overline{P_{overflow}}$ vs normalized traffic intensity ρ

Note: The results for no policing and LB congestion control are too close to distinguish.

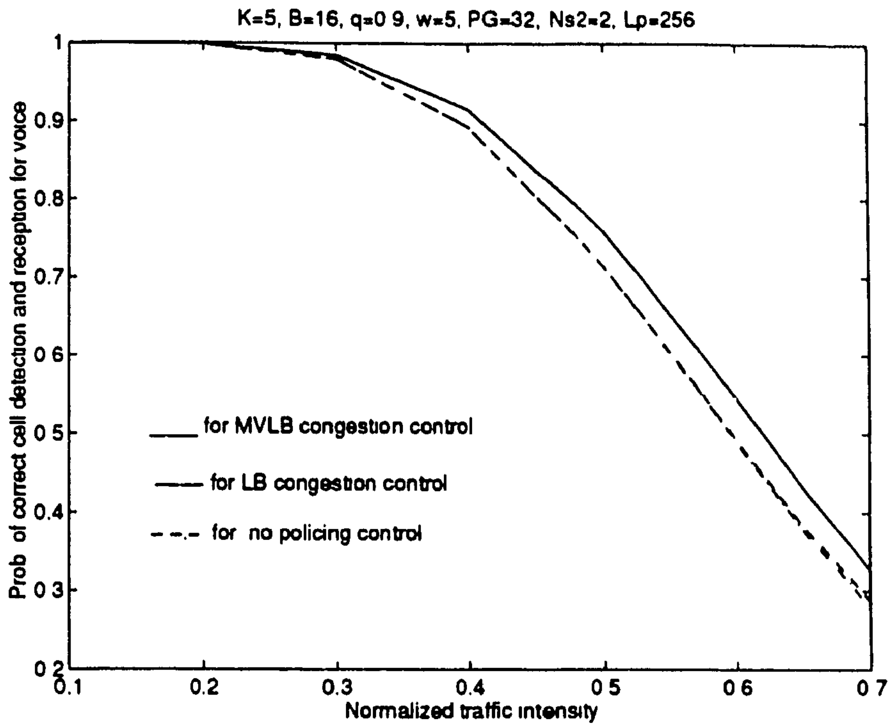


Figure A.14: Voice: Prob. of correct cell detection and reception P_d vs normalized traffic intensity ρ

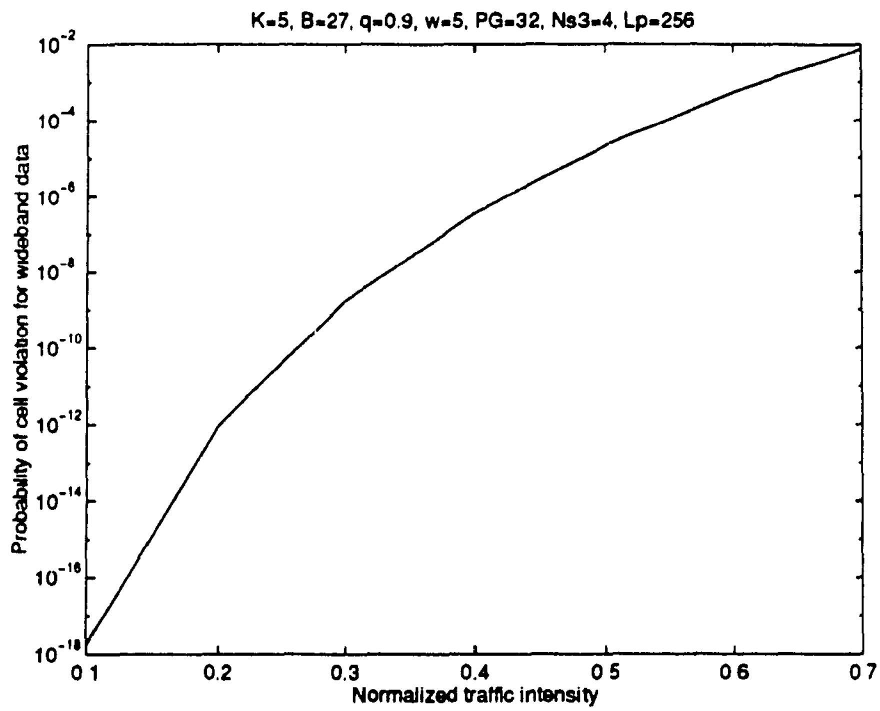


Figure A.15: Wideband data: Prob. of cell violation P_v vs normalized traffic intensity ρ

Note: Probability of cell violation for three schemes is all the same.

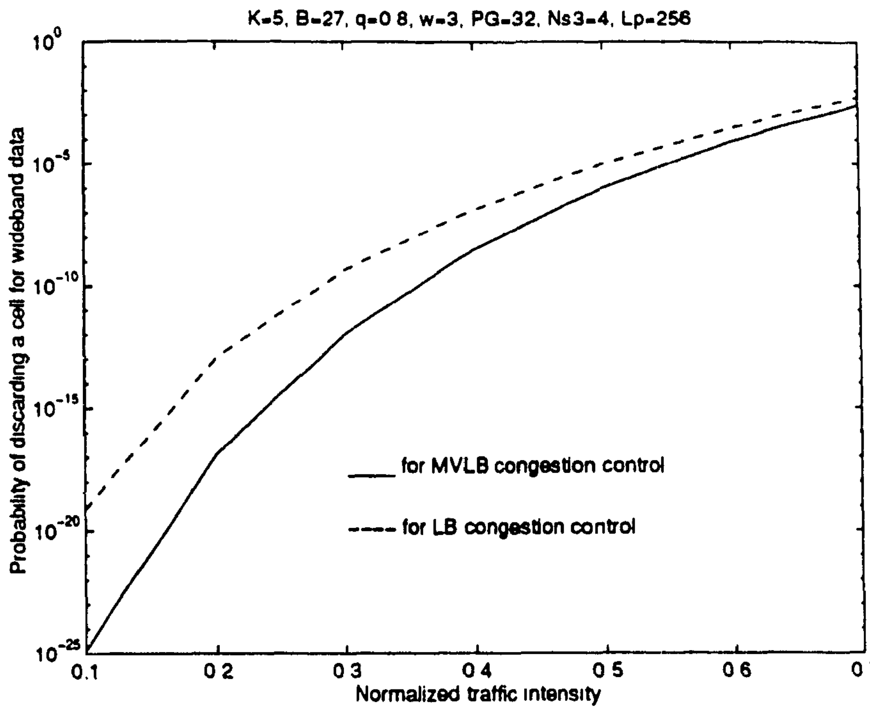


Figure A.16: Wideband data: Prob. of discarding a cell $P_{discard}$ vs normalized traffic intensity ρ

Note: There is no discarding for the no policing control.

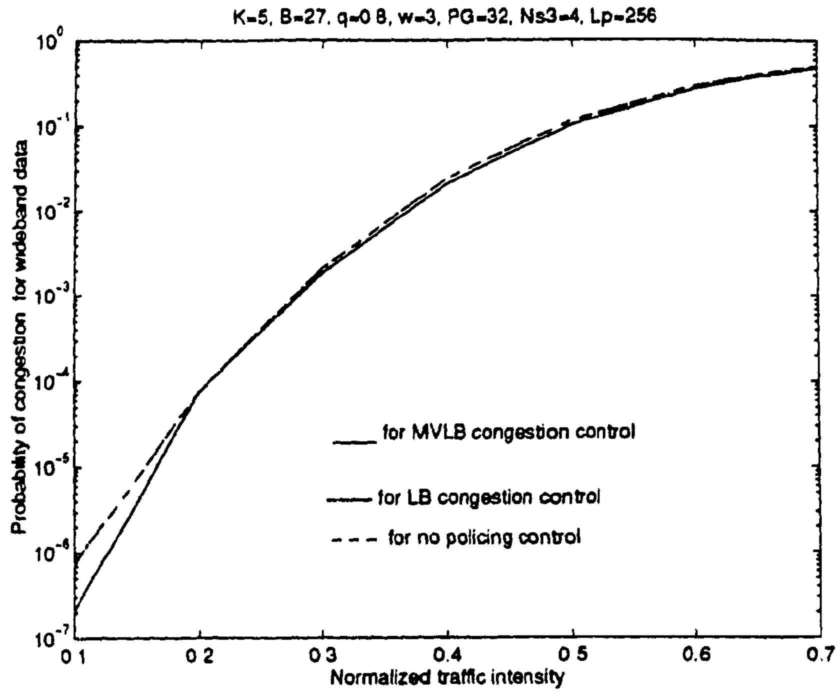


Figure A.17: Wideband data: Prob. of congestion P_{cg} vs normalized traffic intensity ρ

Note: The results for no policing and LB congestion control are too close to distinguish.

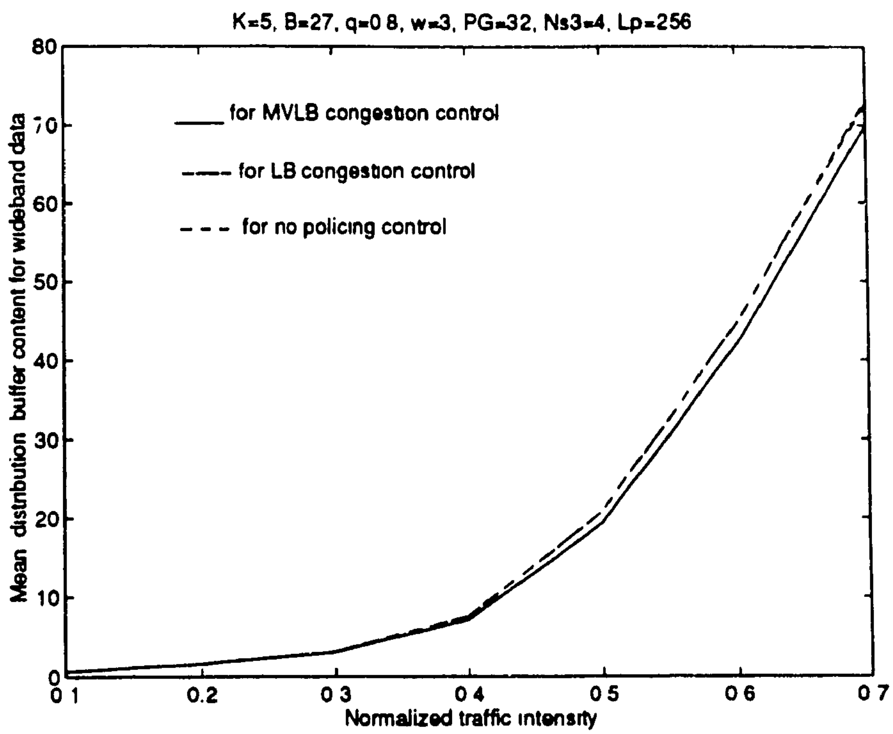


Figure A.18: Wideband data: Mean distribution buffer content $\bar{\Psi}$ vs normalized traffic intensity ρ

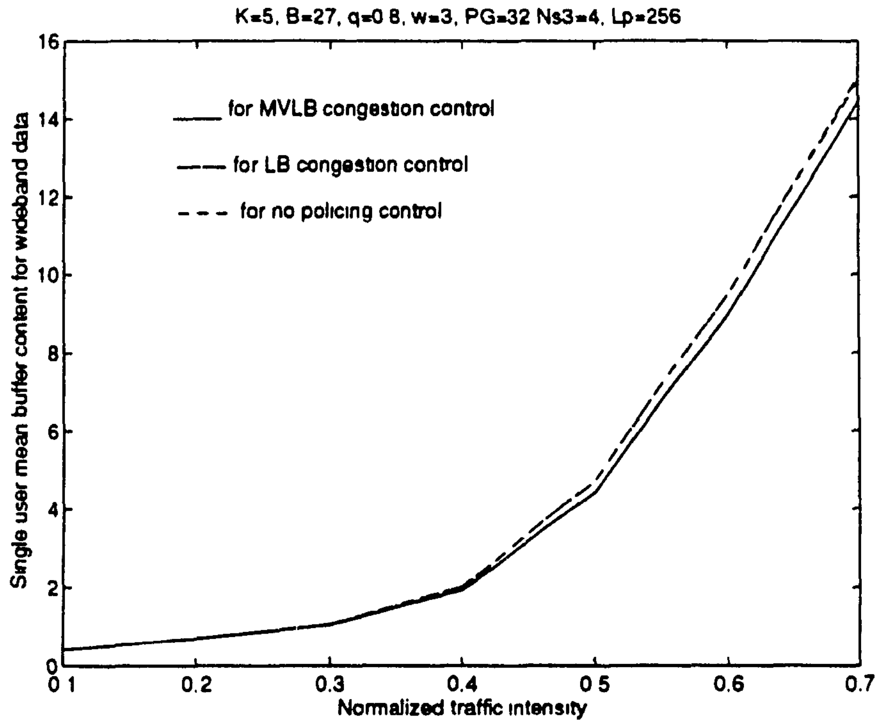


Figure A.19: Wideband data: Single user mean buffer content $\bar{\Phi}$ vs normalized traffic intersity ρ

Note: The results for no policing and LB congestion control are too close to distinguish.

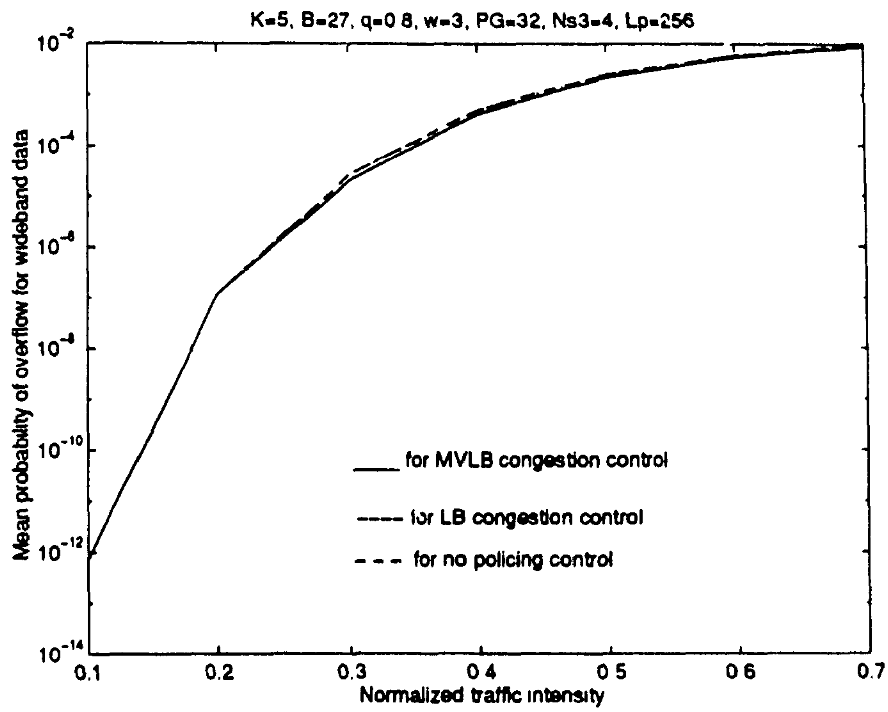


Figure A.20: Wideband data: Mean probability of overflow $\overline{P_{overflow}}$ vs normalized traffic intensity ρ

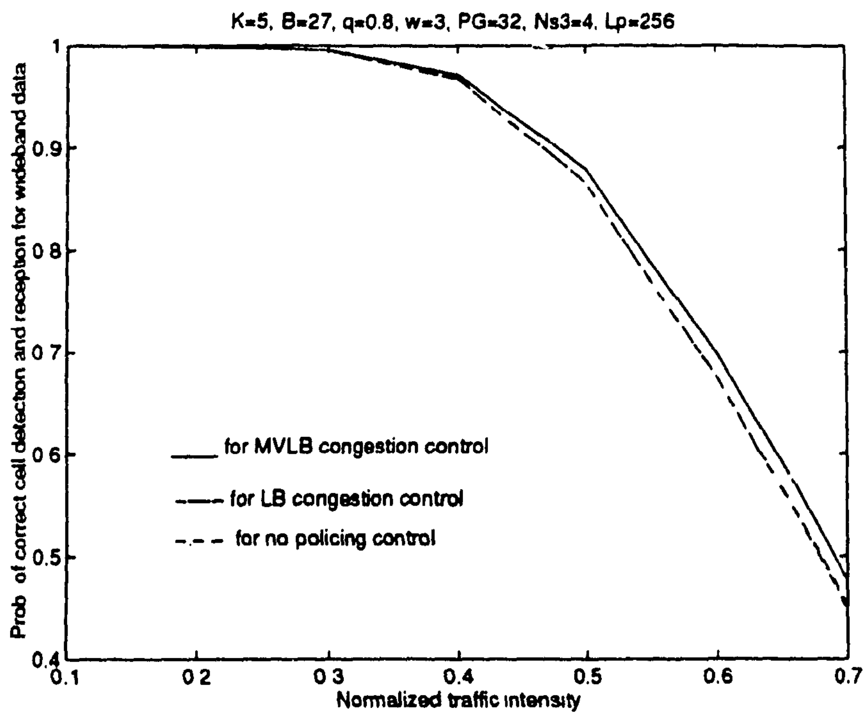


Figure A.21: Wideband data: Prob. of correct cell detection and reception P_d vs normalized traffic intensity ρ

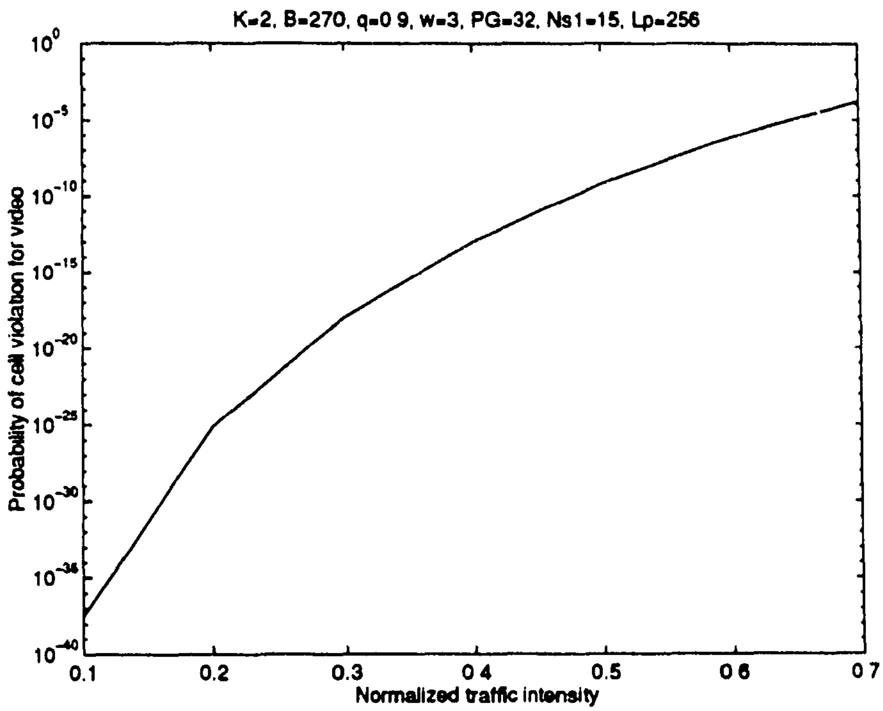


Figure A.22: Video: Prob. of cell violation P_v vs normalized traffic intensity ρ

Note: Probability of cell violation for three schemes is all the same.

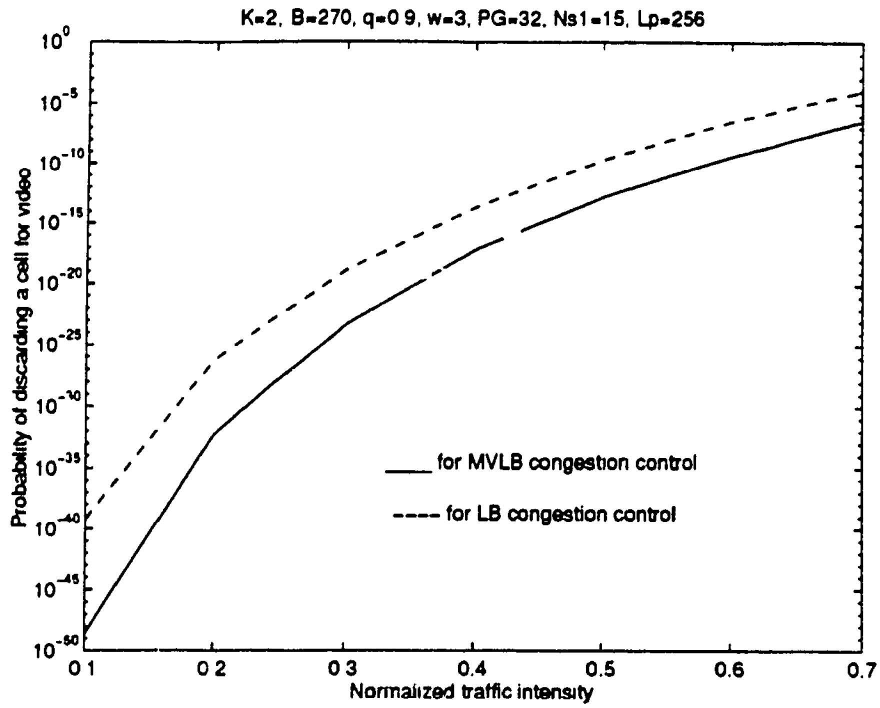


Figure A.23: Video: Prob. of discarding a cell $P_{discard}$ vs normalized traffic intensity ρ

Note: There is no discarding for the no policing control.

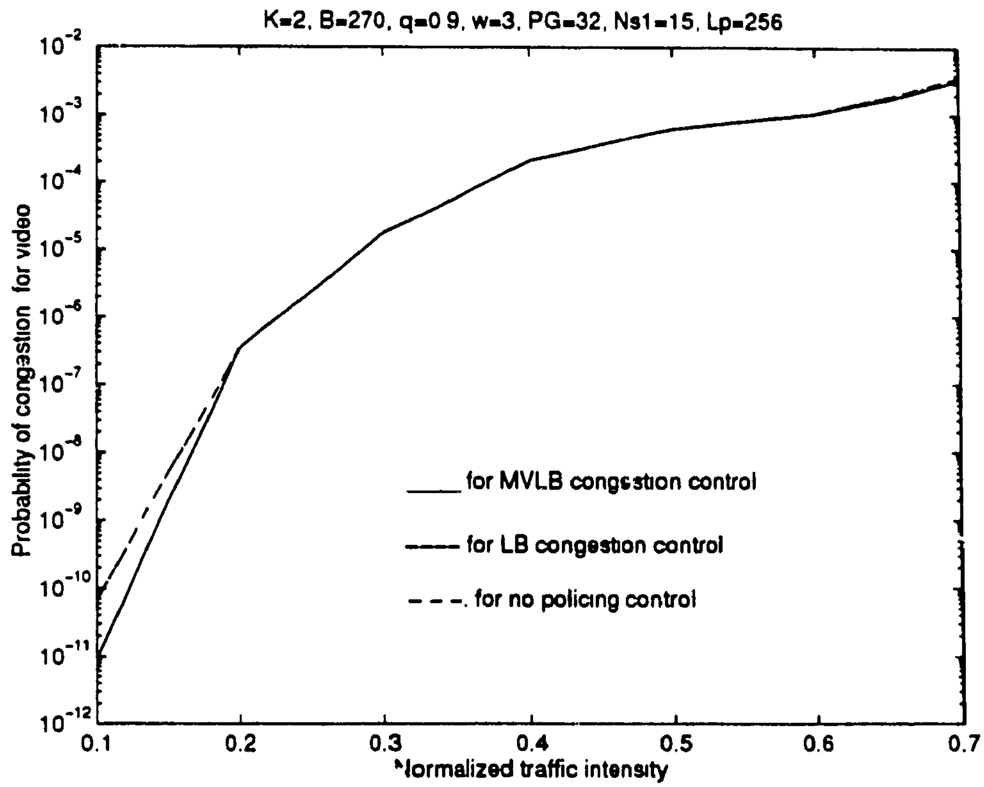


Figure A.24: Video: Prob. of congestion P_{cg} vs normalized traffic intensity ρ

Note: The results for three different congestion control schemes are too close to distinguish.

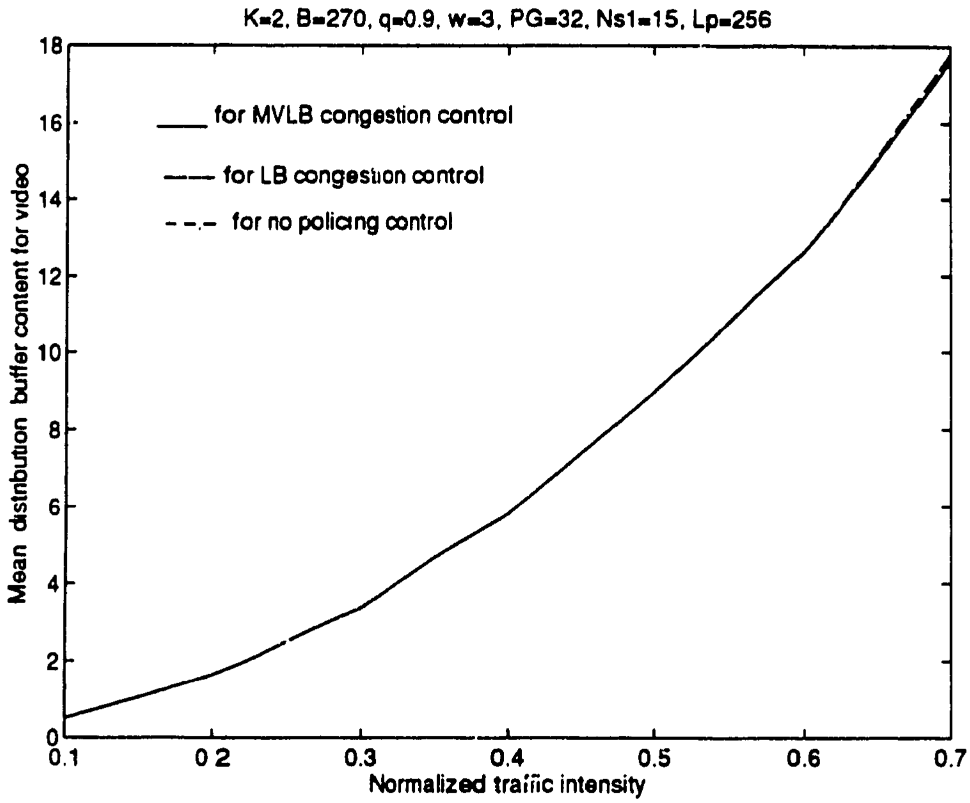


Figure A.25: Video: Mean distribution buffer content $\bar{\Psi}$ vs normalized traffic intensity ρ

Note: The results for three different congestion control schemes are too close to distinguis'

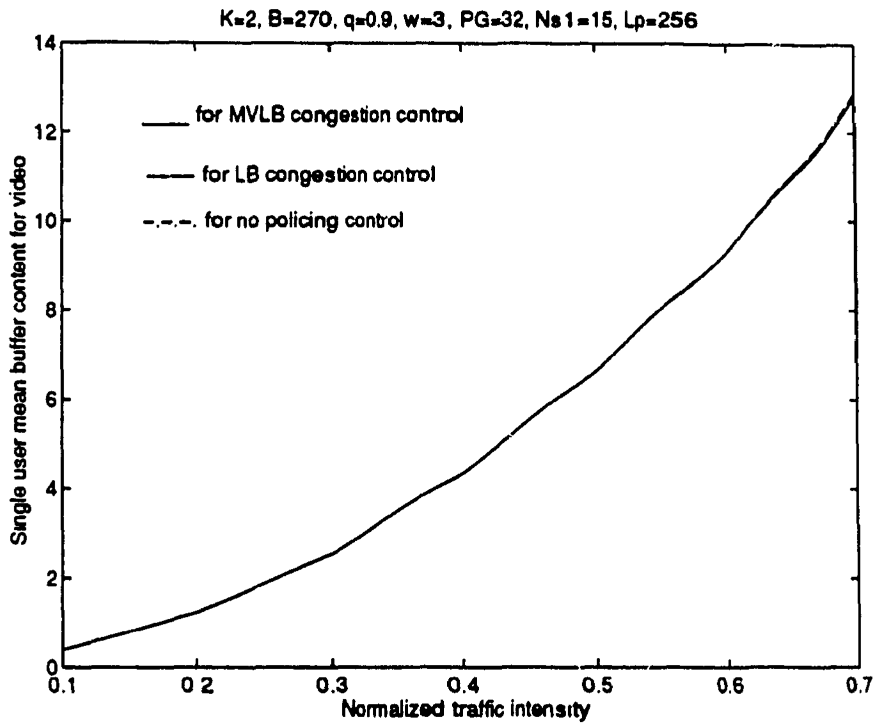


Figure A.26: Video: Single user mean buffer content $\bar{\Phi}$ vs normalized traffic intersity ρ

Note: The results for three different congestion control schemes too close to distinguish.

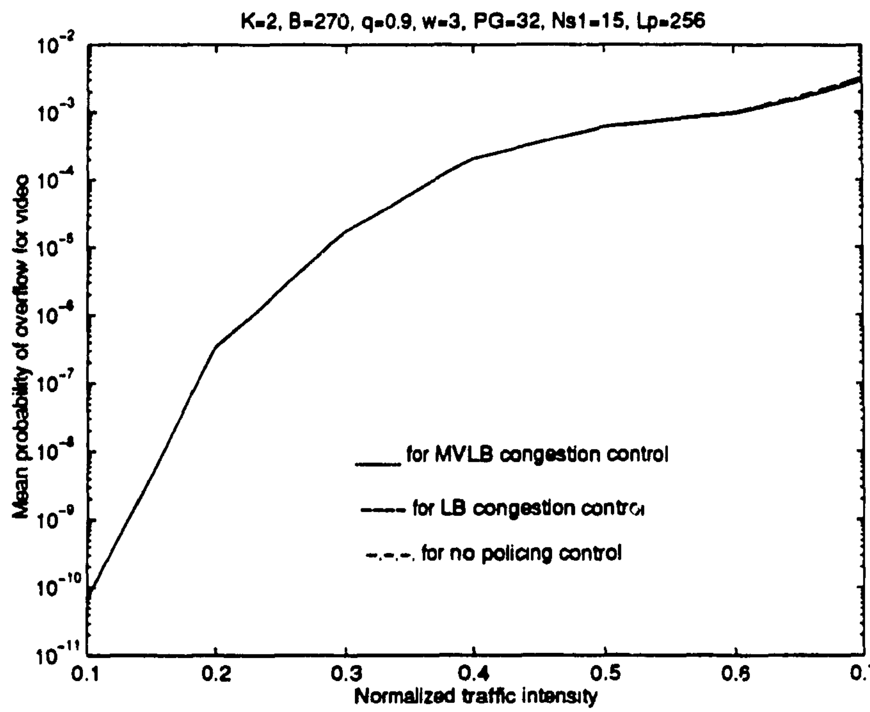


Figure A.27: Video: Mean probability of overflow $\overline{P_{overflow}}$ vs normalized traffic intensity

Note: The results for three different congestion control schemes are too close to distinguish.

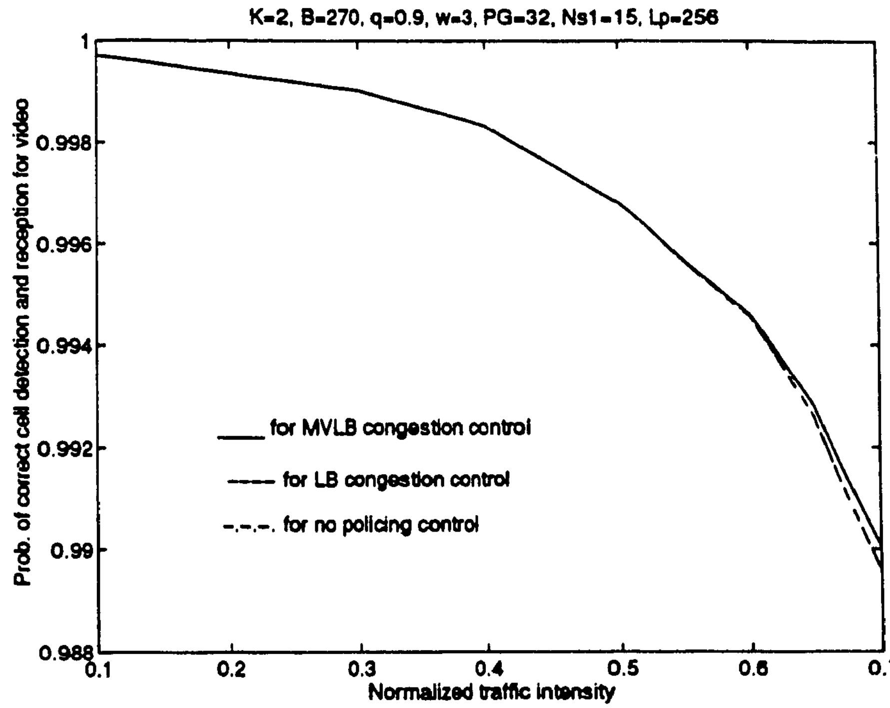


Figure A.28: Video: Prob. of correct cell detection and reception P_d vs normalized traffic intensity ρ

Note: The results for three different congestior
too close to distinguish.

Bibliography

[1] IEEE Workshop on Spread Spectrum Potential Commercial Applications, Montobello, Quebec, Canada, May 1991.

[2] N. Nakagawa and T. Hasegawa, " Spread spectrum for communication application of spread spectrum communication in Japan." *IEEE Trans.*, vol. E74, no. 5 May 1991.

[3] K. S. Gilhousen, I. M. Jakobs, R. Padovina, and L. A. Weaver, " Incresed capacity using CDMA for mobile satellite communications," *IEEE J. Select Areas Commun.*, vol. SAC-8, no. 4, pp. 503-514, May 1990.

[4] J. M. Murser and J. N. Daigle, "Derivation of asynchronous code division multiple access (CDMA) throughput," in *Local Area and Multiple Access Networks.* Computer Science Press, 1986, Ch9.

[5] M. B. Pursley, "Performance evaluation for phase coded spread spectrum multipath access communication-Part I: System analysis"' *IEEE Trans. Commun.*, vol. COM-25, no. 8, pp. 795-799. Aug. 1977.

[6] E. A. Geraniotis and M. B. Pursley, " Performance of coherent direct sequence spread spectrum communications over specular multipath fading channels," *IEEE Trans. Commun.*, vol. COM-33, no. 6, pp. 502-508, June 1985.

[7] H.Ichikawa, M.Aoki, and T.Uchiyama, "High-speed packet switching system for multimedia communication," *IEEE J. Select. Areas Commun.*, vol. SAC-5,pp. 1336-1345, otc. 1987.

[8] A.Hac and H. B. Mutlu, " Synchronous optical network and broadband ISDN protocols," *Computer*, vol. 22, no.11, pp. 26-34, Nov. 1989.

[9] T.Kamitake and T.Suda, " Evaluation of an admission controlscheme for an ATM network considering fluctuations in cell loss rate, "*IEEE GLOBECOM'89*, pp.49.4.1-49.4.7

[10] M.Murata, Y.Oie, T.Suda and H. Miyahara, " analysis of a discrete-time single-server queue with bursty inpits for traffic control in ATM networks, " *Proc. IEEE GLOBECOM '89*, pp.49.5.1-49.5.7

[11] S-Q. Li " Study of information loss in packet voice systems,"*IEEE Trans. Commun.*, vol. 37, pp. 1192-1202, Nov. 1989

[12] J.N.Daigle and J.D.Langford, "Models for analusis of packet voice communications system," *IEEE J. Selec. Areas Commun.*, vol. SAC-4, pp. 847-855, Sept. 1986.

[13] I.Ide, " Superposition of interrupter Poisson processes and its application to packetized voice multiplexers," *Proc. 12th Int. Teletraffic Congress*, Torino, Italy, 1988.

[14] H.Heffes and D.M.Lucantoni, " A Markov modulated characterization of packetized voice and data traffic and ralated statistical multiplexer performance," *IEEE J. Select. Areas Commun.*, vol. SAC-4, pp. 856-868, Sept. 1986.

[15] K.Q.Liao and L.G.Mason," A discrete-time single server queue with a two-level modulated input and its application," *Proc. IEEE GLOBECOM '89*pp. 26.1.1-26.1.6.

[16] P.Sen, B.Maglaris, N.E. Rikili, and D.Anastassiou," Models for packet switching of variable-bit-rate video sources,"*IEEE J. Select. Areas Commun.*, vol.7, pp 865-868, June 1989.

[17] J.M.Holtzman, "Coping with broadband traffic uncertainties: Statistical uncertainty, fuzziness, neural networks," *IEEE Workshop on Computer Commun.*, Data Point, CA, Oct. 1989.

[18] G.Gallassi, G.Rigolio, and L.Fratta, " ATM: Bandwidth assignment and bandwidth enforcement policies," *Proc. IEEE GLOBECOM '89*, pp. 49.6.1-49.6.6.

[19] M.Hirano and N.Watanabe, "Characteristics of a cell multiplexer for bursty ATM traffic," *Proc. IEEE ICC '89*. pp. 13.2.1-13.2.5.

[20] S.Akhtar, " Congestion control in a fast packet switching network," master thesis, Washington University, St. Louis, MD, DEc. 1987

[21] J.S.Turner, "New direction in communications (or which way to the information age?)," *IEEE Commun. Mag.*, vol. 25, pp. 8-15, Oct. 1986.

[22] M.W..Beckner, T.T. Lee and S.E. Minzer, " A protocol and prototype for broadband subscriber access to ISDN's," *Proc. ISS '87*, pp. B6.3.1-B6.3.8.

[23] I.Cidon and I.S.Gopal, " PARIS: An approach to integrated high-speed private networks," *Int. J. Digital & Analog Cabled Systems*, vol. 1,pp. 77-86, Apr.-June, 1988.

[24] A.E.Eckberg, Jr., D.T.Luan, and D.M.Lucantoni, " Meeting the challenge: Congestion and flow control strategies for broadband information transport," *Proc. IEEE GLOBECOM '89*. pp. 49.3.1-49.3.5.

[25] K.Pahlavan and M.Chase " Spread-Spectrum Multiple-Access Performance of Orthogonal Code for Indoor Radio Communications " *IEEE Trans. Commun.*, Vol.38 No.5, pp 574-577, May 1990.

[26] Per K. Enge and Dilip V. Sarwate " Spread -Spectrum Multiple-Access Performance of Orthogonal Codes: Linear Receivers " *IEEE Trans. Commun.*, Vol. Com35. No.12 Dec. 1987.

[27] A. Baiocchi, M. Crosi and M. Listant " Modeling of a Distributed Access Protocol for an ATM Satellite System: An Algorithmic Approach " *IEEE J. Select. Areas Commun.*, Vol. 9. No. 1. pp 65-75, January 1991

[28] Wen-Bin Yang and Evaggelos Geraniotis: " Admission Policies for Voice and Data Traffic in CDMA Packet Radio Networks " *IEEE J. Select. Areas Commun.*, Vol.12, No.4, May 1994 pp 654-664.

[29] Elhakeem A.K., Shilling D.L., Baier W.P. and Nakagawa M. " Code Division Multiple Access Networks I & II "*IEEE J. Select. Areas Commun.*, Vol.12, No.4, May 1994 and Vol.13, No. 5, June 1994

[30] Bohm, S., Elhakeem, A.K., Murthy, K.M., Hachichda, M., and Kadoch, M.: " System Analysis of a Movable Boundary Accessing Technique for Integrated Service Multi-beam Satellite ", *International Journal of Satellite Communications, Wiley*, Vol.12, No.3, May-June 1994 pp 299-312.

[31] S.Bohm, A.K.Elhakeem and M. Hachicha: " Performance of a Prioritized On-Board Basedband Switch for Multiservice Multibeam Satellite " *Inter. J. of Satellite Commun.* Vol. 12 pp 283-297 (1994)

[32] L.C.Palmer and L.W.White: " Demand Assignment in the ACTS LBR System " *IEEE Trans. Commun.* Vol. 38. No. 5. pp 684-692 May 1990.

[33] M. Yabusaki: " Channel Reassignment Algorithm in Multiple Carrier Hopping TDMA Systems " *IEEE Trans. Commun. Vol.*, 41. No. 2 pp 381-390 Feb. 1993.

[34] E. P. Rathgeb: " Modeling and Performance Comparison of Policing Mechanisms for ATM Networks ", *IEEE J. Select. Areas Commun.*, Vol 9. No. 3 pp 325-334 April 1991.

[35] M. Butto, E.Cavallero and A. Tonietti: " Effectiveness of the "Leaky Bucket" Policing Mechanism in ATM Networks " *IEEE J. Select. Areas Commun.*, Vol 9. No. 3 pp 335-342 April 1991.

[36] CCITT, "Draft recommendation I.121-broadband aspects of ISDN," CCITT, TD49 (PLEN), Seoul, Korea, Feb. 1988.

[37] S.E. Minzer, " Broadband ISDN and asybchronous transfer mode (ATM)," *IEEE Commun. Mag.*, vol. 27, pp.17-24, Sept. 1989.

[38] " An Overview of the Application of CDMA to Digital Cellular Systems and Personal Cellular Network," QUALCOMM Inc., May, 1992