Performance Evaluation of Admission Control Policies for High Speed Networks

A Thesis Submitted to the College of
Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in the
Department of Computational Science
University of Saskatchewan
Saskatoon, Saskatchewan

by
Ying H. Cao

May, 1994

Performance Evaluation of Admission Control Policies for High Speed Networks

Candidate: Ying H. Cao

Supervisor: Carey Williamson

Master of Science Thesis

Presented to the College of Graduate Studies and Research

University of Saskatchewan

May, 1994

ABSTRACT

The Broadband Integrated Services Digital Network (B-ISDN) is an all purpose digital network that will provide integrated access to a wide variety of applications in a flexible manner. Asynchronous Transfer Mode (ATM) is the underlying packet transport mechanism for B-ISDN that allows different types of traffic to share the available bandwidth in B-ISDN efficiently.

Admission control in B-ISDN/ATM is responsible for accepting or denying a new connection bandwidth request based on a traffic descriptor for the new connection and the current resources available at each node and link of the determined route. A connection is accepted only if, provided all other connections adhere to their bounds, the quality of service would be met. The challenge for an admission control policy is to determine the maximum set of connections that can be admitted into the network while still ensuring a controllable level of congestion.

The objectives of the thesis are to establish a performance evaluation method for admission control policies, and to use this method, called the loading region, for the quantitative evaluation of four proposed admission control policies for high speed networks. The four policies are evaluated in terms of their dependence on traffic mix, on traffic characteristics, and on class of service requirements. The results show that the loading region is an effective method for the evaluation of admission control policies. For the same input traffic models, different admission control policies allow significantly different amount of traffic into the network. Of the four admission control policies, the Joint Scheduling and Admission Control Policy is the most aggressive policy.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Carey Williamson, for his guidance, encouragement, and financial support throughout the whole duration of this thesis. Dr. Derek Eager and Dr. Carl McCrosky are also to be thanked for their help from the very beginning of their involvement as committee members. Thanks to Dr. Charbel Tannous for carefully reading my thesis and providing valuable comments.

A very special thank is reserved for my parents, my brother and his family. Without their constant love, caring, understanding and encouragements, I cannot accomplish any achievement, including this thesis.

Thanks are extended for the faculty, staff, and fellow graduate students at the Department of Computational Science for their assistance and reassurance when I needed it.

Finally, I owe a great deal to my husband, Peizhong Wang, and my son, Chenchen, who have been so understanding, encouraging, and supportive all the way. They fulfill my life with love, joy, and happiness.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

High speed digital communications networks are an enabling technology for a whole new range of multimedia services and applications. Recently, multimedia computing has emerged as a major area of research motivated by the wide range of potential applications made feasible by combining information sources such as voice, image, text, and video. Potential applications exist in education, banking, advertising, publishing, and medical imaging [32].

The emerging demand for high speed networks and multimedia services and the desire to integrate these and future services on a single flexible network have drawn much attention from researchers. The Broadband Integrated Services Digital Network (B-ISDN) is defined as an all-purpose digital network that will provide integrated access to a wide variety of applications in a flexible and cost-effective manner. The transfer mode chosen by the CCITT as the basis for B-ISDN is the Asynchronous Transfer Mode (ATM) [10, 13, 20, 32]. ATM is a high-bandwidth, low-delay, packet-switching and multiplexing technique [20]. In an ATM network, the usable network capacity can be dynamically assigned on demand.

ATM networks are suitable for both bursty and continuous-bit-rate services. Traffic is said to be bursty, as opposed to continuous-bit-rate, when the traffic generates cells at a peak bit rate for a short period of time (active period) and then enters a period of no transmission (idle period). The peak bit rate is the mean rate of the cells being generated over an active period. The mean bit rate is the mean rate of the cells being generated over an entire measurement period. The ratio of peak bit rate to mean bit rate is one way to classify the traffic type. The larger the ratio is, the burstier is the traffic. When the value of the ratio is approximately one, the traffic is called continuous-bit-rate traffic.

To support continuous-bit-rate services, ATM can operate in a deterministic multiplexing mode such that sufficient switching and transport bandwidth is guaranteed. By operating in a statistical multiplexing mode, however, ATM networks can take advantage of statistical gain among bursty services. ATM networks may be more efficient in carrying bursty services since each service may not require continuous allocation of its peak bandwidth. Hence, more services can be allowed to share the resource at the same time, in a statistical function. Statistical gain is achieved when the number of traffic sources admitted to the network is larger than the number that can be accommodated with peak bandwidth allocation (i.e., deterministic multiplexing). In other words, in statistical multiplexing, the sum of the peak bandwidths for each call is allowed to exceed the channel's capacity. Achieving any statistical gain, of course, introduces a non-zero probability of network congestion.

1.1 Congestion Control

Network congestion occurs when various sources compete for the network resources and these resources are not sufficient to meet the demands [10]. A network in a congested state exhibits poor performance due to increased queueing delay, buffer overflow, packet loss, and packet retransmission. Network congestion can happen at the cell level and at the call level. Cell level congestion usually occurs at a network node due to poor buffer management, as a result of congestion from higher layers, or when long bursts of cells lead to buffer overflow. At the call level, congestion can happen when logical channels request bandwidth that cannot be supported, when a certain call exceeds its reserved bandwidth limits, or when the network admits more calls than the links can handle. The goal of a congestion control algorithm is to regulate traffic flow in the network to prevent, reduce, or control the level of congestion.

Congestion control in ATM is difficult because of the high link speeds involved and the diverse characteristics of the traffic that an ATM based network is expected to support [13]. The congestion control strategies can be classified into two groups: reactive and preventive. Reactive approaches rely on the network to provide feedback signals, either explicit or implicit, indicating the presence or absence of congestion. Sources then dynamically adjust the rate of transmission into the network [31]. Preventive approaches rely on reservation of network resources before the sources can access the network, and these sources are required to remain within their allocated resources [18]. Reactive strategies are not suitable for an ATM network because of the increased propagation delay-bandwidth product in high speed links. In a 1 Gb/s link with a propagation delay of 20 ms for a cross-continental distance of 6000 km, there can be up to 20 Mb in transit at a time. In a reactive control system where a congested network sends choke packets to the senders to stop or slow transmission, up to 20 Mb of data may have to be retransmitted by the time that the choke packet reaches the sender. Large retransmissions can cause severe buffer management problems, which make reactive control strategies inefficient. By contrast, preventive control strategies rely on advance reservation of resources. They provide a fair allocation of bandwidth by requiring that each connection's traffic flow remain within pre-specified bounds appropriate for the supported service. A connection would only be accepted if, provided all connections adhere to their bounds, the transport quality of service (QOS) would be met.

1.2 Admission Control

Admission control is responsible for accepting or denying a connection request based on descriptions of the traffic flow and the current resources (e.g., bandwidth, buffers, and processing) available at each node and link of the determined route [32]. Admission control is based in part on measurements of current network traffic, which might be the result of observation of the actual cell arrival process, or be deduced from the declared traffic descriptors of all connections in progress.

The call acceptance decision made by an admission control policy depends greatly on the declared descriptor values of the call in question. If a new call is accepted, the network resources are reserved for the call until a request for release. The traffic characteristics and required class of service will have great influence on the decision of call acceptance. A certain amount of bandwidth may be sufficient to support the call with a certain class of service, while the same bandwidth may be insufficient to support the same call if the class of service is more stringent [10].

A conservative admission control policy allows relatively low loading on the network to minimize the probability of cell level congestion. A more aggressive policy prefers to accept more calls, reducing the probability of a call being blocked, but may result in congestion. The tradeoff between the probability of blocking a call and the probability of congestion has to be balanced when the acceptance decision is made. The challenge for an admission control policy is to determine the maximum set of arriving calls that can be admitted to the network while still ensuring a manageable level of congestion within the network.

1.3 Thesis Objectives

The objectives of this thesis are to establish a performance evaluation method for admission control policies, and to use this method to evaluate different admission control policies for high speed networks. For the same arriving call, different admission control policies may make different decisions, depending on the policy's operating rule. Also, the performance of an admission control policy varies depending on the loading condition within the network.

To measure the efficiency of an admission control policy, the concept of a *Loading Region* is introduced in the thesis. The loading region measures the number of calls of heterogeneous traffic admitted to the network. The shape and size of a loading region are highly dependent upon the admission control policies, the traffic characteristics, and the quality of service requirements. The Average Expansion Degree is defined in the thesis as a mathematical measurement of the shape and size of a loading region.

This thesis evaluates four different admission control policies by simulation, using the loading region framework as the primary metric for performance evaluation. The evaluation of each policy focuses on its dependence on the traffic mix, the traffic characteristics, and the quality of service requirements. The comparisons of the four admission control policies are also presented in the thesis to illustrate the aggressiveness of each admission control policy, and the sensitivity of each admission control policy to the quality of service constraints and the traffic characteristics.

1.4 Thesis Outline

The remainder of the thesis is organized as follows. Chapter 2 presents background information on ATM-based networks, the multiplexing modes in ATM networks, and the integrated traffic types in ATM networks. Chapter 3 focuses on admission control policies. The principle of admission control and a survey of admission control policies are presented. Chapter 4 presents the experimental methodology, including the network and traffic models used for the simulation, the scheduling algorithms, and the admission control policies evaluated in the thesis. Chapter 5 discusses the statistical behaviour of scheduling algorithms. Chapter 6 presents the performance evaluation results for the four different admission control policies. The conclusions and directions for future research are summarized in Chapter 7.

Chapter 2

Background

In this chapter, a brief discussion of ATM, multiplexing, and multimedia traffic is presented as background information for the thesis.

2.1 Asynchronous Transfer Mode (ATM)

The Asynchronous Transfer Mode (ATM) is a connection-oriented packet transport scheme that is considered by the CCITT and T1 standards committees as the preferred packet transport method for the Broadband Integrated Services Digital Network (B-ISDN) [25]. The ATM technique is versatile and flexible enough to provide the means to transport, at broadband rates, the traffic generated by a wide range of multimedia services. Multimedia services require some combination of text, image, voice, and video storage and communication, and have potential applications in many areas such as education, banking, advertising, publishing, and medical imaging [32].

ATM provides the flexibility to integrate the transport of services with a wide range of bandwidth requirements by assigning fixed-length cells to virtual connections on an as-needed basis [8]. The bandwidth in ATM is divided into small fixed-size packets called cells. A cell consists of a 5-byte header and 48 bytes of data. The header contains the routing and channel identification information. A channel supports a single end-to-end user connection. Since the cells belonging to a channel may share the physical transmission medium with other user connections, the channel is called a virtual channel, and all cells belonging to a channel are identified by a header field called the Virtual Channel Identifier (VCI) [25]. No mapping of time slots to virtual channels at call establishment is performed. Hence, ATM can be responsive to short-term fluctuations and long-term changes in service demand [20].

The virtual channel provides for the dynamic assignment of bandwidth to each service. In other words, the bandwidth required by an application is allocated on demand. This concept of a virtual channel also means that services with highly variable bandwidth requirements need not waste bandwidth by reserving a fixed size channel equal to their maximum transmission rate. For those services that do require guaranteed bandwidth, it is still possible to implement in ATM the notion of reserved bandwidth [25].

ATM also provides the potential to improve bandwidth efficiency through the statistical multiplexing of variable bit rate, or bursty, traffic streams. Gains due to statistical multiplexing, however, may introduce potential cell loss. Also, the buffering that must occur at each multiplexing and switching point in the network may result in unpredictable end-to-end cell delay.

2.2 Multiplexing Mode in ATM

ATM multiplexes or switches cells from different connections onto a common link, independent of the connections' bit rate or burstiness. ATM connections can be allocated resources based on either deterministic multiplexing or statistical multiplexing. Figure 2.1 shows an example of a general ATM multiplexer. This multiplexer assumes N input channels, a peak bit rate P_i on channel i, a buffer for each input link, and an output link capacity L.

If the connections are multiplexed deterministically, then sufficient resources are allocated to each connection to accommodate its peak bit rate. This approach is similar to bandwidth reservation in synchronous transfer mode, with the added flexibility of being able to reserve any peak bit rate required, rather than just multiples of the base channel rate. In deterministic multiplexing mode, the sum of the input peak bit

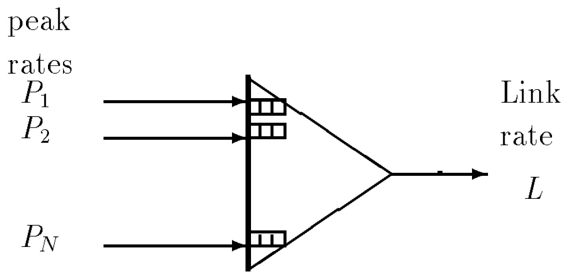


Figure 2.1: ATM Multiplexer

rates is never allowed to exceed the output capacity. That is, the constraint

$$\sum_{i=1}^{N} P_i \leq L \tag{2.1}$$

must be satisfied at all times with deterministic multiplexing.

With deterministic multiplexing, cell level congestion is less of an issue and each connection experiences negligible cell loss and cell delay variation. The major disadvantage of deterministic multiplexing is that efficient bandwidth utilization cannot be achieved when a large proportion of the traffic is bursty. For example, channel utilizations over 45% are rarely achievable. (See Section 5.4 for a brief discussion based on the observations from the simulation for the thesis.)

To overcome the bandwidth inefficiency of deterministic multiplexing, statistical multiplexing mode can be used. Statistical multiplexing takes advantage of the variable information transfer rates generated by many connections to implicitly share resources on the assumption that each connection only requires its peak bit rate for a small fraction of its time. Whenever statistical multiplexing is used, the sum of the connection peak bit rates is allowed to exceed the output link capacity. That is,

$$\sum_{i=1}^{N} P_i \geq L \tag{2.2}$$

Statistical gain is achieved when the number of traffic sources allowed into the network is larger with statistical multiplexing than with deterministic multiplexing.

Hence, the statistical gain directly depends on the bandwidth utilization and traffic characteristics of the input channels. ATM allows for statistical multiplexing at the cell level. For some high-bandwidth bursty services statistical multiplexing has the potential for substantial gains in bandwidth.

With statistical multiplexing, there is a non-zero probability that at some point in time the offered load to a multiplexer will exceed the output link capacity. The provision of buffers can accommodate short overloads. If the condition persists, however, cell-level congestion will occur resulting in increased cell delay, potential cell loss, and increased cell retransmission.

2.3 Multimedia Traffic Characteristics

In B-ISDN, a wide mixture of traffic types must be handled by the same ATM transport network. As shown in Figure 2.2, the bandwidth requirements for B-ISDN services may range from a few kilobits/second for low speed data, to hundreds of megabits/second for video traffic. In addition, traffic streams vary from being reasonably continuous like voice traffic, to being extremely bursty like image and file transfer. The required quality of service (QOS), in terms of cell loss and cell delay, will also vary considerably depending on the application active on each connection. Furthermore, the traffic streams generated by these applications will be difficult to characterize because insufficient information is known about their statistical behaviour. Generally, multimedia traffic can be grouped into two categories: real-time and non-real-time traffic. Voice and video traffic belong to the real-time traffic category, which require stringent constraints on time delay and delay variability, but might be able to tolerate a low cell loss rate depending on the application. Non-real-time traffic, such as computer generated data, is sensitive to cell loss but can tolerate a certain amount of time delay.

Video traffic can be divided into two groups: real-time video and image retrieval. Real-time video, like teleconferencing and distribution video for entertainment, requires bit rates of up to several hundred megabits per second [18]. This type of traffic

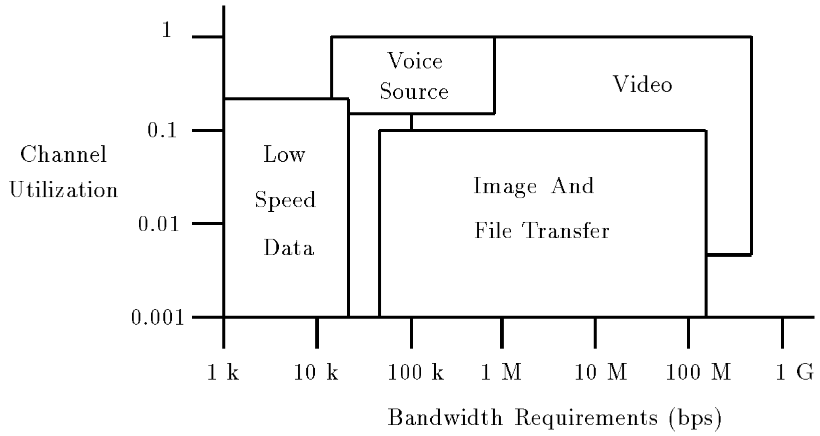


Figure 2.2: Multimedia Traffic Characteristics [7]

is extremely delay sensitive.

Image retrieval applications include medical imaging, computer-aided design, and videotext services [18]. By accessing image banks, these applications generate a significant amount of traffic. Depending on the resolution and coding techniques, each image is coded in a number of bits ranging from hundreds of kilobits to tens of megabits [26]. Image retrieval applications require moderate constraints on delay and are highly loss-sensitive, compared to real-time video.

Voice traffic is delay sensitive but has a much lower bit rate than video traffic, as shown in Figure 2.2. The delay limit for voice traffic is considerably longer (tens of milliseconds) compared to the one for video (a few milliseconds). Voice traffic can also tolerate a certain degree of packet loss (acceptable cell loss is 10^{-4}). Recent research shows that digitized voice traffic can suffer a packet loss probability up to 2% without hurting the packetized voice quality [14].

Data traffic on broadband networks is probably the least predictable type of traffic compared to video and voice. The statistical behaviour of data traffic depends largely on the types of applications and on flow control policies. Data applications can tolerate a certain amount of time delay, but are usually loss sensitive. The retransmission of lost packets is possible because of the delay tolerance, but retransmission should be minimized as much as possible to achieve better network bandwidth utilization.

2.4 Multimedia Traffic Models

The study of any component of an ATM network requires a workload model to examine the performance of designs. Estimates of types of traffic and the traffic characteristics of potential services are one way to obtain a traffic model for use in these studies.

Burstiness and correlation between two successive packet arrivals are the characteristics of many multimedia traffic sources, and will have a significant impact on node performance [32]. Data-oriented applications such as image and file transfer generate traffic flows that move from a burst state, offering a continuous stream of packets at peak bit rate, to an idle state. Packet video sources also generate a stream of correlated packet arrivals at a rate that may vary with time according to the coding scheme and motion in the picture [32].

2.4.1 General ATM Traffic Models

Statistically, ATM traffic sources can be characterized at three levels: the call level, the burst level, and the cell arrival level as shown in Figure 2.3.

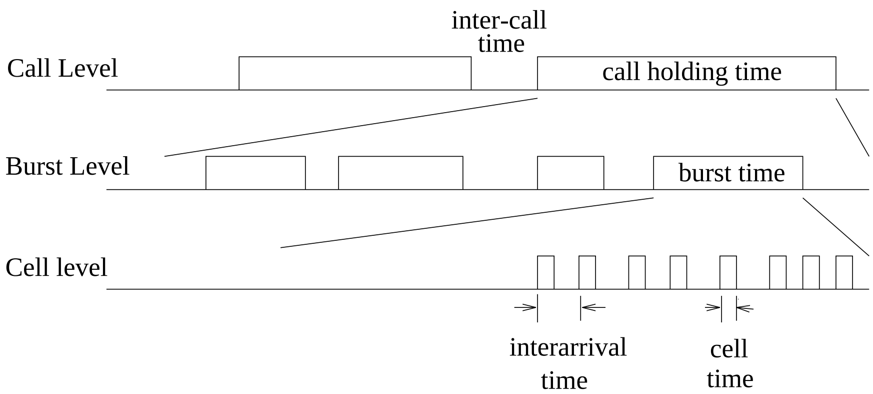


Figure 2.3: ATM Traffic Levels [30]

The call holding time, at the call level, is the time between when the call is set up and when the call is released. The burst time, at the burst level, is defined as the duration that a single piece of information (a video frame or a data file) is transmitted. The cell time, at the cell level, is the time that it takes for a single cell to be transmitted on the network. Since cells are fixed size, each cell transmission time is constant.

A two-state Markovian representation has been proposed by several researchers [25, 32] to model bursty traffic, at both the burst level and the cell level. In the two-state Markov model, one state represents the active period and the other represents the idle period. Figure 2.4 gives the three parameters of such a model: the peak bit rate (P), the mean bit rate (A), and the average burst length (B). The value of A/P denotes the mean-to-peak ratio of the traffic sources. The burstier the source is, the smaller the value is. The unit time in this model is the time to transmit one cell on the outgoing link, which depends on the speed of the outgoing link.

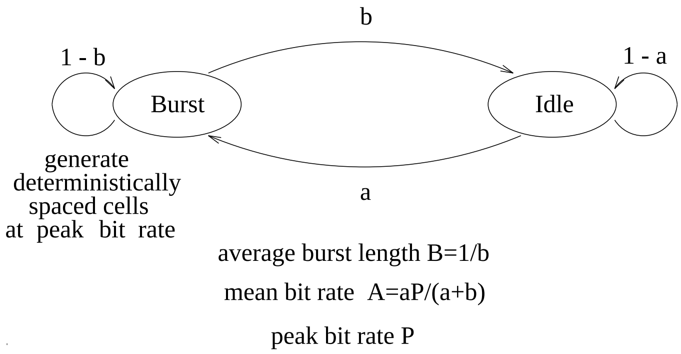


Figure 2.4: Two-state Markov ATM Traffic Flow Model [31]

2.4.2 Video Traffic Models

The models for video traffic are more complex and more dependent on particular coding methods. Several models have been introduced in the literature to characterize video traffic. In this section, only three models will be described.

1. **Two-State Model**

   Hyman [15] introduced a two-state video traffic model with three parameters, which differs from the two-state Markovian model mentioned in the last section. In Hyman's model, a single video call is represented as a periodic random process that is characterized by:

   - Fixed frame duration: $F = 62.5ms$

- Peak bit rate: $P = 10Mb/s$
- Average active period: $C = 25ms$

During the active period, the source emits cells at a rate of P. The active period, C_{active}, is a random variable uniformly distributed between $C_{min} = 10$ ms and $C_{max} = 40$ ms. At the end of the active period, the source stops emitting cells for the duration of the frame $(F - C_{active})$. The mean bit rate of the source is

$$A = \frac{C}{F} * P = 4Mb/s \tag{2.3}$$

The multiplexing of K video calls is done by uniformly distributing the frame start time at intervals of F/K ms.

2. **Autoregressive Model**

Maglaris [19] introduced a Markov process model developed from experimental measurements of motion video. The bit rate $\lambda(n)$ of a single video source during the nth frame is modeled by a first-order autoregressive Markov process generated by the recursive relation:

$$\lambda(n) = a\lambda(n-1) + bw(n) \tag{2.4}$$

where $w(n)$ is a sequence of independent Gaussian random variables and a and b are constants. $w(n)$ has a mean η and a variance 1. If $|a| < 1$ then the process achieves steady state for large n. The steady-state average $E(\lambda)$ and discrete autocovariance of the bit rate $X(n)$ are given by:

$$E(\lambda) = \frac{b}{1-a} * \eta \tag{2.5}$$

and

$$X(n) = \frac{b^2}{1-a^2} * a^n * (1-\eta^2) \qquad n \geq 0 \tag{2.6}$$

By matching the above two equations with measured data the values $a \simeq 0.8781$, $b \simeq 0.1108$, and $\eta = 0.572$ are obtained [19]. Two other parameters, the number of frames per second and the number of pixels in each frame, have to be given when this model is used to generate video traffic.

3. **Gamma Distribution Model**

Heyman [12] claims that for video teleconference traffic, the number of cells per frame follows a gamma (or negative binomial) distribution. His observation is based on a statistical analysis of teleconference data (which had very little camera motion, scene changes, or cuts), and is unlike previous studies that assume a normal distribution.

The probability function for the negative binomial distribution is given by

$$f_n = \begin{pmatrix} n+r-1 \\ r-1 \end{pmatrix} p^r q^n \tag{2.7}$$

for n = 0, 1, 2, $\cdots$, ∞ , where $0 < p < 1$, $q = 1 - p$, and $r > 0$. The parameter r need not be an integer. The mean and variance of this distribution are:

$$m = \frac{r(1-p)}{p} \tag{2.8}$$

$$u = \frac{r(1-p)}{p^2} \tag{2.9}$$

For the measured data, the mean and variance are $m = 130.3067$ and $u = 5536.873$; the method of moments produces the estimates

$$\hat{p} = 0.02353 \qquad and \qquad \hat{r} = 3.140 \tag{2.10}$$

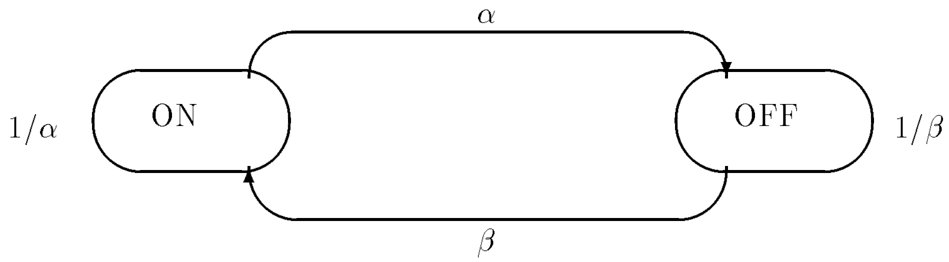
Since the Gamma distribution is the continuous version of the negative binomial distribution, this model can also be described by using the Gamma distribution (see literature [12] for details).

The frame duration is fixed at 40ms (25 frames per second).

2.4.3 Voice Traffic Models

The packet arrival process from a single voice source reveals strong correlations and dependencies among the interarrival times of successive packets [11]. Experimental results have shown that a single voice source alternates between two phases, the active or talking spurt (ON) period and the idle or silent (OFF) period [11]. This process describes a two-way speech conversation where voice packets are generated during the active period and no packets are generated during the idle period.

To model this ON/OFF periodic process requires the examination of not only the packet generation process during the ON period, but also the characteristics of both the ON and OFF periods. The most common method used is the continuous-time Markov chain approach. As shown in Figure 2.6, one state represents the ON period while the other represents the OFF period, where the holding time in each state is exponentially distributed with means $1/\alpha$ and $1/\beta$, respectively. The transition rates from the ON to OFF state and from OFF to ON state are α and β, respectively.



Holding-Time Distribution: Exponential

Arrival Process: Poisson

Figure 2.5: Two-state Continuous-time Markov Chain Model

The packet arrivals follow a Poisson arrival process.

2.4.4 Data Traffic Models

Data traffic models for ATM networks are neither clearly described nor introduced in the literature. Some use the Poisson process to characterize the statistical behaviour of a data stream. Alternatively, a two-state Markov chain is used to model data traffic [25].

For the Poisson process model, the data traffic is described by a Poisson source with a mean bit rate of [15] :

$$A = 1Mb/s$$

For the two-state Markov chain model, the model has an exponentially distributed idle period with mean $1/a$ and the number of packets in a burst is geometrically distributed with mean $1/b$. During a burst, packets are transmitted at the peak rate, giving the peak bit rate (P), mean bit rate (A), and mean burst length (B), as shown in Figure 2.5.

2.5 Summary

ATM networks will provide the means to transport, at broadband rates, the traffic generated by a wide range of multimedia services. ATM also provides potential gains in bandwidth efficiency by statistically multiplexing bursty traffic. This chapter presented background information on ATM networks, ATM multiplexing mode, multimedia traffic characteristics, and a survey of multimedia traffic models.

To control the traffic flow in ATM efficiently, admission control policies are used to prevent network congestion. The next chapter discusses the admission control techniques proposed for high speed networks.

Chapter 3

Admission Control

The call level admission control procedure provides the main method of controlling the traffic flow within the network. When a request for a new call is received, the call admission procedure makes the decision to either accept or deny the call request. A call is accepted only if, for all calls in the network at the time, the transport QOS will not be violated. If the call is accepted, resources are reserved until a request for release. The two key factors that influence the call acceptance decision are the traffic characteristics (e.g., a traffic descriptor that specifies the bandwidth required for the call), and the required class of service (e.g., a descriptor that specifies the tolerable levels of delay and loss for cell delivery). A certain amount of bandwidth may be sufficient to support the call with a certain class of service, while the same bandwidth may be insufficient to support the same call if the class of service is more stringent [10].

3.1 Admission Control Principle

The admission control function is invoked for each link on the route between the originating point of a call in an ATM network and the destination [23]. Admission control decides for each link on the route whether a request for a new connection will be accepted or not. A new call will be accepted only if the QOS estimate, based on a set of traffic parameters for the new connection and a corresponding set for established traffic on the link, is within an acceptable level. The new connection is established and allowed access to the network only if it has been accepted on each link on the investigated route.

The complexity of heterogeneous traffic in an ATM network makes the admission acceptance decision a very challenging one. Some experimental work shows that under certain conditions, there can be a strong dependence of performance and node buffer requirements on the burst length distribution, as well as the average burst length itself [32]. Since a connection's exact traffic characteristics are very difficult to predict, admission control decisions may be based on simple approximating or "upper-bounding" ATM traffic models that include only the key traffic parameters that significantly affect performance, such as the peak bit rate, the mean bit rate, and the average burst length. However, as a connection is followed along its route through the network, its flow characteristics could change, depending on the degree of interference due to buffering at each node through which it is routed [9, 32]. Golestani [9] showed that well-regulated and smoothed packet streams at the edge of a network could become clustered together and tend to form longer and longer bursts inside the network, destroying the original smoothness and increasing the chance of buffer overflow [9]. Golestani proposed a queueing mechanism that preserves the smoothness of a flow along a path. Therefore, the assumption that an individual connection's traffic flow is not altered appreciably by any network node is reasonable and can also make the admission control procedure easier to design.

The multiplexing mode has a great impact on admission control policies. In deterministic multiplexing mode, the decision on the admission of new calls can be based purely on the peak bandwidth required by the call and the available bandwidth in the network at the time of call arrival. In statistical multiplexing mode, the call acceptance decision for a newly arriving call is more complicated because the peak bandwidth requirements are allowed to exceed the output link capacity. To determine the bandwidth availability, QOS parameters must be monitored at each ATM multiplexer and switch. Since ATM networks carry a heterogeneous mix of traffic, not only long-term delay and loss at the multiplexer will be used, but the dynamic change in traffic conditions due to the bursty sources has to be taken into consideration.

3.2 Admission Control Strategies

There has been much research into admission control policies for ATM networks with heterogeneous traffic. The admission control policies that will be discussed in this document can be classified into two groups: schedulable-region-based and quality-estimate-based policies.

3.2.1 Schedulable-region-based Admission Control Policies

The task of a scheduler is to determine in what order the queued cells of different traffic classes get transmitted into the network. Admission control policies focus on the acceptance of incoming calls, and the scheduler, a cell level traffic control strategy, controls the competition for the output link among cells of different traffic classes at an ATM multiplexer or switch.

At low loads, a simple first-come-first-serve (FCFS) scheduling strategy may be adequate to satisfy all QOS requirements, since minimal queueing will occur. At higher loads, a more complicated scheduling strategy may be required to resolve contention between different traffic classes (e.g., cells from a video traffic stream may have priority over cells from a data traffic stream). At extremely high loads, the queueing delay for all traffic streams may be too high to meet any of the QOS constraints, regardless of the scheduling strategy used.

The schedulable region of a queueing system, intuitively, is the set of points in the space of possible loads for which the quality of service can be guaranteed [15, 16]. The schedulable region indicates the possible loads for which a scheduling algorithm guarantees quality of service. The size and shape of a schedulable region are highly dependent upon the scheduling algorithm employed, the QOS parameters under consideration, and the characteristics of the traffic.

The schedulable region has been used in admission control strategies [15, 16]. By estimating the size of the schedulable region, the admission control procedure can decide whether a new call should be accepted or rejected. The admission controllers are constrained by the need for the network to guarantee the required QOS at the cell level to all calls admitted into the network, and by limits on the call level QOS constraint such as call blocking probability as well. The acceptance decision will be made based on the information provided to admission controllers: the boundaries of the schedulable region specified by the scheduler, and the call arrival and departure rates associated with each type of traffic [16]. The advantage of such an approach is that the admission control procedure can be shielded from the flood of detailed cell-level information. The QOS can be guaranteed at the cell-level simply by keeping the number of calls within the schedulable region. Detailed admission control policies and simulation results are discussed in the literature [16].

3.2.2 Quality-estimate-based Admission Control Policies

Because of the resource sharing at each network node, the performance of calls already in progress may be affected by the admission of a new call. The current node load will also determine the quality of service that can be provided to the new incoming call.

The prediction of how the QOS performance for the calls already in the network will be affected by the admission of a new incoming call is very difficult to make. This difficulty is partly caused by the uncertainties involved in modeling the statistical behaviour of many types of traffic sources. More importantly, however, it is due to the complicated way in which different traffic streams interact with each other within a packet network [9]. In an integrated multimedia service environment, these issues are further complicated by the high transmission speeds and the diverse mix of traffic types and service requirements.

Several admission control policies proposed in recent years are based on measurements that estimate the possible effect of adding a new call into the network. Several measurement criteria such as *virtual cell loss rate, clipping rate and average* *packet loss gap, upper bound on cell loss probability, and effective bandwidth* have been used in different admission control policies. The definitions and applications of these measurements will be described below.

1. **Virtual cell loss rate**

   The virtual cell loss rate admission control policy operates on a logically bufferless fluid flow network model. Cell losses due to overflow occur if and only if an aggregate peak rate P exceeds the link capacity L. When the aggregate peak bit rate P is less than the link capacity L, cell loss is assumed to never happen [21]. The virtual cell loss probability is the ratio of excess traffic and offered traffic.

   Simulation results [21] show that, for both homogeneous and heterogeneous traffic, the virtual cell loss probability in the bufferless flow model is always higher than the real cell loss probability p in the same flow model with a finite size of buffer. When the average burst length B is increased, the virtual cell loss probability provides an upper bound on the cell loss probability p in the queueing model. Furthermore, p converges to the value of virtual cell loss when B goes to infinity [21].

   The virtual cell loss probability can be used as an acceptance criteria for an admission control policy. Virtual bandwidth and virtual link capacity schemes allocate bandwidth based on the peak bit rate P_i, mean bit rate A_i for Class i traffic, and virtual cell loss probability. The detailed algorithm can be found in [21, 29].

   The advantages of the virtual cell loss model include its explicit formulation, and easy policing for both peak bit rate and mean bit rate [21].

2. **Clipping probability and average packet loss gap**

   A set of QOS constraints has been introduced in the literature [6] for real-time traffic to measure the quality of service received by such calls. For each call, the packet clipping probability and the average packet loss gap are defined as an acceptance criteria for admission control. The packet clipping probability is the probability of packets being discarded due to exceeding the delay boundary. The average packet loss gap is defined as the average number of consecutive packets lost for a given call [6]. With high correlation among arriving packets, this set of measurements may be of practical interest for an integrated network that supports video, voice, and data traffic.

The clipping probability and average packet loss gap are the performance degradation criteria that are used by admission control to make the acceptance decision for a new real-time call. The admission controller attempts to maximize the number of calls while guaranteeing the quality of service for each call. The admission control procedure will check if the quality of service for each call is acceptable for the new proposed loads. If all the calls admitted in the network and the new call can have an acceptable quality of service, the new connection will be established at the node [6].

The disadvantage of this approach is that it is only suitable for Poisson arrival type of traffic because the estimation of clipping probability of average packet loss gap is based on an MMPP model.

3. **Upper bound on cell loss probability**

Saito [27] proposes an admission control policy based on estimating an upper bound on cell loss probability. The upper bound on cell loss probability is the boundary of the total cell loss probabilities in the network for all classes of traffic, which is always greater than or equal to the real total cell loss probabilities for all classes. The calculation of the upper bound on cell loss probability is based on the QOS constraints and traffic descriptor for each class of traffic, such as cell loss and delay constraints, the peak bit rate, and mean bit rate. From these parameters, the distribution of the number of arrival cells for each call of each class of traffic can be estimated and the upper bound on cell loss probability is calculated.

The admission control policy can use the upper bound on cell loss probability as an acceptance criteria. If, including the new call, the upper bound on cell loss probability is in the acceptable level for all the calls, the new connection will be established and rejected otherwise.

The advantage of this scheme is that the control is independent of the classification of calls and the number of call classes [27]. Thus, the scheme is effective when the number of call classes is fairly large. The disadvantage of this scheme is that it treats different classes of traffic the same. In other words, the admission control policy based on upper bound on cell loss probability has to meet the most stringent QOS constraint among all classes of traffic, for every call involved. It is not desirable for multimedia traffic with a wide range of QOS requirements.

4. **Effective Bandwidth**

Elwalid [4] proposes an admission control policy based on effective bandwidth estimation. The effective bandwidth is the estimated bandwidth requirement for an n-state traffic source, which is usually higher than the mean bandwidth requirement and lower than the peak bandwidth requirement. The calculation of the effective bandwidth for each n-state source is based on the QOS constraints and traffic descriptor, such as cell delay and loss constraints, the peak bit rate, the mean bit rate, and the transition probabilities from one state to another.

For a special type of two-state source, the ON/OFF source with exponentially distributed ON and OFF periods (where P is the peak bit rate, A is the mean bit rate, α is the probability of transition from OFF state to ON state, and β is the probability of transition from ON state to OFF state), the effective bandwidth function is:

$$g(z) = \frac{1}{2z}[Pz + \alpha + \beta] - \frac{1}{2z}\sqrt{(Pz + \alpha + \beta)^2 - 4\beta Pz} \tag{3.1}$$

where $z \in [-\infty, 0]$, and $g(z) \in [A, P]$. Let the buffer size, denoted by M, be bounded by cell delay constraints. Let the channel capacity be L. Let the cell loss probability boundary be p. When the buffer overflow probability is less than p (e.g., the QOS constraints are met), the value of $\frac{log(p)}{M}$ is called the admission criterion ζ.

The total effective bandwidth $\sum_i g_i(z)$ of heterogeneous traffic is computed from $g_i(z)$ for the ith type of traffic [5]. If $\sum_i g_i(\zeta) < L$, then the new call is accepted. Otherwise the new call is rejected.

.

3.3 Summary

The admission control procedure attempts to prevent cell-level congestion from happening by controlling the number of calls admitted to the network. Several admission control policies are discussed in this chapter. The multiplexing mode has a great impact on the performance of an admission control strategy. Operated in deterministic multiplexing mode, an admission control policy is conservative. An aggressive admission control policy, operating in statistical multiplexing mode, accepts more calls by taking the risk of violating QOS constraints. The next chapter presents the simulation setup for the quantitative evaluation of several admission control policies.

Chapter 4

Experimental Methodology

This chapter discusses the simulation setup, including the network model, the traffic models, the scheduling algorithms, and the admission control policies.

4.1 Network Model

The simulation model used in the experiments is shown in Figure 4.1. The ATM network is modeled as a multiplexer with up to N heterogeneous input sources and one output port. The input and output link capacities are $L = 150Mb/s$. The multiplexer checks the input ports in a round robin fashion. The arriving cells are multiplexed onto the output port, which has three sets of buffers and one server. Each set of buffers is a completely shared buffer with a finite capacity of M cells assigned to a specific class of traffic (e.g., video, voice, and data). The size of the video buffer (M_{video}) is 2124 cells, the size of the voice buffer (M_{voice}) is 5310 cells, and the size of the data buffer (M_{data}) is 10620 cells. The value of M is large enough for each class of traffic so that no buffer overflow occurs. In other words, cell loss for a class of traffic occurs only when the cell delay for the class exceeds the cell delay boundary, and never due to buffer overflow. The processing time at the multiplexer is assumed to be zero, since it is independent of the class of traffic and the traffic model used. Therefore, cell delay is only due to queueing at the server. The time unit in the simulation is the time to transmit one cell on the outgoing link (2.83 microseconds).

Input cells are serviced by the server, and then forwarded to the destination. If the network server is idle at the time of a cell arrival, the newly arriving cell will go directly to the server. When the server is busy, arriving cells will be put into the buffer according to its class of traffic. After a cell has been served, the server transmits the cell to the destination and selects the next cell to be served from the buffers based on certain scheduling rules. Two scheduling algorithms, Deterministic Scheduling and Deadline Scheduling, are examined in the thesis. The detailed algorithms are discussed in Section 4.3.

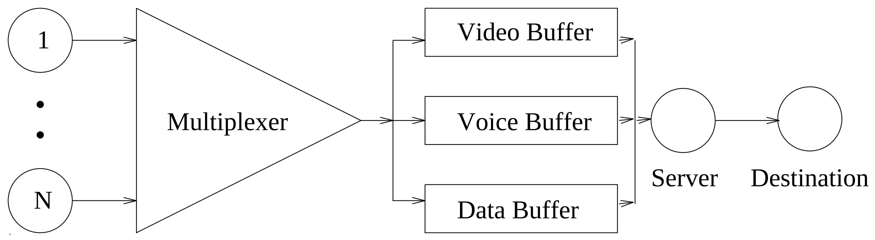


Figure 4.1: Network Model

4.2 Traffic Models

Three classes of traffic are used in the experiments: Class I for video traffic, Class II for voice traffic, and Class III for data traffic. The QOS constraints and the traffic models for each traffic class are presented in the following subsections.

4.2.1 Quality of Service Constraints

The quality of service constraints for the three types of traffic (Class I, II, III) are defined by a six-tuple $(\varepsilon^I, \eta^I; \varepsilon^{II}, \eta^{II}; \varepsilon^{III}, \eta^{III})$, where ε represents cell delay constraints, and η represents cell loss constraints.

Class I traffic has QOS constraints specified by ε^I and η^I. The maximum tolerable cell delay (in milliseconds) at the multiplexer is ε^I. Class I requires a cell loss probability that is less than or equal to η^I.

Class II traffic has QOS constraints specified by ε^{II} and η^{II}. The maximum tolerable cell delay (in milliseconds) at the multiplexer is ε^{II}. Class II requires a cell loss probability that is less than or equal to η^{II}.

Class III traffic has QOS constraints specified by ε^{III} and η^{III}. The maximum tolerable cell delay (in milliseconds) at the multiplexer is ε^{III}. Class III requires a cell loss probability that is less than or equal to η^{III}. For Class I, Class II and Class III traffic, no retransmission is allowed.

The call level QOS constraints in the experiments include the call blocking probability, the call arrival rate, and the call departure rate. The blocking probability is 0.1. Call arrivals are assumed to follow a Poisson process with an average rate of λ^i calls per second for Class i traffic. The time that a call sends a request for connection is called the call arrival time. The time that a call is accepted by the admission controller is called the call admission time. The time that a call stops generating cells is called the call departure time. The time that a call sends a request for connection release is called the call disconnection time. The incoming calls will be either accepted immediately or blocked immediately. Call durations follow an exponential distribution, and the call departure process is Poisson with a mean of μ^i calls per second for Class i traffic. Cells of a departing call that are queued in the buffer are always processed to completion before the call disconnects. These cells are carried to the destination if the delay constraint is met, otherwise discarded.

4.2.2 Traffic Model

1. **Video Traffic Model**

   Two video traffic models have been used in the simulation experiments for Class I traffic. Both models are based on Hyman's two-state model (Section 2.4.2), but with two different parameter settings to explore the impact of different traffic characteristics. Video Model1 uses Hyman's original parameter settings ($F = 62.5ms, P = 10Mb/s, C = 25ms$).

   Video Model2 is defined using:

   - Fixed frame duration: $F = 62.5ms$

- Peak bit rate: $P = 50Mb/s$
- Average active period: $C = 5ms$

Video Model2 has a higher peak rate ($50Mb/s$ versus $10Mb/s$) and different burst behaviour than Video Model1. The active period, C_{active} is a random variable uniformly distributed between $C_{min} = 2ms$ and $C_{max} = 8ms$. At the end of the active period, the source stops emitting cells for the duration of the frame ($F - C_{active}$). The mean bit rate of a single video source remains the same as Video Model1 with $A = 4Mb/s$.

2. **Voice Traffic Model**

A single voice call (Class II traffic) is modeled as an ON-OFF source. It is characterized by an exponentially distributed active period, in which cells are generated with constant rate $P = 64\ Kb/s$. C_{on} and C_{off} are two statistically independent random variables with a negative exponential distribution describing the ON-OFF behaviour of the source. $\overline{C_{on}} = 352$ ms and $\overline{C_{off}} = 650$ ms denote the average values of ON period and OFF period, respectively [15]. The average rate for one such source is given by:

$$A = \frac{C_{on}}{C_{on} + C_{off}} * P = 22.5\ Kb/s \tag{4.1}$$

The average burst length is $B = 53\ cells$.

3. **Data Traffic Model**

A single data call (Class III traffic) is modeled by a two-state Markov model as shown in Figure 2.4. In this experiment, the three parameters, average burst length B, the peak link rate P, and the average link rate A, are chosen as

- $B = 4150\ cells$
- $P = 5\ Mb/s$
- $A = 1\ Mb/s$

4.3 Scheduling Algorithms and Admission Control Policies

Two scheduling algorithms have been examined in order to explore the statistical behaviour of scheduling in the ATM multiplexer. A scheduling algorithm that operates in deterministic multiplexing mode, Deterministic Scheduling, and a scheduling algorithm that works in statistical multiplexing mode, Deadline Scheduling, are presented in the following subsection. The definition and description of joint scheduling and admission control policies is discussed in Section 4.3.2.

4.3.1 Scheduling Algorithms

Two scheduling algorithms have been studied in the experiments.

1. **Deterministic Scheduling**

   Deterministic scheduling operates in deterministic multiplexing mode. The number of input sources is bounded by the rule that the total peak bandwidth requirements by input sources can never exceed the output link capacity. The video cells of Class I have the highest priority among the three classes. The priority of voice cells of Class II is higher than data cells of Class III. The voice cells can be transmitted to the destination only when no video cells are in the video buffer. The data cells will be transmitted when the other two buffers are empty. Within the same class of traffic, cells are transmitted in a first-come-first-serve fashion. Cells of the same class that arrive during the same slot are transmitted in round robin fashion.

2. **Deadline Scheduling**

   Deadline scheduling operates in statistical multiplexing mode. Let $D = \{d^I, d^{II}, d^{III}\}$ be a set of cell delay boundaries for Class I, II, and III traffic respectively. Let $N^i = \{n_1, n_2, \cdots, n_k\}$ be a set of cells in the buffer of Class i, and let w^i_j be the waiting time for cell j in the buffer of Class i. The waiting time is defined as the time period between the time that a cell arrived at the buffer and time

t. The maximum waiting time at time t is defined as:

$$W^i = MAX\{w_1^i, w_2^i, \cdots, w_k^i\} \tag{4.2}$$

The deadline for Class i is defined as:

$$V^i = d^i - W^i \tag{4.3}$$

The cells of traffic class with the smallest deadline value (i.e., $Min\{V^I, V^{II}, V^{III}\}$) are given the highest priority. The cells of the traffic class with the largest deadline value have the lowest priority. The cells with the highest priority are transmitted first. The cells with the lowest priority are transmitted last. Within the same priority class, cells are transmitted in first-come-first-serve fashion. V^i is re-evaluated every one millisecond.

4.3.2 Admission Control Policies

Four admission control policies will be evaluated in the thesis.

1. **Deterministic Policy**

   The deterministic control policy operates in the deterministic multiplexing mode. Let $M = \{m_1, m_2, m_3\}$ be a set of traffic classes, $N^i = \{n_1^i, n_2^i, \cdots, n_y^i\}$ be a set of calls admitted into the network for class i traffic, P_j^i be the peak rate for the jth call of Class i admitted into the network and P_{j+1}^i be the call waiting to be admitted. The sum $\sum_{j \in N^i} P_j^i$ is the total bandwidth requirement for all the admitted calls of Class i under deterministic multiplexing. The sum $\sum_{i \in M}(\sum_j P_j^i)$ is the total bandwidth requirement for all the calls admitted into the network of all three classes. If the following expression:

   $$\sum_{i \in M}\Big(\sum_{l \in (N^i, j+1)} P_l^i\Big) \leq L \tag{4.4}$$

   is true, then the call P_{j+1}^i is admitted. Otherwise, the call is blocked.

2. **Joint Scheduling and Admission Control Policy**

   In the joint scheduling and admission control policy, a schedulable region based on the Deadline Scheduling algorithm is predefined, measured, and stored with the admission control procedure. Incoming calls are always admitted when the number of calls in the network does not exceed the boundary of the schedulable region.

3. **Effective Bandwidth Policy**

   The Effective Bandwidth Policy discussed in Section 3.2.2 is used in the experiment. Although the video model used in the experiments is not a two state model with exponentially distributed active and idle states, it is assumed that the video model can be approximated with such a two-state model. The same effective bandwidth function equation (Equation 3.1) was thus used for the video model in the simulation to give an approximation of the bandwidth requirements. The evaluation results of this policy is conservative (i.e., it results in an effective bandwidth estimate that is close to the peak bandwidth requirement). The results for this policy may depend on the video model used.

4. **Virtual Bandwidth Policy**

   The Virtual Bandwidth admission control policy is based on virtual cell loss probability estimates as described in Chapter 3. The virtual cell loss probability is the ratio of excess traffic T and average traffic load O. The virtual cell loss probability v for homogeneous traffic is defined in the following equation [21], where the number of inputs is defined by N.

$$v = T/O \tag{4.5}$$

$$T = \sum_{(n*P-L)\geq 0}^{n=N} p(n)(n*P-L) \tag{4.6}$$

$$O = N*A \tag{4.7}$$

where P is the peak bit rate, A is the mean bit rate, L is the output link capacity, and $p(n)$ is the probability that n out of N sources are active. That is,

$$p(n) = \begin{pmatrix} N \\ n \end{pmatrix} \left(\frac{A}{P}\right)^n \left(1 - \frac{A}{P}\right)^{N-n} \tag{4.8}$$

The virtual bandwidth V_B is defined based on the virtual cell loss probability, as follows:

$$V_B = A/L' \tag{4.9}$$

where L' is the maximum link capacity satisfying a specific v condition for the homogeneous traffic, which can be derived from Equation 4.6.

For two types of traffic, the virtual cell loss probability for Class j type of traffic can be estimated by the following equations [21]:

$$v_j = T_j/O_j \tag{4.10}$$

$$T_j = \sum_{(n_1P_1+n_2P_2)-L\geq 0}^{n_1=N_1} \sum^{n_2=N_2} \left[\frac{(p_1(n_1)p_2(n_2))(((n_1P_1 + n_2P_2) - L)P_jn_j)}{n_1P_1 + n_2P_2}\right] \tag{4.11}$$

$$O_j = A_j * N_j \tag{4.12}$$

$$p_i(n_i) = \begin{pmatrix} N_i \\ n_i \end{pmatrix} \left(\frac{A_i}{P_i}\right)^{n_i} \left(1 - \frac{A_i}{P_i}\right)^{N_i-n_i} \tag{4.13}$$

where N_i is the number of Class i calls, A_i is the mean bit rate for Class i traffic, P_i is the peak bit rate for Class i traffic, and L is the output link capacity.

The virtual bandwidth for Class j type of traffic can be calculated following the same procedure as for the homogeneous traffic.

The virtual bandwidth admission control policy will accept or deny a new call based on the new call's virtual bandwidth requirement. If the available bandwidth is large enough for the new call, the new call will be accepted. Otherwise, the call is denied.

4.4 Summary

In this chapter, the simulation setup, including the network model, the traffic models, the scheduling algorithms to be studied, and the four admission control policies to be evaluated, have been discussed. The statistical behaviour of the scheduling algorithms is discussed in Chapter 5. The performance evaluation of the admission control policies is presented in Chapter 6.

Chapter 5

Statistical Behaviour of Scheduling Algorithms

In this chapter, the results of simulation experiments that illustrate several properties of scheduling algorithms and their associated schedulable regions are presented. The schedulable region for a given system depends on three factors: the QOS constraints, the statistical characteristics of the offered traffic, and the choice of the scheduling algorithm.

5.1 Schedulable Region

The schedulable region represents the possible traffic load on the network for which the QOS constraints can be guaranteed. To obtain the schedulable region for Deterministic Scheduling and Deadline Scheduling, respectively, the network model in Section 4.1 has been used. The call level QOS constraints are not considered in this part of the experiments since the scheduling algorithm is a cell level traffic control strategy and traffic sources stay in the network until the end of the simulation. The QOS constraint set defined in Section 4.2.1 has been given the value of :

$$QOS\ SET = (1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9}) \tag{5.1}$$

5.1.1 Schedulable Region for Deterministic Scheduling

The schedulable region for Deterministic Scheduling is calculated for a traffic mixture of video and voice traffic. Let N^I and N^{II} be the number of video sources and voice sources respectively, P^I and P^{II} be the peak bit rate for the video traffic model (Video Model1) and the voice traffic model, respectively, and $\mathcal{N}^I$ and $\mathcal{N}^{II}$ be the maximum number of video sources and voice sources, respectively. For every given value of $N^I \in [0, \mathcal{N}^I]$, the value N^{II} is calculated by using the following:

$$(N^I * P^I \; + \; N^{II} * P^{II}) \leq L \tag{5.2}$$

P^I and P^{II} determine the shape and size of the schedulable region. Figure 5.1 shows the schedulable region for Deterministic Scheduling. Deterministic Scheduling can allow up to 15 video calls, or up to 2340 voice calls, or any intermediate combination of voice and video calls that remains within the schedulable region.

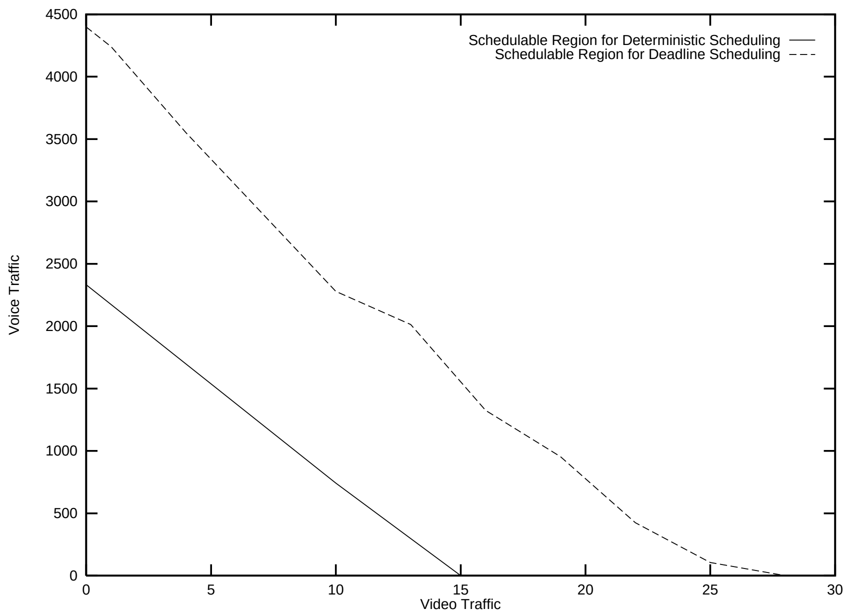


Figure 5.1: Schedulable Region for Deterministic Scheduling and Deadline Scheduling for the Video and Voice Traffic Scenario

5.1.2 Schedulable Region for Deadline Scheduling

The schedulable region for Deadline Scheduling has been determined for three traffic scenarios: video and voice, video and data, and data and voice. Several steps have been followed to determine each schedulable region. The schedulable region for the video and voice traffic scenario is used as an example to illustrate the steps.

Let N^I and N^{II} be the number of video sources and voice sources, respectively, where $N^I \in [0, \mathcal{N}^I]$ and $N^{II} \in [0, \mathcal{N}^{II}]$. Let the QOS constraints for video and voice be $Q^I = (\varepsilon^I, \eta^I)$, $Q^{II} = (\varepsilon^{II}, \eta^{II})$, respectively. Then the schedulable region can be determined as follows:

- **STEP 0**: Set $N^I = 0$; $N^{II} = 0$.

- **STEP 1**: For N^I video traffic sources and N^{II} voice traffic sources, simulate the scheduler and determine the cell delay and cell loss performance for the two traffic classes from simulation results, respectively.

- **STEP 2**: If (Q^I is violated or Q^{II} is violated) then goto STEP 3.

  Otherwise:

  Set $N^{II} \leftarrow N^{II} + 1$; If $N^{II} \leq \mathcal{N}^{II}$ then goto STEP 1; otherwise EXIT.

- **STEP 3**:

  - Plot $(N^I,\ N^{II} - 1)$ as a boundary point for the schedulable region;
  - Set $N^{II} = 0$; $N^I \leftarrow N^I + 1$;
  - If $N^I \leq \mathcal{N}^I$, then goto STEP 1; otherwise EXIT.

The total number of cells sent by the two types of sources in each simulation run were 14,500,000 cells representing a simulation time of 60,000 milliseconds. In addition, five replications of the simulation were carried out for each schedulable region using different random number seeds. The final schedulable region plotted is the average from five simulation runs, as shown in Figure 5.1. The confidence interval

Table 5.1: Confidence Intervals for the Mean of the Schedulable Region for the Deadline Scheduling Algorithm

Video Sources	Voice Sources	99% Confidence Interval
0	4402	[4397, 4407]
1	4242	[4235, 4249]
4	3552	[3545, 3558]
7	2915	[2912, 2917]
10	2279	[2270, 2288]
13	2014	[2009, 2018]
16	1325	[1317, 1332]
19	954	[947, 961]
22	424	[419, 427]
25	106	[103, 109]
28	0	[0, 0]
30	0	[0, 0]

for a confidence level of 99% has been calculated for the mean. Table 5.1 shows the confidence intervals.

The schedulable regions for the other two traffic scenarios (Video and Data, and Data and Voice) can be found in Appendix A (Figures A.1 and A.2).

The schedulable region for Deadline Scheduling in Figure 5.1 is much higher and wider than the schedulable region for Deterministic Scheduling. Because it operates in statistical multiplexing mode, the Deadline Scheduling algorithm handles much higher traffic load and allows more traffic sources to share the output channel. With the same input traffic models, Deterministic Scheduling handles less than one half as many sources as statistical scheduling. These simulation results show that with statistical multiplexing, ATM networks can provide more efficient bandwidth allocation and achieve better bandwidth utilization for bursty traffic sources.

5.2 Dependency of Scheduling Algorithms on QOS Constraints

Simulation results show that QOS constraints have no influence on the shape or size of the schedulable region for Deterministic Scheduling, and only the peak bit rate of incoming traffic affects the schedulable region for Deterministic Scheduling. Therefore, this section will only focus on the sensitivity of Deadline Scheduling to QOS constraints. The discussion is in two parts: dependency on cell delay constraints, and dependency on cell loss constraints.

5.2.1 Dependency on Cell Delay Constraints

Two sets of QOS constraints have been used to explore the sensitivity of the Deadline Scheduling algorithm to the cell delay parameter. The two sets of QOS constraints are defined by:

$$QOS\ SET_1 = (\varepsilon_1^I, \eta_1^I; \varepsilon_1^{II}, \eta_1^{II}; \varepsilon_1^{III}, \eta_1^{III}) = (1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9}) \tag{5.3}$$

$$QOS\ SET_2 = (\varepsilon_2^I, \eta_2^I; \varepsilon_2^{II}, \eta_2^{II}; \varepsilon_2^{III}, \eta_2^{III}) = (3ms, 10^{-9}; 6ms, 0.001; 15ms, 10^{-9}) \tag{5.4}$$

$QOS\ SET_2$ relaxes the delay constraints of $QOS\ SET_1$ by a factor of 3.

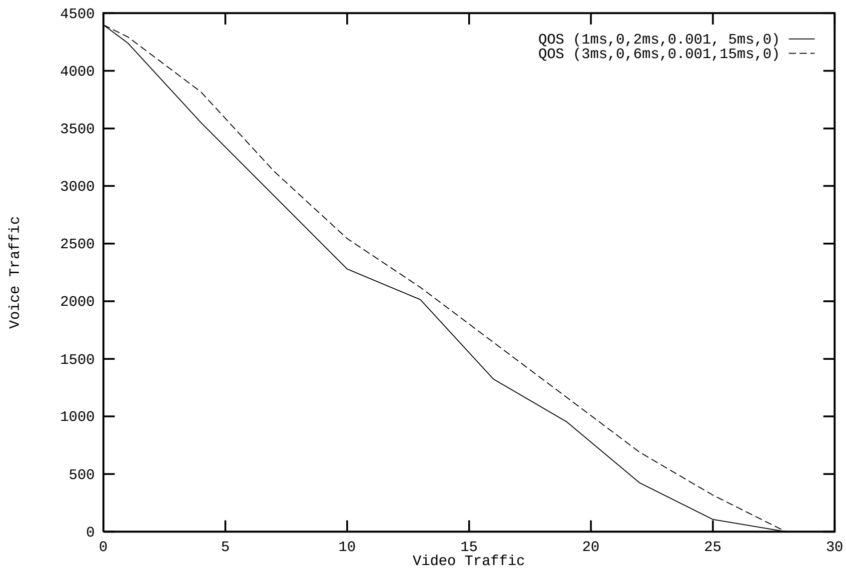


Figure 5.2: The Comparison of Schedulable Regions Under $QOS\ SET_1$ and $QOS\ SET_2$ for Deadline Scheduling for the Video and Voice Traffic Scenario

Figure 5.2 shows the comparison of the schedulable regions for the video and voice traffic scenario under these two sets of QOS constraints. When $QOS\ SET_2$ applies, the schedulable region increases in size.

The increase in the schedulable region for $QOS\ SET_2$ indicates that Deadline Scheduling can accommodate more traffic sources compared to the region with $QOS\ SET_1$. For example, if the number of video sources is $N^I = 20$, then relaxing the delay constraints from $QOS\ SET_1$ to $QOS\ SET_2$ would allow the number of voice sources to increase from 848 to 1007.

The increase in the size of the schedulable region can be expressed mathematically as *Utility Gain.* Let $N^I = (0, 1, 2, \cdots)$ and $N^{II} = (0, 1, 2, \cdots)$ be the number of video sources in the system for $QOS\ SET_1$ and $QOS\ SET_2$, respectively. Let $S^I = (S_0^I, S_1^I, \cdots, S_{\mathcal{N}^1}^I)$ and $S^{II} = (S_0^{II}, S_1^{II}, \cdots, S_{\mathcal{N}^1}^{II})$ be the set of maximum number of voice sources for every given N^I and N^{II}, respectively. The utility gain, then, can be defined as

$$U = \left(\sum_{i=0}^{\mathcal{N}^I} (S_i^{II} - S_i^I) \right) \tag{5.5}$$

The utility gain is the total number of traffic sources increased when QOS constraints are relaxed. Overall, there is significant utility gain (212 voice sources) for Deadline Scheduling when $QOS\ SET_2$ is used for the video and voice traffic scenario. Similar results are obtained for the traffic scenarios of video and data, and data and voice, as shown in Appendix A (Figures A.3 and A.4).

5.2.2 Dependency on Cell Loss Constraints

The second part of this experiment studies the sensitivity of the schedulable region to cell loss constraints for Deadline Scheduling. Two relaxed cell loss QOS settings for video traffic have been used: $\eta_1^I = 0.0001$ and $\eta_2^I = 0.01$. Figure 5.3 presents the simulation results for the video and voice traffic scenario for the given cell loss constraint values, along with the schedulable region for cell loss $= 10^{-9}$ from Figure 5.1. The comparison shows that as the cell loss constraint is relaxed, the schedulable region increases in size, and the utility gain increases. The stricter the cell loss constraint is, the smaller the schedulable region will be. In other words, if video

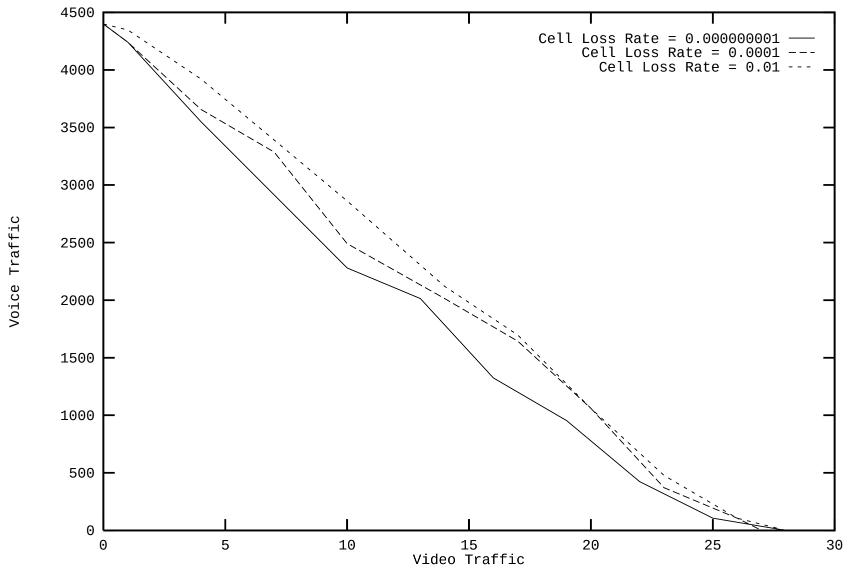


Figure 5.3: The Comparison of Schedulable Regions With Relaxed Cell Loss Constraints for Deadline Scheduling for the Video and Voice Traffic Scenario

traffic sources could be coded to tolerate higher levels of cell loss [30] then more traffic sources could be admitted to the network.

5.3 Dependency of Scheduling Algorithms on Traffic Characteristics

Video Model2 was used in this simulation experiment to explore the sensitivity of both Deterministic Scheduling and Deadline Scheduling algorithms to the input traffic characteristics.

5.3.1 Deterministic Scheduling

When Video Model2 was used in the simulation, the schedulable region observed is much lower and narrower than the region for Video Model1. Figure 5.4 illustrates the comparison of these two schedulable regions. From Figure 5.4, the number of voice sources that can be accommodated by Deterministic Scheduling decreases greatly when the peak bit rate for video source increases from 10 Mb/s to 50 Mb/s. Deter-

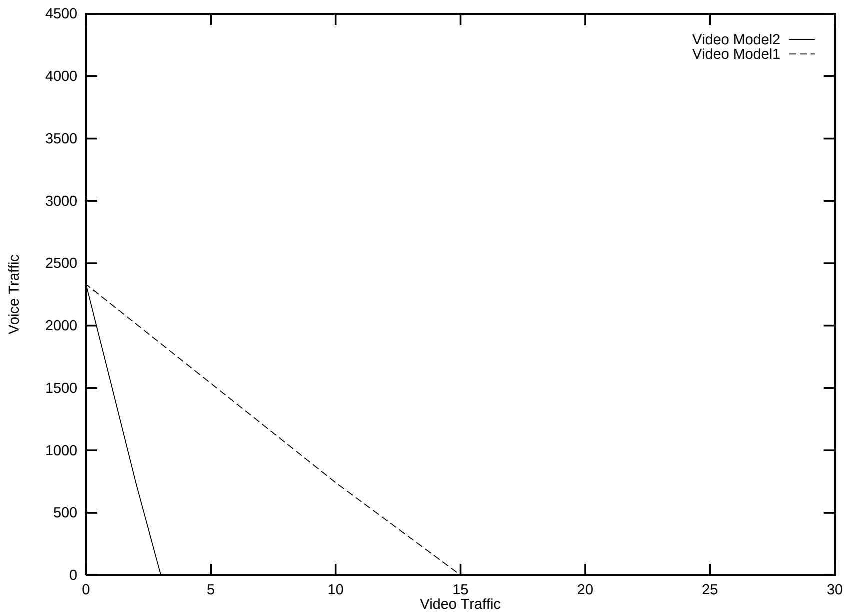


Figure 5.4: The Comparison of Schedulable Regions With Different Traffic Characteristics for Deterministic Scheduling for the Video and Voice Traffic Scenario

ministic Scheduling is clearly sensitive to traffic characteristics because the maximum number of traffic sources is limited by the link capacity L.

5.3.2 Deadline Scheduling

Similar results were obtained when Video Model2 was used in the simulation to plot the schedulable region for Deadline Scheduling. Figure 5.5 presents the sensitivity of Deadline Scheduling to the traffic characteristics. The comparison of the schedulable regions for traffic scenarios of video and data can be seen in Appendix A (Figure A.5). In Figure 5.5, the schedulable region is significantly smaller for Video Model2 than for Video Model1. Since Video Model2 has higher peak rates and shorter bursts than Video Model1, fewer video sources can be accommodated.

Compared to the schedulable region for Deterministic Scheduling (Figure 5.4), the schedulable region for Deadline Scheduling (Figure 5.5) is higher and wider. The shorter active period (2ms to 8ms) of the video sources has allowed the statistical scheduler to handle more traffic sources than the deterministic scheduler does by taking advantage of the periodically required bandwidth request from these video sources.

More importantly, however, these simulation results suggest that realistic traffic models are very important to simulation studies of congestion and admission control. The simulation results for the schedulable region in this case are highly dependent on the input traffic characteristics.

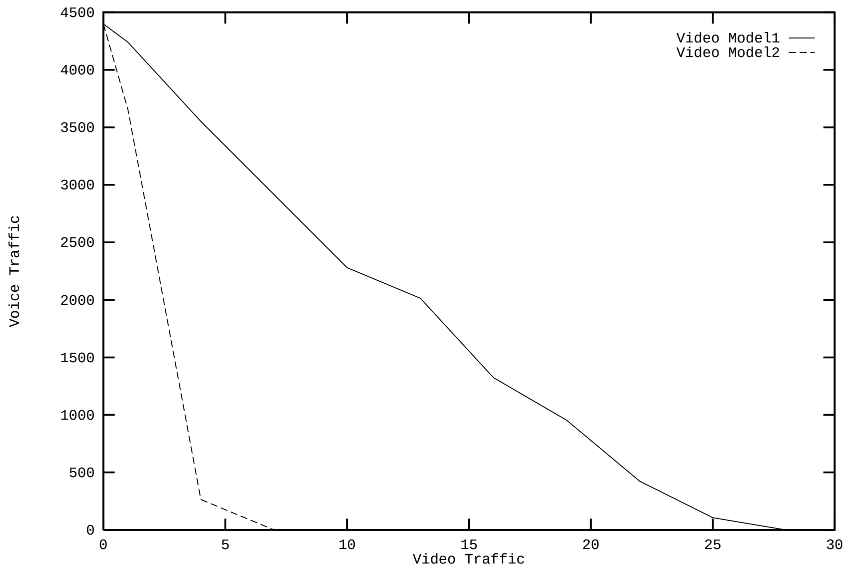


Figure 5.5: The Comparison of Schedulable Regions With Different Traffic Characteristics for Deadline Scheduling for the Video and Voice Traffic Scenario

5.4 Statistical Gain

As discussed in Chapter 2, with deterministic multiplexing, the channel utilization could be very low when traffic sources are bursty. To achieve better channel utilization, ATM networks provide a statistical multiplexing mode that takes advantage of the variable information transfer rates generated by many connections to implicitly share resources.

Statistical gain (an increase in effective network utilization) is achieved when statistical multiplexing mode is used. Compared to a maximum channel utilization of 38.3% obtained by using Deterministic Scheduling (Figure 5.6), channel utilization as high as 82% (Figure 5.7) can be achieved when statistical multiplexing mode is used by the Deadline Scheduling algorithm.

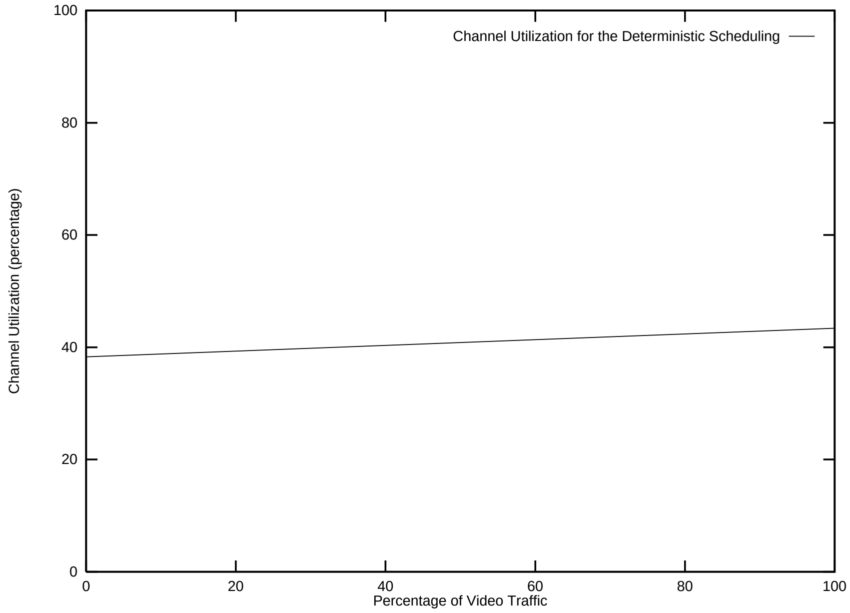


Figure 5.6: Channel Utilization for Deterministic Scheduling for the Video and Voice Traffic Scenario With $QOS\ SET = (1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9})$

Figures 5.6 and 5.7 illustrate the channel utilization as a function of the percentage of video sources in the network. When no video sources are in the system (i.e., 0% of video sources), the channel utilizations are 38.3% for Deterministic Scheduling, and 66.9% for Deadline Scheduling. When there were no voice sources in the network (i.e., 100% video sources), the utilizations are 43.4% for Deterministic Scheduling, and 82% for Deadline Scheduling. The higher the percentage of video sources in the system, the better the improvement in channel utilization is.

5.5 Summary

The statistical behaviour of scheduling algorithms have been carried out as part of the work for the thesis. The scheduling algorithm controls the traffic streams at the cell level. The simulation results show that the different scheduling algorithms have different schedulable regions. Understanding the statistical behaviour of scheduling algorithms is the basis that leads to the further study of Joint Scheduling and Admis-

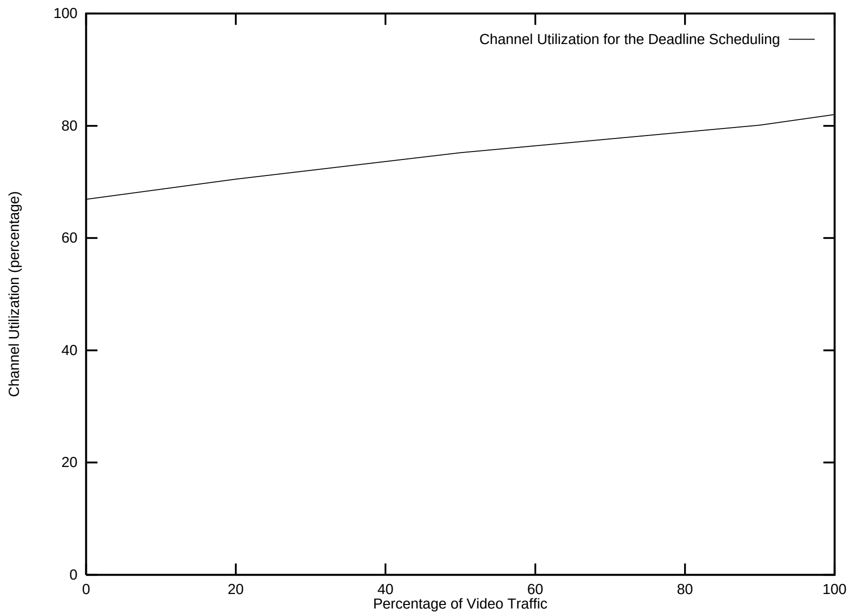


Figure 5.7: Channel Utilization for Deadline Scheduling for the Video and Voice Traffic Scenario With $QOS\ SET\ = (1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9})$

sion Control Policy. The loading region introduced in Chapter 6 is the metric that evaluates the efficiency of an admission control policy quantitatively. More experiments have been carried out in the thesis to evaluate four different admission control policies. Chapter 6 discusses the performance evaluation of the four admission control policies.

Chapter 6

Performance Evaluation of Admission Control Policies

In this chapter, the loading region and the Expansion Degree are introduced as a means to evaluate the performance of admission control policies quantitatively. The second part of this chapter presents the performance evaluation of four admission control policies using the loading region and the Expansion Degree as the framework.

6.1 Loading Region

Several quantitative evaluation methods such as admissible region and bandwidth comparison have been introduced in the literature [16, 21, 23, 27, 29] to evaluate the performance of admission control policies. These methods are insufficient to indicate, for an admission control policy, the level of aggressiveness, the level of the assurance of the QOS constraints for calls admitted into the network, and the operation of the multiplexing mode.

To overcome these weaknesses, the loading region is introduced in the thesis as a means for the quantitative evaluation of admission control policies. The loading region is classified further into three parts: Green, Yellow, and Red regions. The three loading regions represent three levels of aggressiveness for an admission control policy. A conservative admission control policy operates in the green region, where the calls are admitted deterministically into the network so that the QOS constraints for all classes of traffic are guaranteed. An aggressive admission control policy operates in the yellow region, where the calls are admitted statistically into the network and the QOS constraints for all classes of traffic are met. The red region indicates the danger sign that the number of calls admitted into the network exceeds the manageable level. While the schedulable region indicates the maximum traffic load condition that could be handled by an ATM multiplexer at the cell-level, the loading region shows maximum traffic intensity that is manageable by an admission control policy at the call-level.

6.1.1 Definition

A loading region indicates the number of calls of heterogeneous traffic that can be admitted into the network. The shape and size of a loading region will be highly dependent upon QOS constraints, admission control policies, and traffic characteristics. The QOS constraints can be divided into two levels: cell level and call level. The cell level constraints may include cell loss, cell delay, and maximum average user time delay. The call level constraints may be indicated by call blocking probability, call arrival rates, and call departure rates.

A loading region is called a green region whenever for all calls of all types of traffic admitted in the network under the load condition indicated by the region, the total bandwidth requirements for all calls does not exceed the output link capacity and the QOS constraints are guaranteed to be satisfied. A loading region is called a yellow region when the total bandwidth requirements for all calls exceed the output link capacity and the QOS constraints can be met for all calls of all types of traffic admitted in the network under the load condition indicated by the region. When the number of calls admitted in the network falls in the red region, the QOS performance constraints cannot be achieved at all.

The three types of loading regions can be defined more formally as follows. For classes of traffic $i \in$(I, II, III), let Q^i be the set of cell level QOS constraints for class

i, λ^i be the call arrival rate of calls/second for class i, and T^i be the set of call level QOS constraints for class i. A green region, denoted by G, can be defined as:

$$\begin{aligned} G \;=\; \{\; (\lambda^I,\ \lambda^{II},\ \lambda^{III}) \;\mid\; & \text{admitted deterministically with guaranteed} \\ & (Q^I,\ Q^{II},\ Q^{III}) \text{ and } (T^I,\ T^{II},\ T^{III})\} \end{aligned} \tag{6.1}$$

A yellow region, denoted by Y, can be defined as:

$$\begin{aligned} Y \;=\; \{\; (\lambda^I,\ \lambda^{II},\ \lambda^{III}) \;\mid\; & \text{admitted statistically with guaranteed} \\ & (Q^I,\ Q^{II},\ Q^{III}) \text{ and } (T^I,\ T^{II},\ T^{III})\} \end{aligned} \tag{6.2}$$

A red region, denoted by R, can be defined as:

$$\begin{aligned} R \;=\; \{\; (\lambda^I,\ \lambda^{II},\ \lambda^{III}) \;\mid\; & (Q^I,\ Q^{II},\ Q^{III}) \text{ or/and } (T^I,\ T^{II},\ T^{III}) \\ & \text{cannot be guaranteed}\} \end{aligned} \tag{6.3}$$

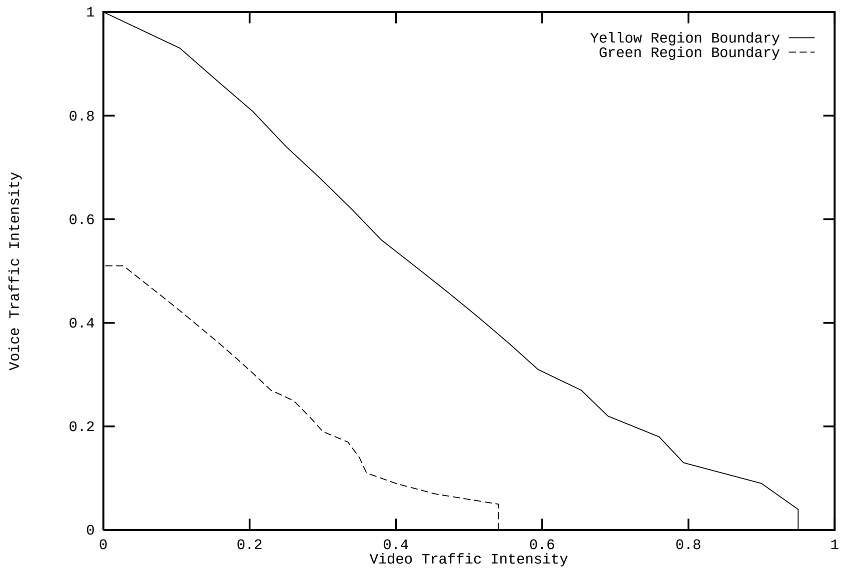


Figure 6.1: The Loading Region for the Joint Scheduling and Admission Control (JSAC) Policy With $QOS\ SET = (1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9})$ for the Video and Voice Traffic Scenario

Figure 6.1 illustrates a loading region for the Joint Scheduling and Admission

Control Policy. Generally, when an admission control policy operates in the green region, the number of calls that can be admitted into the network must be under the upper boundary of the region. The green region shows the proper operating region for an admission control policy that deals with less bursty traffic or continuous-bit-rate traffic. The yellow region indicates the extra traffic load that can be handled by an admission control policy operating in statistical multiplexing mode, compared to the traffic load, shown by the green region, that can be handled by the same admission control policy when it operates in deterministic multiplexing mode. The size and the shape of the yellow region also shows the degree of aggressiveness for an admission control policy. The bigger and higher is the yellow region, the more aggressive an admission control policy is. In any area of the red region, the QOS constraints will not be met. As indicated in Figure 6.1, the size and shape of the three regions (green, yellow, and red) may be significantly different.

The loading region will be used as a metric to quantitatively evaluate and compare the efficiency of different admission control policies.

6.1.2 Illustration

To illustrate the loading region, the traffic intensity has been used to measure the load condition in the network. Given $\mathcal{N}^i$ [1] as the maximum number of calls of Class i and μ^i as the call departure rate for Class i, the Class i traffic intensity is defined as

$$\rho^i = \frac{\lambda^i}{\mathcal{N}^i * \mu^i} \tag{6.4}$$

The total system traffic intensity can be defined as

$$\rho = \rho^I + \rho^{II} \tag{6.5}$$

[1] The value of $\mathcal{N}^i$ in the experiments are defined by the schedulable region from Chapter 5. In general, the value of $\mathcal{N}^i$ can be determined by the ratio of the output link capacity and the mean bit rate of the Class i traffic.

for the video and voice traffic scenario. The percentage of traffic intensity due to Class I is called the *Traffic Intensity Coefficient, c,* and can be expressed by the following equation:

$$c = \frac{\rho^I}{\rho^I + \rho^{II}} = \frac{\rho^I}{\rho} \tag{6.6}$$

where $\rho^I \in [0.0,\ 1.0]$, and $\rho^{II} \in [0.0,\ 1.0]$. For each value of ρ^I, the function

$$\rho^I \ = \ \frac{c}{1-c}\rho^{II} \tag{6.7}$$

is defined as the *Intensity Detecting Line*, where $\rho^I \in [0.0,\ 1.0]$ and $\rho^{II} \in [0.0,\ 1.0]$. The Intensity Detecting Line is a straight line of slope $\frac{c}{1-c}$, passing through the origin.

The loading region is then calculated as follows. The values of μ^I and μ^{II} were fixed at 1.0 throughout the simulation. The values of λ^I and λ^{II} were proportionately raised to increase the total traffic intensity ρ along the Intensity Detecting Line. The intensity ρ^I and ρ^{II} were increased to ρ_0^I and ρ_0^{II}, respectively, to a point where a violation of call-level QOS constraint of blocking probability occurred. The pair $(\rho_0^I,\ \rho_0^{II})$ is called the *Detecting Point*, denoted by t. The procedure was repeated for 21 values of $c = [0.0,\ 0.05,\ 0.10,\ \cdots,\ 1.0]$, which formed 21 different Intensity Detecting Lines, to identify each corresponding Detecting Point, $t_i \in [t_0,\ t_1,\ \cdots,\ t_{20}]$. The t_is are then connected by a line to define the boundary of the loading region. In addition, for every simulation run, a different random seed was given, and five simulation runs were carried out for each loading region. The final loading region plotted is the average from five simulation runs.

The same rules were followed when the loading regions for the other two traffic scenarios (video and data, and data and voice) were plotted.

6.2 Average Expansion Degree

The average expansion degree is a measurement that indicates approximately the average distance between two loading regions on 21 Intensity Detecting lines. If the Average Expansion Degree of the loading region l_1 over the loading region l_2 is greater than 1.0, the loading region l_1 is bigger and higher than the loading region l_2. If the Average Expansion Degree of the loading region l_1 over the loading region l_2 is less than 1.0, the loading region l_2 is smaller and narrower than the loading region l_2.

Let t_i^1 and t_i^2 be the Detecting Point on the Intensity Detecting Line i, $i \in [0, 1, \cdots, 20]$, for the loading region l_1 and the loading region l_2, respectively. The distance between the origin and the Detecting Point t_i^1 is defined as the *Expansion Distance* on the Intensity Detecting Line i for the loading region l_1 , denoted by u_i^1:

$$u_i^1 = \sqrt{(\rho_i^I)^2 + (\rho_i^{II})^2} \tag{6.8}$$

Similarly, the Expansion Distance on the Intensity Detecting Line i for the loading region l_2 is defined as u_i^2.

The *Expansion Degree* of the loading region l_1 over the loading region l_2, denoted by d_i, on the Intensity Detecting Line i is defined as the ratio of the two Expansion Distances u_i^1 and u_i^2:

$$d_i = \frac{u_i^1}{u_i^2} \tag{6.9}$$

The *Average Expansion Degree*, $\overline{d}$, of the loading region l_1 over the loading region l_2 is defined as:

$$\overline{d} = \frac{1}{21}\sum_{i=0}^{20} d_i \tag{6.10}$$

Assume that the two loading regions l_1 and l_2 will not cross with each other. When the Average Expansion Degree, $\overline{d}$, of the loading region l_1 over the loading region l_2 is greater than 1.0, the average Expansion Distance of $(u_0^1, u_1^1, \cdots, u_{20}^1)$ for the loading region l_1 is longer than the average Expansion Distance of $(u_0^2, u_1^2, \cdots, u_{20}^2)$ for the loading region l_2. The loading region l_1 is higher and wider than the loading region l_2. When the Average Expansion Degree, $\overline{d}$, of the loading region l_1 over the loading region l_2 is less than 1.0, the average Expansion Distance of $(u_0^1, u_1^1, \cdots, u_{20}^1)$ for the loading region l_1 is shorter than the average Expansion Distance of $(u_0^2, u_1^2, \cdots, u_{20}^2)$ for the loading region l_2. The loading region l_1 is smaller and narrower than the loading region l_2. When the Average Expansion Degree of the loading region l_1 over the loading region l_2, $\overline{d}$, is equal to 1.0, the loading region l_1 is approximately the same as the loading region l_2.

Let $\overline{d_{1,2}}$ represent the Average Expansion Degree of the loading region l_1 over the loading region l_2; $\overline{d_{3,4}}$ represent the Average Expansion Degree for the loading region l_3 over the loading region l_4. By the definitions above, if

$$\overline{d_{3,4}} > \overline{d_{1,2}} > 1.0$$

then, the distance that the loading region l_3 shifts from the loading region l_4 is greater, or bigger, than the distance that the loading region l_1 shifts from the loading region l_2. Following the same rule, if

$$\overline{d_{3,4}} < \overline{d_{1,2}} < 1.0$$

then, the distance that the loading region l_3 shifts from the loading region l_4 is greater than the distance that the loading region l_1 shifts from the loading region l_2.

The loading region and the Average Expansion Degree are used as the framework in the following sections to evaluate the performance of four different admission control policies quantitatively.

6.3 Deterministic Policy

The principle of bandwidth allocation for the Deterministic Policy is to meet the peak bandwidth need of the traffic, and the total bandwidth allocations for all the traffic in the network can never exceed the capacity of the output link. The features of this policy obtained from simulation are discussed in the following subsections.

6.3.1 Loading Region of the Deterministic Policy

The loading region for the Deterministic Policy is illustrated by using the method discussed in Section 6.1.2. Figure 6.2 presents the simulation results. In the figure, the green region boundary and the yellow region boundary are identical. In other words, the loading region for the Deterministic Policy includes only two sub-regions: the green region and the red region. This observation illustrates the property that the Deterministic Policy does not allow the connections to operate in statistical multiplexing mode. Instead, the Deterministic Policy allocates the bandwidth based on the connection's peak bandwidth requirement, and limits the total bandwidth allocation for all the connections to the output link capacity.

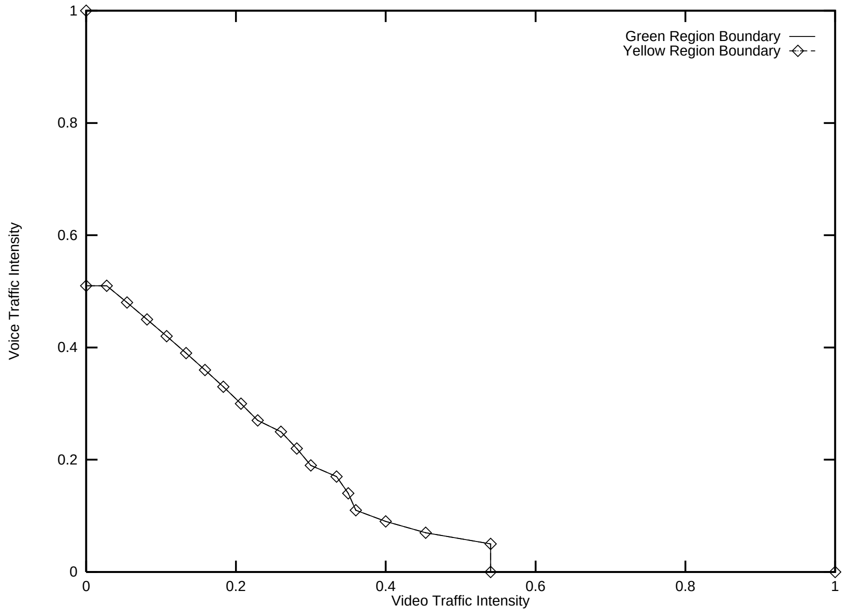


Figure 6.2: The Loading Region for the Deterministic Policy With $QOS\ SET = (1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9})$ for the Video and Voice Traffic Scenario

This property also indicates that no statistical gain can be achieved when the Deterministic Policy is used. The simulation results suggest that the Deterministic Policy is not efficient when the traffic in the network is bursty.

The figure also shows that the maximum traffic intensity for video traffic is only 0.54, and the maximum traffic intensity for voice traffic is only 0.51. The simulation results suggest that the Deterministic Policy operates at a low traffic intensity level.

6.3.2 Sensitivity of the Deterministic Policy to QOS Constraints

The sensitivity of the Deterministic Policy to the QOS constraints is discussed in two parts: the sensitivity to cell delay constraints, and the sensitivity to cell loss constraints.

1. **Sensitivity to Cell Delay Constraints**

   Two sets of QOS constraints were used in the experiments to study the sensitivity of the Deterministic Policy to cell loss constraints. The two sets of QOS constraints are defined in Equations 5.3 and 5.4.

   Figures 6.3 shows the comparison of the loading regions for the Deterministic Policy for the two sets of QOS constraints.

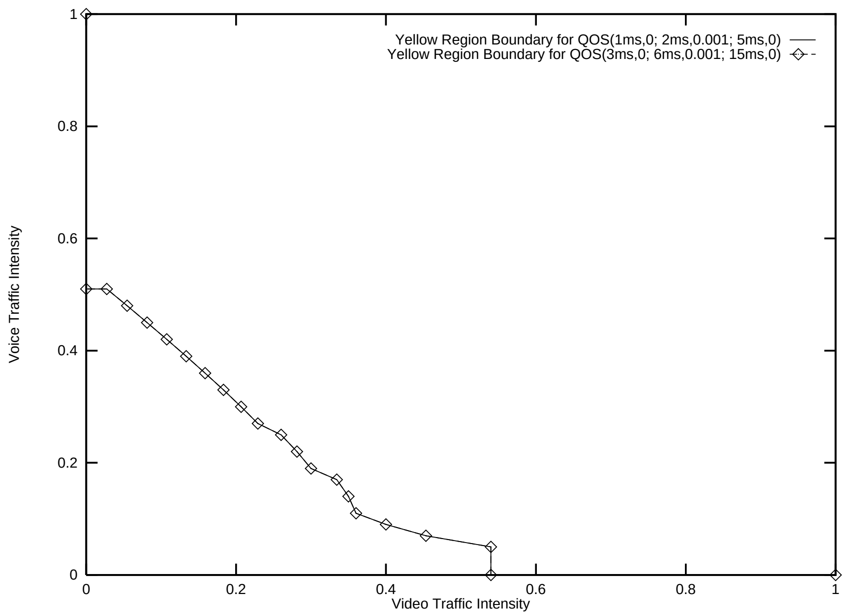


Figure 6.3: The Sensitivity of the Loading Region to the Cell Delay Constraints for the Deterministic Policy for the Video and Voice Traffic Scenario

   In Figure 6.3, the loading region boundary (green boundary) experiences no change at all when the cell delay constraints are relaxed from 1 ms, 2 ms, and 5 ms for the video, voice, and data traffic, respectively, to 3 ms, 6 ms, and 15 ms for video, voice, and data, respectively. The results indicate that the

Deterministic Policy is not sensitive to the cell delay constraints. This property is due to the bandwidth allocation rules of the policy. There is no consideration of the cell delay constraints in the policy: the bandwidth is simply allocated at the peak bandwidth needs for a connection.

2. **Sensitivity to Cell Loss Constraints**

   The second part of the experiments examines the sensitivity of the Deterministic Policy to cell loss constraints. Two relaxed cell loss QOS settings were used in the simulation: $\eta_1^I = 0.0001$ and $\eta_2^I = 0.01$. Figure 6.4 illustrates the comparison of the loading regions for the given cell loss values, along with the loading region for the cell loss $= 10^{-9}$ from Figure 6.2.

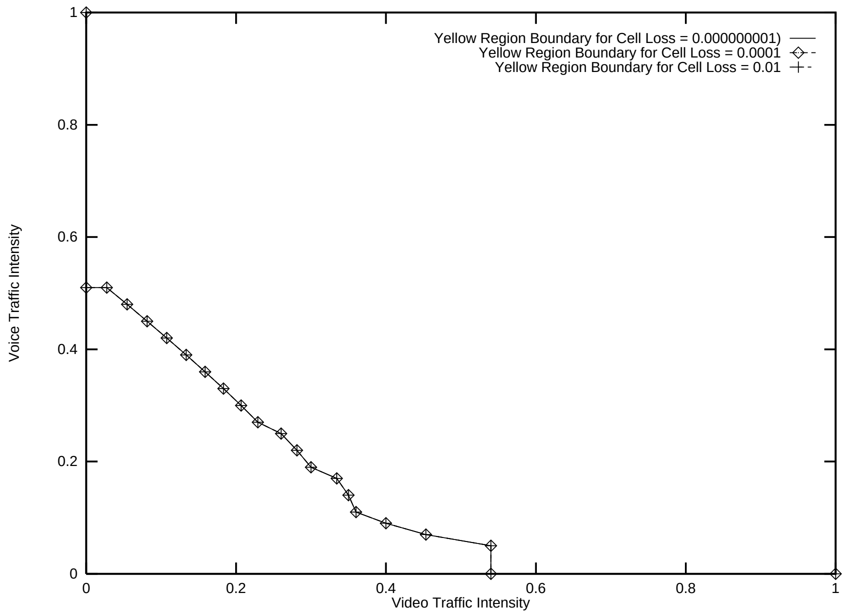


Figure 6.4: The Sensitivity of the Loading Region to the Cell Loss Constraints for the Deterministic Policy for the Video and Voice Traffic Scenario

When the cell loss constraints for video traffic are relaxed from cell loss $= 10^{-9}$, to cell loss $= 0.0001$ and cell loss $= 0.01$, respectively, the loading region boundary for the Deterministic Policy shown in Figure 6.4 remains the same. In other words, the Deterministic Policy is not sensitive to cell loss constraints.

6.3.3 Sensitivity of the Deterministic Policy to Traffic Characteristics

Two video models (listed in Section 4.2.2) are used in the experiments to study the sensitivity of the Deterministic Policy to traffic characteristics.

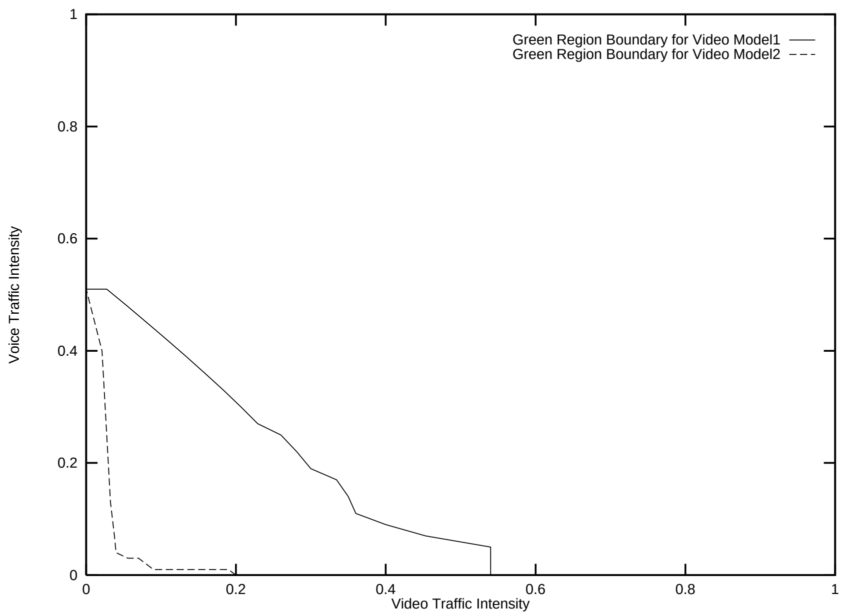


Figure 6.5: The Sensitivity of the Loading Region to the Traffic Characteristics for the Deterministic Policy for the Video and Voice Traffic Scenario

Figure 6.5 compares the loading regions for the two video traffic models. The loading region (green region) for Video Model2 is much smaller and narrower than the one for Video Model1, and the maximum video traffic intensity in Figure 6.5 decreases from 0.54 to 0.19 when the peak bit rate for video traffic increases from $10Mb/s$ to $50Mb/s$. The results suggest that the Deterministic Policy is sensitive to the traffic characteristics.

The comparison shown in Figure 6.5 also demonstrates that the maximum traffic intensity allowed in the network by the Deterministic Policy depends greatly on the traffic characteristics, especially the peak bit rate of the traffic.

6.3.4 Discussion

The results from the previous subsections show that the loading region for the Deterministic Policy has no yellow region, operates in a low level of traffic intensity, is not sensitive to QOS constraints, but is sensitive to traffic characteristics.

For bursty traffic, the Deterministic Policy is not efficient because this policy does not have the statistical multiplexing mode to allow the traffic to share resources on the assumption that each connection only requires its peak bit rate for a small fraction of its time. Therefore, no statistical gain can be achieved by using this policy. When the traffic is less bursty, this policy operates more efficiently, since each connection requires its peak bit rate for a longer time, compared to the burstier traffic. The less bursty the traffic is, the more efficiently this policy operates.

In this policy, only two major factors have been taken into account when bandwidth allocation is made: the peak bit rate and the capacity of the output link. The bandwidth allocation is not adjusted according to the changing of the QOS constraints. In other words, the same amount of bandwidth is allocated to an individual connection regardless of its QOS constraints.

The loading region is somewhat sensitive to changes in traffic characteristics. The higher the peak bit rate of the traffic, the lower the traffic intensity allowed into the network because of the peak bandwidth allocation policy. The mean bit rate of a connection has no impact on the bandwidth allocation rules. The sensitivity of this policy to traffic characteristics emphasizes, again, that bandwidth allocation decision for a connection is based only on the peak bit rate of the traffic and the capacity of the output link.

6.4 Joint Scheduling and Admission Control Policy (JSAC)

The JSAC Policy is an algorithm that combines the scheduling algorithm and admission control policy. Based on the schedulable region predefined and stored in the database, the JSAC Policy admits a connection if and only if the traffic load is within the schedulable region. The evaluation results of the JSAC Policy are discussed in the following subsections.

6.4.1 Loading Region of the JSAC Policy

The loading region for the JSAC Policy is obtained by following the procedure in Section 6.1.2. Figure 6.1 illustrates the results. The loading region shown in the figure is clearly divided into three sub-regions: the green region which is bounded by the dotted line, the yellow region which is bounded by the solid line, and the red region which is above the solid line. The maximum video and voice traffic intensities are 0.54 and 0.51, respectively, for the green region. The maximum video and voice traffic intensities are 0.96 and 1.0, respectively, for the yellow region. When the traffic intensity for video is below 0.54 or the traffic intensity for voice is lower than 0.51 or combined video and voice intensity is lower than the dotted line in the figure, the JSAC Policy operates in the green region. When the traffic intensity in the network is between the dotted line and the solid line, the JSAC Policy is admitting the connections in the statistical multiplexing fashion that is indicated by the yellow region. The JSAC Policy does not allow traffic intensity higher than the limit shown by the solid line (i.e., going into the red region). The yellow region for the JSAC Policy also represents the amount of traffic intensity gains (statistical gain) that can be achieved by admitting the connections statistically.

The loading regions for the other two traffic scenarios (i.e., video and data, and data and voice) were also obtained by following the same procedure. These loading regions are shown in Appendix B.

6.4.2 Sensitivity of the JSAC Policy to QOS Constraints

To explore the sensitivity of the JSAC Policy to QOS constraints, two types of experiments have been carried out: sensitivity to cell delay constraints, and sensitivity to cell loss constraints.

The experiments show that when the two sets of QOS constraints have been used, the green region boundaries corresponding to each set of QOS constraints are the same. In the remaining subsection, only the yellow region boundary will be discussed.

1. **Sensitivity to Cell Delay Constraints**

   Two sets of QOS constraints represented by Equations 5.3 and 5.4 were used in the experiments. Figure 6.6 illustrates the simulation results.

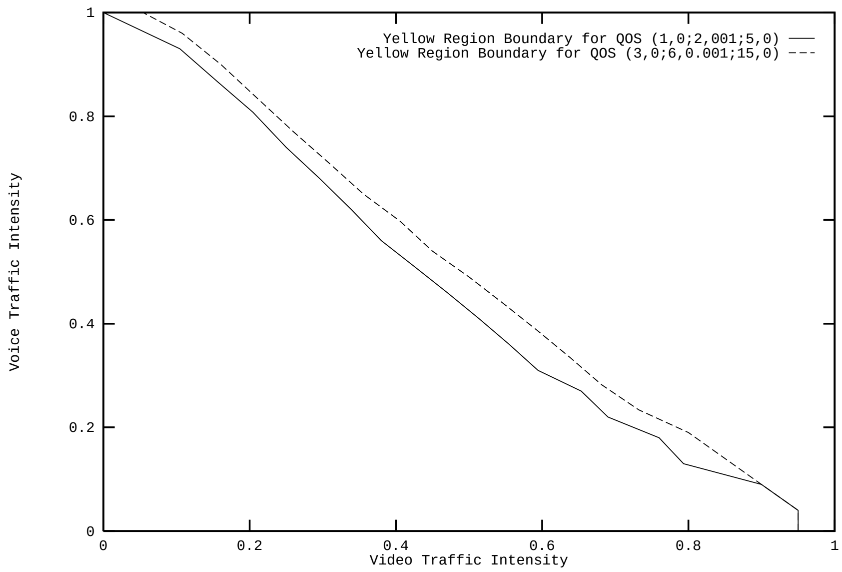


Figure 6.6: The Sensitivity of the Loading Region to the Cell Delay Constraints for the JSAC Policy for the Video and Voice Traffic Scenario

The yellow region boundary corresponding to the $QOS\ SET_2$, represented by the dotted line in the figure, is wider and higher than the one for $QOS\ SET_1$, plotted by the solid line. When the video traffic intensity is higher than 0.9 in the network, the yellow region boundary for the $QOS\ SET_2$ shows no improvement over the one for $QOS\ SET_1$ as the cell delay parameter is relaxed.

The increase of the loading region in size for $QOS\ SET_2$ indicates that the JSAC Policy can admit higher traffic intensity compared to the region with $QOS\ SET_1$. Taking the video traffic intensity = 0.5 as an example, the voice traffic intensity increases from 0.43 to 0.5 when the cell delay parameter relaxed from $QOS\ SET_1$ to $QOS\ SET_2$.

Similar results for traffic scenarios of video and data, and data and voice are shown in Appendix B.

As in the discussion in Section 5.2.1, when $QOS\ SET_2$ applies, the schedulable region for Deadline scheduling algorithm increases in size, and the utility gain increases. The JSAC Policy reflects the schedulable region's sensitivity regarding the cell delay relaxation by widening and enlarging the loading region.

2. **Sensitivity to Cell Loss Constraints**

   The experiments of sensitivity to the cell loss constraints were carried out by using two relaxed cell loss QOS constraint settings for video traffic: $\eta_1^I = 0.0001$ and $\eta_2^I = 0.01$. Figure 6.7 shows the simulation results for the video and voice traffic scenario for the given cell loss parameter values, along with the loading region for cell loss $= 10^{-9}$ from Figure 6.1.

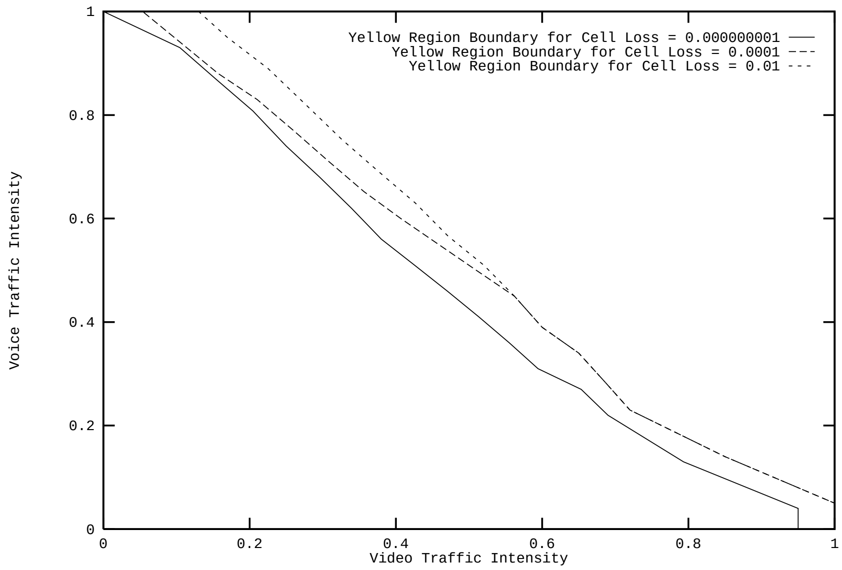


Figure 6.7: The Sensitivity of the Loading Region to the Cell Loss Constraints for the JSAC Policy for the Video and Voice Traffic Scenario

The comparison shows that as the cell loss parameter is relaxed, the loading region increases in size. The more relaxed the cell loss constraint is, the wider and higher the loading region will be.

When the video traffic intensity in the network is higher than 0.55, the loading region does not increase further in size when the cell loss parameter is relaxed from 0.0001 to 0.01. This result matches the trend of the schedulable regions shown in Figure 5.3 in Section 5.2.2.

The comparison results for traffic scenarios of video and data, and data and voice can be found in Appendix B.

6.4.3 Sensitivity of the JSAC Policy to Traffic Characteristics

To study the sensitivity of the JSAC Policy to traffic characteristics, Video Model1 and Video Model2 have been used in the experiments.

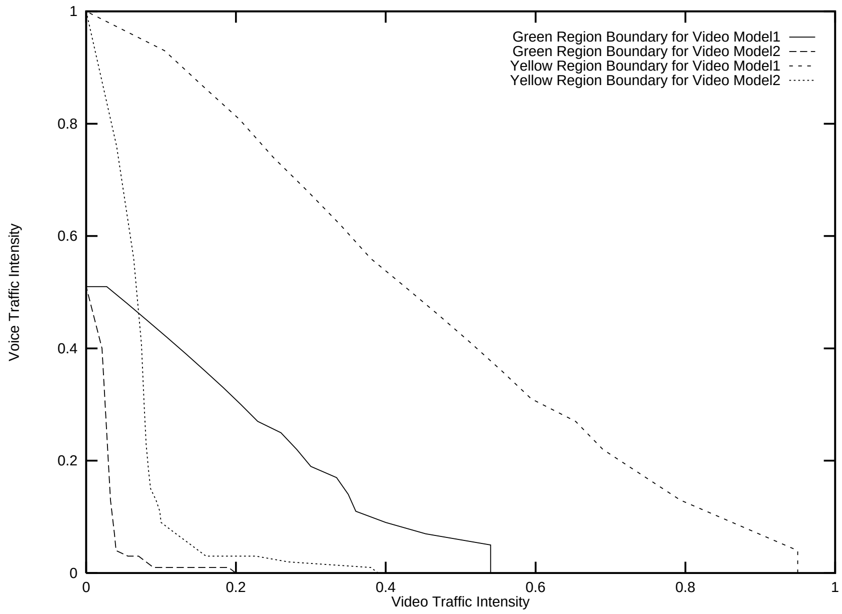


Figure 6.8: The Sensitivity of the Loading Region to Traffic Characteristics for the JSAC Policy for the Video and Voice Traffic Scenario

Figure 6.8 illustrates the simulation results of the sensitivity to traffic characteristics for the video and voice traffic scenario. The following discussions is in two parts: sensitivity of the green region and sensitivity of the yellow region.

1. **Sensitivity of the Green Region to Traffic Characteristics**

   When Video Model2 applies in the experiments, the green region, shown in the figure as the leftmost dashed line, becomes much lower and narrower than the green region for Video Model1, which is represented by the solid line in the figure. When the video traffic intensity increases from 0.0 to 0.073, the maximum allowed voice traffic intensity in the network decreases sharply from 0.51 to 0.015. The maximum voice traffic intensity declines slowly from 0.015 to 0.0 as the video traffic intensity increases from 0.073 to 0.2. The comparison results are consistent with the results for the Deterministic Policy.

2. **Sensitivity of the Yellow Region to Traffic Characteristics**

   A similar trend can be found for the yellow regions when Video Model2 applies. The yellow region is much smaller and narrower for Video Model2 than for Video Model1. A sharp decrease of the maximum voice intensity in the network (from 1.0 to 0.034) occurs when the video traffic intensity increases from 0.0 to 0.16, and is followed by a slow decline from 0.034 to 0.0.

   The consistent changes of the green regions and the yellow regions to different video models suggest that the JSAC Policy is sensitive to traffic characteristics.

6.4.4 Discussion

The simulation results show that the loading region for the JSAC Policy is high and wide and is clearly divided into three sub-regions: the green, the yellow, and the red regions, and the JSAC Policy is sensitive to QOS constraints and traffic characteristics.

The wide and large green and yellow regions for the JSAC Policy indicate that this policy provides the flexibility of supporting both deterministic and statistical multiplexing modes according to the network loading condition and types of the traffic, such as bursty traffic and continuous-bit-rate traffic.

The nature of the JSAC Policy is to combine the scheduling algorithm and the admission control policy so that the maximum traffic intensity can be achieved based on the schedulable region, predefined and stored in the database, which indicates the maximum amount of traffic that can be handled by a switch or a multiplexer for a certain kind of traffic scenario. The high traffic intensity that could be admitted into the network by the JSAC Policy, indicated by the loading region in Figure 6.1, demonstrates this nature well.

The JSAC Policy is sensitive to the QOS constraints, including the cell delay constraints and the cell loss constraints. When the cell delay or the cell loss constraints are relaxed, the yellow region becomes higher, which allows more traffic to be admitted into the network. The shifting trends of the yellow region corresponding to the QOS constraints relaxation match well with the shifting trend for the schedulable regions for QOS constraint relaxation.

The JSAC Policy is also sensitive to traffic characteristics. The loading region, both the green and the yellow region, are much lower and narrower for Video Model2 than the corresponding regions for the Video Model1.

6.5 Effective Bandwidth Policy

The Effective Bandwidth Policy estimates approximate bandwidth needs for each connection, based on the description of peak bit rate, mean bit rate, the burstiness, and the QOS constraints. Bandwidth is allocated according to the estimation. This policy is only suitable for n-state types of traffic. The advantage of this policy is its simple calculation. The complexity of the calculation depends only on the number of states in the traffic models. The properties of this policy are discussed in the following subsections.

6.5.1 Loading Region of the Effective Bandwidth Policy

The loading region for the Effective Bandwidth Policy for the video and voice scenarios is illustrated by following the same procedure discussed in Section 6.1.2. Figure 6.9 shows the results.

In the figure, the loading region has the green, the yellow, and the red region. The yellow region is very narrow and has no significant traffic intensity improvement over the green region when the video traffic intensity is greater than 0.25. The narrow

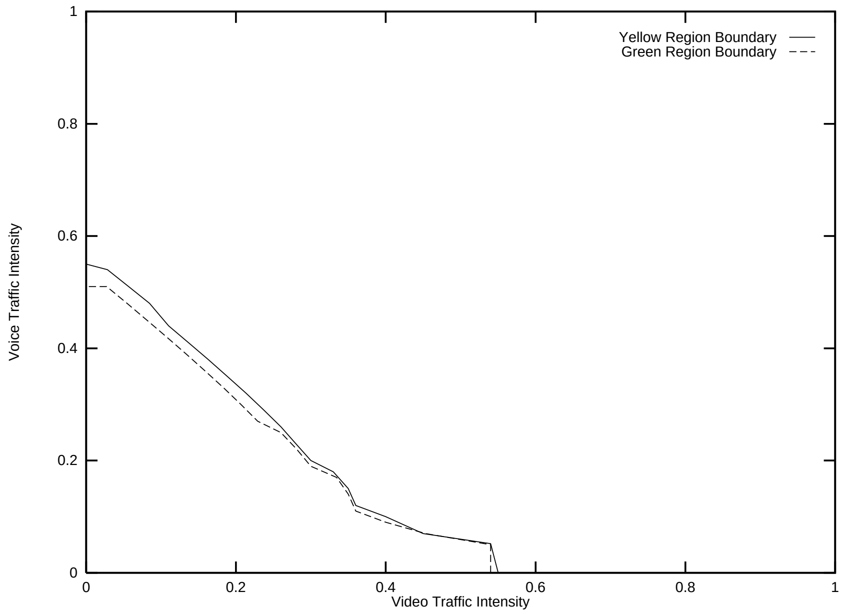


Figure 6.9: The Loading Region for the Effective Bandwidth Policy With $QOS\ SET = (1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9})$ for the Video and Voice Traffic Scenario

yellow region indicates that the Effective Bandwidth Policy, as used in this thesis, is a conservative policy and the estimation of the bandwidth needs for a connection is close to the peak bandwidth of the connection. This observation also suggests that the higher the video traffic intensity is, the closer the yellow region boundary is to the green region.

The loading region for the other two traffic scenarios of video and data, and data and voice were also obtained by following the same procedures. These loading regions are shown in Appendix C.

The loading region of the Effective Bandwidth Policy for traffic scenario of data and voice is also conservative. This result shows that the video model is not the main cause of the conservative effective bandwidth estimations for the other two traffic scenarios of video and voice, and video and data.

6.5.2 Sensitivity of the Effective Bandwidth Policy to QOS Constraints

Two experiments were carried out to explore the sensitivity of the Effective Bandwidth Policy to QOS constraints: sensitivity to cell delay constraints, and sensitivity to cell loss constraints.

1. **Sensitivity to Cell Delay Constraints**

   Two QOS constraints, defined in Equations 5.3 and 5.4 were used in the simulation. Figure 6.10 represents the loading regions for the video and voice traffic scenario with two cell delay constraint settings.

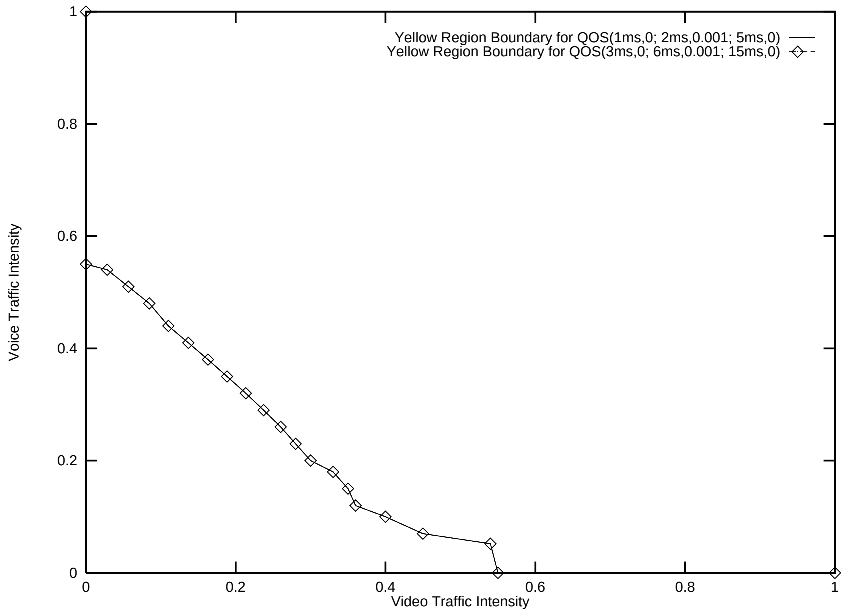


Figure 6.10: The Sensitivity of the Loading Region to the Cell Delay Constraints for the Effective Bandwidth Policy for the Video and Voice Traffic Scenario

   The yellow region boundary has no change when $QOS\ SET_2$ applies, compared to the yellow region boundary for $QOS\ SET_1$. This result indicates that the cell delay relaxation has no impact on the estimation of the bandwidth allocation. This can be seen from Equation 4.5 and the admission criterion ζ, discussed in Section 4.3.2. When the cell delay parameters are relaxed, the buffer sizes and the admission criterion ζ are changed corresponding to the new values of the cell delay parameters. But the changes are not significant enough to reduce the amount of bandwidth allocation in the estimates for each individual connection. As a result, no higher traffic intensity is admitted into the network after all.

The sensitivity of the loading region to cell delay constraints for the other two traffic scenarios of video and data, and data and voice are in Appendix C.

2. **Sensitivity to Cell Loss Constraints**

   Two cell loss QOS constraint settings of $\eta_1^I = 0.0001$ and $\eta_2^I = 0.01$ for video traffic were used in this simulation to assess the sensitivity to cell loss constraints. Figure 6.11 shows the simulation results for the video and voice traffic scenario.

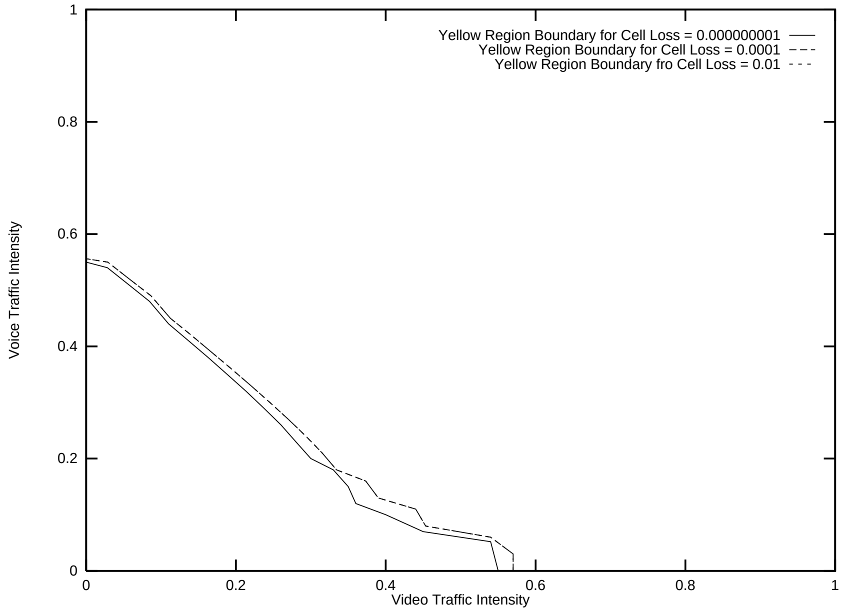


Figure 6.11: The Sensitivity of the Loading Region to the Cell Loss Constraints for the Effective Bandwidth Policy for the Video and Voice Traffic Scenario

Figure 6.11 shows that the yellow region is wider and higher when the cell loss parameter is relaxed from the value 10^{-9} to the value 0.0001. A slight traffic intensity gain is achieved. When the cell loss parameter is relaxed further to 0.01, the yellow region experiences no further change compared to the yellow region for cell loss = 0.0001.

This observation suggests that the Effective Bandwidth Policy is sensitive to the cell loss constraint but not sensitive enough to respond significantly to small changes in the cell loss parameter value.

The simulation results for traffic scenarios of video and data, and data and voice are illustrated in the figures shown in Appendix C.

6.5.3 Sensitivity of the Effective Bandwidth Policy to Traffic Characteristics

The experiments exploring the sensitivity of the Effective Bandwidth Policy to traffic characteristics include two parts: sensitivity of the green region to the traffic characteristics and sensitivity of the yellow region to the traffic characteristics. In the experiment, the two video models given in Section 4.2.2 were used. Figure 6.12 illustrates the simulation results for the video and voice traffic scenario. The results for traffic scenarios of video and data, and data and voice can be found in Appendix C.

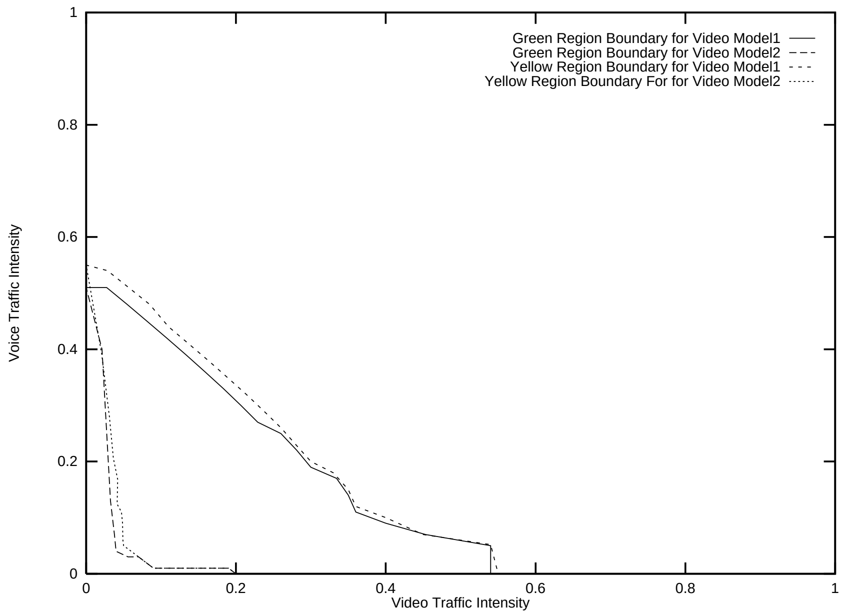


Figure 6.12: The Sensitivity of the Loading Region to the Traffic Characteristics for the Effective Bandwidth Policy for the Video and Voice Traffic Scenario

1. **Sensitivity of the Green Regions to Traffic Characteristics**

   The green region boundary becomes much lower and narrower for Video Model2 than for Video Model1. The maximum video traffic intensity decreases from 0.54 to 0.2. The maximum voice traffic intensity declines from 0.51 to 0.01 when the video traffic intensity increases from 0.0 to 0.09, and the maximum voice traffic intensity stays near 0.01 when the video traffic intensity increases from 0.09 to 0.2.

The declining trend of the green region for Video Model2 is similar with the one for the JSAC Policy, as discussed in the previous section.

2. **Sensitivity of the Yellow Region to Traffic Characteristics**

   The yellow region for Video Model2 is much smaller and narrower than the one for Video Model1. As the video traffic intensity in the network increases, a sharp decline of the yellow region for Video Model2 occurs, compared to a slow decline in the yellow region for Video Model1.

   The yellow region for Video Model2 is very narrow and very small, with no significant traffic intensity improvement when the voice traffic intensity is between 0.3 and 0.51, and when the video traffic intensity is greater than 0.07.

   The similarity of the sensitivity of the green regions and the yellow regions to traffic model changes among the loading region for the three policies suggests that the peak bit rate of an individual connection has extremely important impact on the maximum allowed traffic intensity in the network.

6.5.4 Discussion

The simulation results show that the loading region for the Effective Bandwidth Policy has a narrow yellow region, and the Effective Bandwidth Policy is not sensitive to the cell delay constraints but is sensitive to the cell loss constraints and traffic characteristics.

The narrow yellow region suggests that the Effective Bandwidth Policy, as used in this thesis, is a conservative policy that does not offer much operating margin for the policy to admit bursty traffic statistically. The estimated bandwidth need for an individual connection is close to its peak bandwidth. There is not much traffic intensity improvement compared to the green region, especially when the video traffic intensity is greater than 0.245.

The Effective Bandwidth Policy is not very sensitive to cell delay constraints. Although the cell delay constraints have been considered as a factor to bound the buffer size for the bandwidth estimates, the response to the cell delay changes in the formula has little impact on the decision making.

The Effective Bandwidth Policy is sensitive to the cell loss constraints, but not sensitive enough to respond to small changes in the cell loss value.

The Effective Bandwidth Policy is sensitive to the traffic characteristics. When the peak bit rate of the video traffic changes from 10Mb/s to 50Mb/s, both the green and the yellow region become much narrower and smaller. This result is similar to the observation for the JSAC Policy for video model changes.

6.6 Virtual Bandwidth Policy

The Virtual Bandwidth Policy is based on virtual cell loss probability estimates using a bufferless model. The following subsections present the properties of this admission control policy.

6.6.1 Loading Region of the Virtual Bandwidth Policy

The loading region for the Virtual Bandwidth Policy is obtained using the procedure discussed in Section 6.1.2. Figure 6.13 is the illustration of the simulation results for the video and voice traffic scenario.

In the figure, the loading region is divided into three sub-regions: the green, the yellow, and the red regions. The yellow region is fairly wide and high, with the maximum video traffic intensity 0.67 and the maximum voice traffic intensity 0.95. The large yellow region indicates that the Virtual Bandwidth Policy supports bursty traffic well by allowing the connections into the network statistically, and has great potential to improve the overall traffic intensity in the network.

The calculation of the virtual cell loss is accurate, but the complexity of the

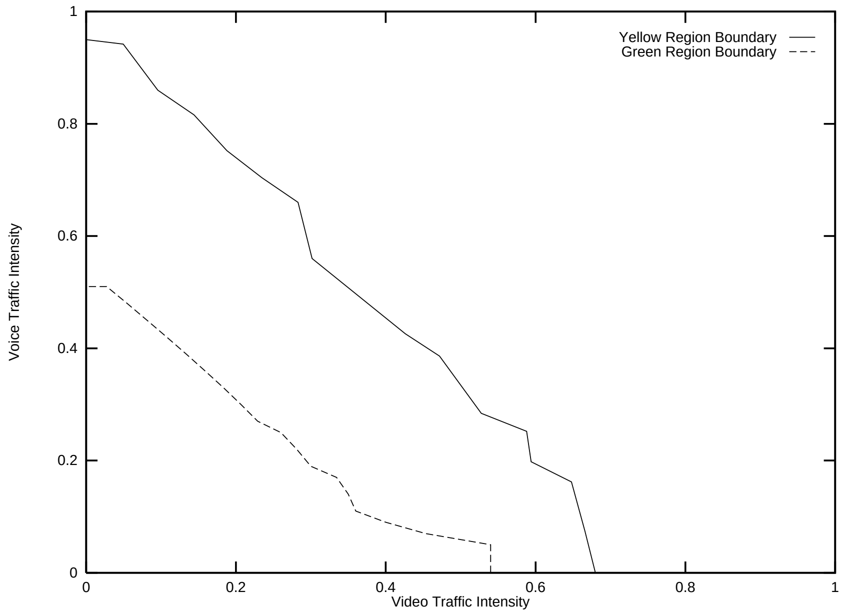


Figure 6.13: The Loading Region for the Virtual Bandwidth Policy With $QOS\ SET = (1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9})$ for the Video and Voice Traffic Scenario

calculation depends on how many connections are in the network. In other words, the more connections there are in the network, the more complicated the calculation is. In the simulation, the time for a decision on acceptance of a connection needs approximately four times as much as for the other policies.

The loading regions for traffic scenarios of video and data, and data and voice can be found in Appendix D.

6.6.2 Sensitivity of the Virtual Bandwidth Policy to QOS Constraints

Similar to the Deterministic Policy, the Virtual Bandwidth Policy does not consider the cell delay constraints in the connection acceptance. When the cell delay constraints are relaxed, the loading region remains the same. The remaining discussion focuses only on the sensitivity of the Virtual Bandwidth Policy to cell loss constraints.

Two sets of cell loss constraints, $\eta_1^I = 0.0001$ and $\eta_2^I = 0.01$, are used in the experiments to assess the sensitivity of the Virtual Bandwidth Policy to the cell loss constraints. Figure 6.14 shows the simulation results, along with the yellow region for cell loss = 10^{-9} in Figure 6.12, for the video and voice traffic scenario.

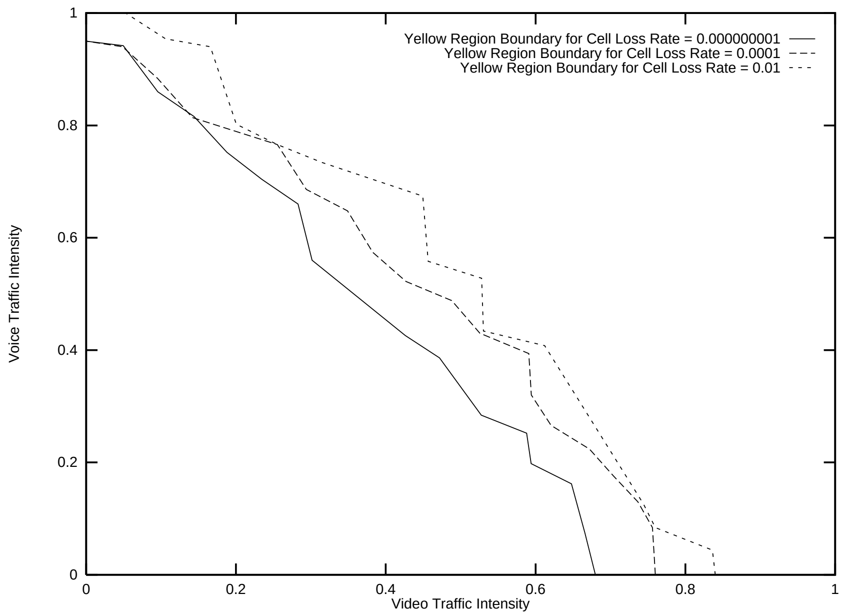


Figure 6.14: The Sensitivity of the Loading Region to the Cell Loss Constraints for the Virtual Bandwidth Policy for the Video and Voice Traffic Scenario

The yellow region is much higher and wider when the cell loss constraint is relaxed from 10^{-9} to 0.0001, particularly when the video traffic intensity is higher than 0.2 in the network. While the maximum voice traffic intensity remains the same, the maximum video traffic intensity increases from 0.67 to 0.76 when the cell loss parameter is relaxed from 10^{-9} to 0.0001. When the cell loss parameter is further relaxed from 0.0001 to 0.01, the yellow region is larger and wider than the one for the cell loss = 0.0001. The maximum video traffic intensity increases from 0.76 to 0.84, and the maximum voice traffic intensity increases from 0.95 to 1.0, compared to the yellow region for cell loss = 0.0001. The results show that the Virtual Bandwidth Policy is very sensitive to the cell loss constraints.

The sensitivity of the loading region for the Virtual Bandwidth Policy to the cell loss constraints for traffic scenarios of video and data, and data and voice are listed in Appendix D.

6.6.3 Sensitivity of the Virtual Bandwidth Policy to Traffic Characteristics

The two video models from Section 4.2.2 are used in the experiments. Figure 6.15 presents the simulation results for the traffic scenario of video and voice. The discussions is in two parts: sensitivity of the green region to traffic models, and sensitivity of the yellow region to traffic models.

The simulation results of sensitivity of the Virtual Bandwidth Policy to traffic characteristics for traffic scenarios of video and data, and data and voice are listed in Appendix D.

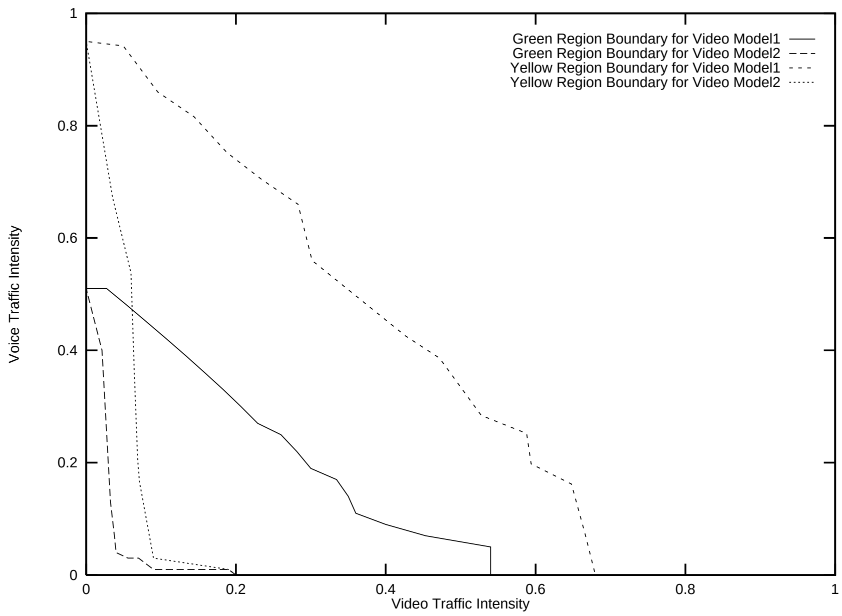


Figure 6.15: The Sensitivity of the Loading Region to the Traffic Characteristics for the Virtual Bandwidth Policy for the Video and Voice Traffic Scenario

1. **Sensitivity of the Green Region to Traffic Characteristics**

   The green region for Video Model2 is much narrower and smaller than the one for Video Model1, as shown in Figure 6.15. The great change of the green region in size and shape when Video Model2 applies demonstrates that the total traffic intensity allowed in the network will be reduced greatly if some of the traffic has a very high peak bit rate. The sensitivity of the green regions to the traffic models is similar to the other three admission policies.

2. **Sensitivity of the Yellow Region to Traffic Characteristics**

   Similar to the changes for the green regions after Video Model2 is applied, the yellow region becomes very narrow and small. The maximum video traffic intensity decreases sharply from 0.67 to 0.2 when the video traffic model changed from Video Model1 to Video Model2.

   The same changing trend when different video model applies emphasizes that the peak bit rate of the traffic has great impact on the traffic intensity allowed in the network.

6.6.4 Discussion

The experiment results show that the loading region for the Virtual Bandwidth Policy is high and wide and clearly divided into three sub-regions: the green, the yellow, and the red regions, and the Virtual Bandwidth Policy is not sensitive to the cell delay constraints but is very sensitive to the cell loss constraints and traffic characteristics.

The loading region of the Virtual Bandwidth Policy shows that this policy supports bursty traffic well by admitting the connections statistically. A high traffic intensity is achieved, as illustrated by the wide and large yellow region in Figure 6.13. This policy is also flexible enough to operate both deterministically and statistically.

The Virtual Bandwidth Policy is not sensitive to the cell delay constraints because it does not take the cell delay as a factor for the virtual cell loss estimates. For two identical types of traffic with different cell delay constraints, the bandwidth allocation is the same. The disadvantage of this type of acceptance decision making is that the priority of the connections based on the cell delay constraints is ignored by this policy.

The sensitivity of the Virtual Bandwidth Policy to the cell loss constraints are shown by the large shifting of the yellow region boundary when the cell loss parameters are relaxed from 10^{-9} to 0.0001 to 0.01. This change in the size and shape of the yellow region indicates that more traffic can be admitted into the network to share the resources when the cell loss constraints are not stringent.

The sensitivity of the Virtual Bandwidth Policy to the traffic characteristics shows the same characteristics as the other admission control policies. This similarity suggests that the peak bit rate of an individual connection has a large influence on the acceptance decision making.

6.7 Comparison of the Four Admission Control Policies

The four admission control policies have been discussed individually in the previous sections. This section focuses on direct comparisons of the four policies, including how aggressive a policy is, how sensitive a policy is to the QOS constraints or to the traffic characteristics, and how well the channel utilization achieved by a policy is, amongst the four. To mathematically express the difference between two loading region boundaries, the Average Expansion Degree is used in the following subsections.

6.7.1 Aggressiveness of Admission Control Policies

The yellow regions for the four admission control policies are illustrated in Figure 6.16. In the figure, the yellow region for the JSAC Policy is much higher and wider than the yellow regions for the other three policies. Therefore, the JSAC Policy provides the highest and most flexible traffic intensity to the connections. The yellow region for the Virtual Bandwidth Policy is much bigger and higher than the yellow regions for the Effective Bandwidth Policy and the Deterministic Policy. The Deterministic Policy has the lowest and narrowest yellow region amongst the four policies. By using the yellow region for the Deterministic Policy as a basis, the Average Expansion Degree of the other three yellow regions over the yellow region for the Deterministic Policy can be calculated to mathematically compare the aggressiveness of the loading regions.

The Average Expansion Degree of the yellow region for the JSAC Policy over the yellow region for Deterministic Policy, $\overline{d_{J,D}}$, is 1.82. The Average Expansion Degree of the yellow region for Virtual Bandwidth Policy over the Deterministic Policy, $\overline{d_{V,D}}$, is 1.77. The Average Expansion Degree of the yellow region for Effective

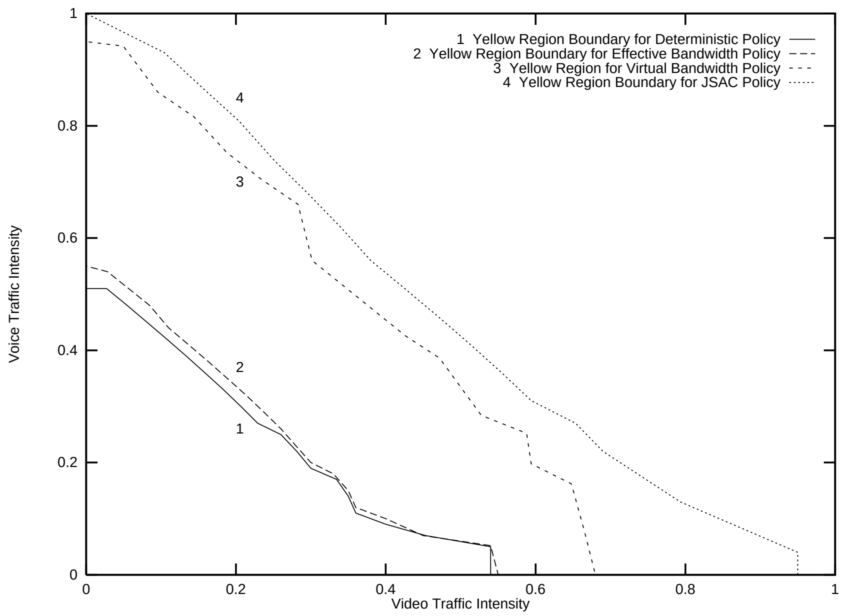


Figure 6.16: The Comparison of the Yellow Regions for Four Admission Control Policies with QOS $(1ms, 10^{-9}; 2ms, 0.001; 5ms, 10^{-9})$ for Video and Voice Traffic Scenario

Bandwidth Policy over the Deterministic Policy, $\overline{d_{E,D}}$, is 1.06. The JSAC Policy is the most aggressive policy amongst the four policies with the highest $\overline{d_{J,D}} = 1.82$. The Virtual Bandwidth Policy is second most aggressive policy with the $\overline{d_{V,D}} = 1.77$. The Effective Bandwidth Policy is only more aggressive than the Deterministic Policy with $\overline{d_{E,D}} = 1.06$. The Deterministic Policy is the least aggressive policy amongst the four policies.

6.7.2 Sensitivity to QOS Constraints

The sensitivity of individual admission control policies to QOS constraints have been discussed in the previous sections. This subsection focuses on a mathematical comparison of how sensitive each of the four admission control policies are to the QOS constraints. The discussion is in two parts: sensitivity to cell delay, and sensitivity to cell loss.

1. **Sensitivity to Cell Delay Constraints**

   To compare how sensitive an admission control policy is to the cell delay constraints, the Average Expansion Degree of the Yellow region for $QOS\ SET_2$

over the yellow region for $QOS\ SET_1$ ($\overline{d_D}$, $\overline{d_J}$, $\overline{d_E}$, $\overline{d_V}$), is calculated for the Deterministic Policy, the JSAC Policy, the Effective Bandwidth Policy, and the Virtual Bandwidth Policy, respectively. Table 6.1 lists the results.

Table 6.1: Average Expansion Degree for Cell Delay Sensitivity for the Four Admission Control Policies

Admission Control Policy	Average Expansion Degree
Deterministic Policy	$\overline{d_D} = 1.0$
JSAC Policy	$\overline{d_J} = 1.06$
Effective Bandwidth Policy	$\overline{d_E} = 1.0$
Virtual Bandwidth Policy	$\overline{d_V} = 1.0$

The JSAC Policy is the only policy that shows sensitivity to the cell delay constraint. This conclusion can be drawn from both the loading region figures and Table 6.1 of the Average Expansion Degree.

The Deterministic Policy and the Virtual Bandwidth Policy don't consider the cell delay constraint as a factor to the call acceptance decision making. The total traffic that can be admitted by these two policies into the network for a relaxed cell delay constraint is identical with the total traffic for a non-relaxed cell delay constraint.

The Effective Bandwidth Policy takes the cell delay constraint into account when the call acceptance decision is made, but fails to allocate the bandwidth different when the cell delay constraint is relaxed.

The insensitivity of an admission control policy to the cell delay constraint shows the failure of responding to the different cell delay requirements of multimedia traffic. Therefore, no traffic intensity improvement and channel utilization gain is achieved when the cell delay constraint is relaxed.

2. **Sensitivity to Cell Loss Constraints**

   The Average Expansion Degree (AED) of the yellow region for the cell loss = 0.0001 over the yellow region for the cell loss = 10^{-9}, and the Average Expansion Degree of the yellow region for the cell loss = 0.01 over the yellow region for the cell loss = 0.0001, is calculated for each admission control policy. Table 6.2 lists the results.

Table 6.2: Average Expansion Degree for Cell Loss Sensitivity for the Four Admission Control Policies

Admission Control Policy	AED (0.0001:10^{-9})	AED (0.01:0.0001)
Deterministic Policy	$\overline{d_D} = 1.0$	$\overline{d_D} = 1.0$
JSAC Policy	$\overline{d_J} = 1.056$	$\overline{d_J} = 1.034$
Effective Bandwidth Policy	$\overline{d_E} = 1.071$	$\overline{d_E} = 1.0$
Virtual Bandwidth Policy	$\overline{d_V} = 1.094$	$\overline{d_V} = 1.057$

When the cell loss constraint is relaxed from 10^{-9} to 0.0001, all policies except the Deterministic Policy are sensitive to the cell loss relaxation. The Virtual Bandwidth Policy is the most sensitive policy towards the cell loss relaxation with $\overline{d_V} = 1.094$, and the JSAC Policy is the least sensitive policy among the three with $\overline{d_J} = 1.056$.

When the cell loss constraint is relaxed further from 0.0001 to the 0.01, only two policies, the JSAC Policy and the Virtual Bandwidth Policy, are sensitive to the cell loss relaxation. The Virtual Bandwidth Policy is more sensitive to the cell loss relaxation ($\overline{d_V} = 1.057$) than the JSAC Policy ($\overline{d_J} = 1.034$).

These three policies, the JSAC Policy, the Effective Bandwidth Policy, and the Virtual Bandwidth Policy, are more sensitive to the first cell loss constraint relaxation (from 10^{-9} to 0.0001) than to the second cell loss constraint (from 0.0001 to 0.01). This observation suggests that the degree of the sensitivity of an admission control policy matches the degree of the cell loss constraint relaxation. For the first cell loss constraint relaxation, the degree of the cell loss constraint is $0.0001/10^{-9} = 10^5$; and the degree of the cell loss constraint is $0.01/0.0001 = 10^2$ for the second cell loss constraint relaxation. The degree of the sensitivity to the first and second cell loss constraint relaxation are $\overline{d_J} = 1.056$ and $\overline{d_J} = 1.034$, respectively, for the JSAC Policy; are $\overline{d_E} = 1.071$ and $\overline{d_E} = 1.0$, respectively, for the Effective Bandwidth Policy; and are $\overline{d_V} = 1.091$ and $\overline{d_V} = 1.057$, respectively, for the Virtual Bandwidth Policy. In other words, the more relaxed the cell loss constraint is, the more sensitive an admission control policy is to the relaxation. The less relaxed the cell loss constraint is, the less sensitive an admission control policy is to the relaxation.

The relaxation of the cell loss constraint allows the admission control policies to admit more traffic into the network and achieves higher traffic intensity gains and higher channel utilization.

6.7.3 Sensitivity to Traffic Characteristics

To quantitatively compare sensitivity of the four admission control policies, the Average Expansion Degree of the yellow region for Video Model2 over the yellow region for Video Model1 is calculated for each of the four admission control policies. Table 6.3 shows the results.

Table 6.3: Average Expansion Degree for Traffic Characteristics Sensitivity for the Four Admission Control Policies

Admission Control Policy	Average Expansion Degree
Deterministic Policy	$\overline{d_D} = 0.266$
JSAC Policy	$\overline{d_J} = 0.317$
Effective Bandwidth Policy	$\overline{d_E} = 0.271$
Virtual Bandwidth Policy	$\overline{d_V} = 0.233$

The four policies are all sensitive to the traffic characteristics. Each shows a dramatic decrease in the yellow region when Video Model2 is used. The Virtual Bandwidth Policy has the most sensitivity to the traffic characteristics among the four policies with $\overline{d_V} = 0.233$. The JSAC Policy is the least sensitive policy towards to the traffic characteristics with $\overline{d_J}$ value 0.317.

The sensitivity of an admission control policy to Traffic characteristics indicates that much lower traffic intensity can be admitted into the network when the peak bit rate for video traffic is increased from 10 Mb/s to 50 Mb/s, even though the mean bit rate for video traffic is remained the same.

6.7.4 Channel Utilization

The average channel utilization for each policy has been monitored in the simulation. Figure 6.17 presents the average channel utilization as a function of the percentage of video traffic in the network for the four admission control policies.

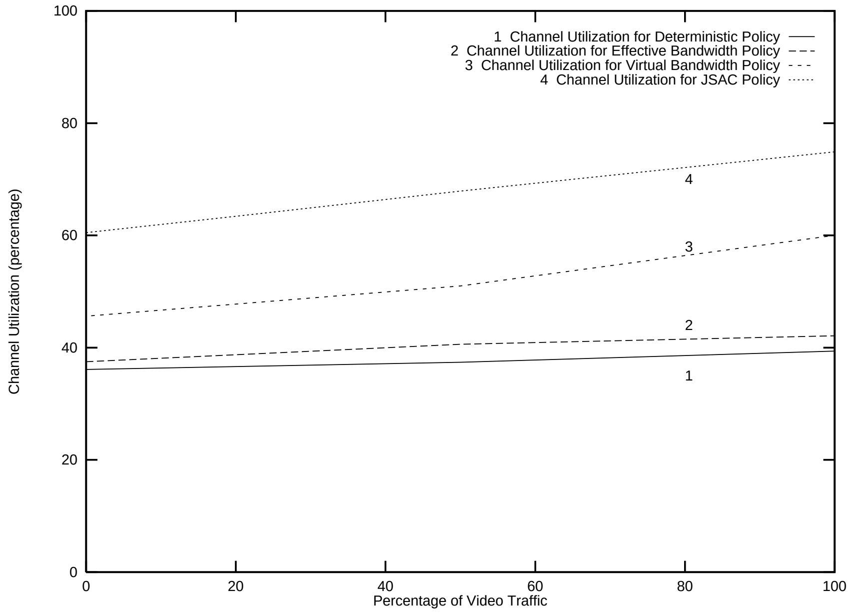


Figure 6.17: Channel Utilization for the Four Admission Control Policies for the Video and Voice Traffic Scenario

The JSAC Policy achieves the highest channel utilization among the four policies. When no video traffic is in the system (zero percent of video traffic), the channel utilization is 60.4%. The channel utilization increases steadily to 74.9% when the video traffic achieves 100%, for the JSAC Policy. The more video traffic in the system, the higher the channel utilization achieved. The figure shows the same trend for the other three policies. The Deterministic Policy has the lowest channel utilization with 36.1% channel utilization for zero percent of video traffic and 39.4% channel utilization for 100% video traffic.

The channel utilization performance achieved by the four admission control policies match well with the aggressiveness of the four policies (Section 6.6.1). The most aggressive policy, the JSAC Policy, has the highest channel utilization. The second most aggressive policy, the Virtual Bandwidth Policy, achieves higher channel utilization than the Effective Bandwidth and the Deterministic Policies do. The Deterministic Policy, the least aggressive policy, has the lowest channel utilization.

6.8 Discussion

The comparison results discussed in the previous section show that the JSAC Policy is the most aggressive policy amongst the four admission control policies with the highest Average Expansion Degree of 1.82. The Deterministic Policy is the most conservative policy among the four policies with the lowest Average Expansion Degree of 1.0.

Table 6.4 summarizes the sensitivity of the four policies to QOS constraints and traffic characteristics. The JSAC Policy is the only policy that shows sensitivity

Table 6.4: Sensitivity of the Four Policies to QOS Constraints and Traffic Characteristics

Admission Control Policy	AED Delay	AED Loss		AED Traffic
		0.0001:10^{-9}	0.01:0.0001	
Deterministic Policy	$\overline{d_D} = 1.0$	$\overline{d_D} = 1.0$	$\overline{d_D} = 1.0$	$\overline{d_D} = 0.266$
JSAC Policy	$\overline{d_D} = 1.06$	$\overline{d_D} = 1.056$	$\overline{d_D} = 1.034$	$\overline{d_J} = 0.317$
Effective Bandwidth Policy	$\overline{d_D} = 1.0$	$\overline{d_D} = 1.071$	$\overline{d_D} = 1.0$	$\overline{d_E} = 0.271$
Virtual Bandwidth Policy	$\overline{d_D} = 1.0$	$\overline{d_D} = 1.094$	$\overline{d_D} = 1.057$	$\overline{d_V} = 0.233$

to cell delay constraint with the Average Expansion Degree greater than 1.0. The Virtual Bandwidth Policy is the most sensitive policy to cell loss constraint and traffic characteristics.

6.9 Summary

The loading region has been introduced in this chapter to quantitatively measure the performance of the four admission control policies. The size and shape of the loading region represent the aggressiveness of an admission control policy. The four admission control policies have been evaluated individually and compared with the other policies, in the aspects of sensitivity to the QOS constraints and traffic characteristics. The

JSAC Policy is the most aggressive policy of the four policies. The summary of the thesis and the conclusions drawn from the experiments are discussed in Chapter 7.

Chapter 7

Summary and Conclusions

In this final chapter, a summary of the thesis is presented along with its conclusions. The suggestions for future work are in the final section.

7.1 Summary

The research carried out in this thesis focused on two main objectives:

1. to establish a performance evaluation method for admission control policies
2. to use this method to evaluate different admission control policies for high speed networks.

Chapter 4 presented the experimental methodology, including the network model, the traffic models, the scheduling algorithms, and the admission control policies studied in this thesis, upon which a series of experiments were undertaken.

The statistical behaviour of scheduling algorithms has been studied as part of the work for the thesis. Understanding the statistical behaviour of scheduling algorithms forms the basis that leads to the further study of the JSAC Policy. A scheduling algorithm that operates in statistical multiplexing mode can handle much higher traffic load and achieve much higher channel utilization than a scheduling algorithm that operates in deterministic multiplexing mode. For the scheduling algorithm that operates in statistical multiplexing mode, the burstier the traffic is, the higher the channel utilization that can be achieved. The amount of traffic sources that can be handled by the statistical scheduling algorithm varies depending on different QOS requirements and traffic characteristics.

The loading region and the Average Expansion Degree are introduced in this thesis to evaluate quantitatively different admission control policies. A loading region indicates the number of calls of heterogeneous traffic that can be admitted into the network. The Average Expansion Degree is a measurement that indicates approximately the average distance between two loading regions. The loading region and the Average Expansion Degree are used as the framework in the experiments to evaluate aggressiveness and sensitivity of the four different admission control policies, individually and compared with the other policies.

The four admission control policies, the Deterministic Policy, the JSAC Policy, the Effective Bandwidth Policy, and the Virtual Bandwidth Policy, are evaluated in this thesis, individually and compared with the other policies, in terms of aggressiveness, sensitivity to the QOS constraints, and sensitivity to traffic characteristics. The JSAC Policy is the most aggressive policy and achieves the highest channel utilization among the four policies. Different characteristics are observed when sensitivity to cell delay constraint, and to cell loss constraint are assessed, respectively, for the four admission control policies. Only the JSAC Policy is sensitive to the cell delay constraint and all policies except the Deterministic Policy are sensitive to the cell loss constraint. All four admission control policies are sensitive to traffic characteristics. This observation suggests that the peak bit rate of an individual connection has great influence on the acceptance decision making.

7.2 Conclusions

The research presented in this thesis leads to the following conclusions.

1. The loading region is a powerful method to evaluate the performance of admission control policies. The size and shape of the yellow region can represent the aggressiveness of an admission control policy. The higher and wider the yellow

region is, the more aggressive is the admission control policy. The loading region captures the sensitivity of an admission control policy to QOS constraints and traffic characteristics, by showing the shifting of the region after the restriction changes. The Average Expansion Degree is a quantitative measurement that shows how sensitive and aggressive an admission control policy is compared to the other policies.

2. The different admission control policies have very different loading regions. An admission control policy should be flexible enough to allow more bursty traffic being admitted into the network so that higher traffic intensity and channel utilization can be achieved; and to respond to the different QOS requirements so that more traffic can be admitted into the network when the QOS constraints are relaxed. The yellow region indicates the extra traffic intensity that can be admitted into the network statistically, compared to the traffic intensity indicated by the green region. The higher and wider is the yellow region, the more traffic intensity is admitted into the network; the more aggressive is the admission control policy; and the higher the channel utilization is achieved.

3. The peak bit rate of an individual traffic has great impact on the call acceptance decision making. The higher is the peak bit rate, the smaller and narrower is the loading region (including the green and the yellow region).

7.3 Future Work

The research presented in this thesis can be extended in many ways. First, the traffic scenario of three dimensions or more can be used in the experiments to explore the characteristics of the admission control policies under more complicated traffic mix environment.

Second, other admission control policies can be evaluated in the experiments to examine the capability of the loading region and the Average Expansion Degree as a quantitative evaluation method.

Finally, an extended definition for the Average Expansion Degree can be established to quantitatively measure two loading regions that cross with each other, so that the aggressiveness of the admission control policy can be accurately calculated and assessed.

Bibliography

[1] A. Baiocchi, N. Melazzi, M. Listanti, A. Roveri, and R. Winkler, "Modeling Issues on an ATM Multiplexer Within A Bursty Traffic Environment," *Proceedings IEEE INFOCOM'91*, Bar Harbour, Florida, pp 83-91, April 1991.

[2] R. Bolla and F. Davoli, "Dynamic Hierarchical Control of Resource Allocation in an Integrated Services Broadband Network," *Computer Networks and ISDN Systems 25 (1993)*, pp 1079-1087, 1993.

[3] R. Chipalkatti, J. Kurose, and D. Towsley, "Scheduling Policies for Real-Time and Non-Real-Time Traffic in a Statistical Multiplexer," *Proceedings IEEE INFOCOM'89*, Ottawa, Canada, pp 774-783, April 1989.

[4] A. I. Elwalid and D. Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High-Speed Networks," *IEEE/ACM Transactions on Networking*, Vol. 1, No. 3, pp 329-343, June 1993.

[5] A. I. Elwalid and D. Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High-Speed Networks," *IEEE INFOCOM'93*, San Francisco, CA, pp 256-265, March 1993.

[6] J. Ferrandiz and A. Lazar, "Admission Control for Real-Time Packet Sessions," *Proceedings IEEE INFOCOM' 91*, Bar Harbour, Florida, pp 553-559, April 1991.

[7] A. Gersht and K. Lee, "A Congestion Control Framework for ATM Networks," *Proceedings IEEE INFOCOM'89*, Ottawa, Canada, pp 774-783, April 1989.

[8] H. Gilbert, O. Aboul-Magd, and V. Phung, "Developing a Cohesive Traffic Management Strategy for ATM Networks," *IEEE Communications Magazine*, Vol. 29, No. 10, pp 36-45, October 1991.

[9] S. Golestani, "Congestion-Free Transmission of Real-Time Traffic in Packet Networks," *Proceedings IEEE INFOCOM' 90*, San Diego, CA, pp 527-536, December 1990.

[10] I. Habib and T. Saadawi, "Controlling Flow and Avoiding Congestion in Broadband Networks," *IEEE Communications Magazine*, Vol. 29, No. 10, pp 46-53, October 1991.

[11] I. Habib and T. Saadawi, "Multimedia Traffic Characteristics in Broadband Networks," *IEEE Communications Magazine*, Vol. 30, No. 7, pp 48-54, July 1992.

[12] D. Heyman, A. Tabatabai, and T. Lakshman, "Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 2, No. 1, pp 49-59, March 1992.

[13] D. Hong and T. Suda, "Congestion Control and Prevention in ATM Networks," *IEEE Network Magazine*, Vol. 5, No. 4, pp 10-15, July 1991.

[14] J. Huang and S. Lee, "MHTP - A Multimedia High-speed Transport Protocol," *Proceedings IEEE Globecom '92*, Orlando, Florida, pp 1364-1368, December 1992.

[15] J. Hyman, A. Lazar, and G. Pacifici, "Real-Time Scheduling with Quality of Service Constraints," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 7, pp 1052-1063, September 1991.

[16] J. Hyman, A. Lazar, and G. Pacifici, "Joint Scheduling and Admission Control for ATS-Based Switching Nodes," *Proceedings ACM SIGCOMM' 92*, Baltimore, MD, pp 223-234, August 1992.

[17] J. LeBoudec, "The Asynchronous Transfer Mode: A Tutorial," *Computer Networks and ISDN Systems 24(1992)*, pp 279-309, 1992.

[18] H. Liu, "Traffic Shaping for Congestion Control in High Speed ATM Networks," Master's Thesis, Department of Computational Science, University of Saskatchewan, August 1992.

[19] B. Maglaris, D. Anastassiou, P. Gunnar Karlsson, and J. Robbins, "Performance Models of Statistical Multiplexing in Packet Video Communications," *IEEE Transaction on Communications*, Vol. 36, No. 7, pp 834-843, July 1988.

[20] S. Minzer, "Broadband ISDN and Asynchronous Transfer Mode (ATM)," *IEEE Communications Magazine*, Vol. 27, No. 9, pp 17-24, September 1989.

[21] T. Murase, H. Suzuki, S. Sato, and T. Takeuchi, "A Call Admission Control Scheme for ATM Networks Using a Simple Quality Estimate," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 9, pp 1461-1470, December 1991.

[22] R. Nagarajan, J. Kurose, and D. Towsley, "Approximation Techniques for Computing Packet Loss in Finite-Buffered Voice Multiplexers," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 3, pp 368-377, April 1991.

[23] C. Rasmussen, J. Serensen, K. Kvols, and S. Jacobsen, "Source-Independent Call Acceptance Procedures in ATM Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 3, pp 351-358, April 1991.

[24] E. Rathgeb, "Modeling and Performance Comparison of Policing Mechanisms for ATM Networks," *IEEE Journal of Selected Areas in Communications*, Vol. 9, No. 3, pp 325-334, April 1991.

[25] K. Reid, "Priority Queueing in Asynchronous Traffic Mode Switches," Master's Thesis, Department of Computational Science, University of Saskatchewan, July 1992.

[26] J. Roberts, "Variable-Bit-Rate Traffic Control in B-ISDN," *IEEE Communications Magazine*, Vol. 29, No. 9, pp 50-56, Sept 1991.

[27] H. Saito and K. Shiomoto, "Dynamic Call Admission Control in ATM Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 7, pp 982-989, September 1991.

[28] H. Saito, M. Kawarasaki, and H. Yamada, "An Analysis of Statistical Multiplexing in an ATM Transport Network," *IEEE Journal of Selected Areas in Communications*, Vol. 9, No. 3, pp 359-367, April 1991.

[29] H. Suzuki, T. Murase, S. Sato, and T. Takeuchi, "A Burst Traffic Control Strategy for ATM Networks," *Proceedings IEEE INFOCOM' 90*, San Diego, CA, pp 505-515, December 1990.

[30] A. Syed, "Quantitative Evaluation of Video Coding Algorithms for ATM-Based Networks," Master's Thesis, Department of Computational Science, University of Saskatchewan, October 1993.

[31] C. Williamson, "Optimizing File Transfer Response Time Using the Loss-Load Curve Congestion Control Mechanism," *Proceedings of ACM SIGCOMM'93*, San Francisco, California, pp 117-126, Sept 1993.

[32] G. Woodruff and R. Kositpaiboon, "Multimedia Traffic Management Principles for Guaranteed ATM Network Performance," *IEEE Journal on Selected Areas in Communications*, Vol. 8, No. 3, pp 437-446, April 1990.

[33] H. Yamada and S. Sumita, "A Traffic Measurement Method and its Application for Cell Loss Probability Estimation in ATM Networks," *IEEE Journal of Selected Areas in Communications*, Vol. 9, No. 3, pp 315-323, April 1991.

Appendix A

Schedulable Regions

Appendix A lists the schedulable regions for the Deadline Scheduling algorithm for the traffic scenarios of video and data, and data and voice; sensitivity of the schedulable region to QOS constraints for the two traffic scenarios of video and data, and data and voice; and the sensitivity of the schedulable region to the traffic characteristics for the traffic scenario of video and data.

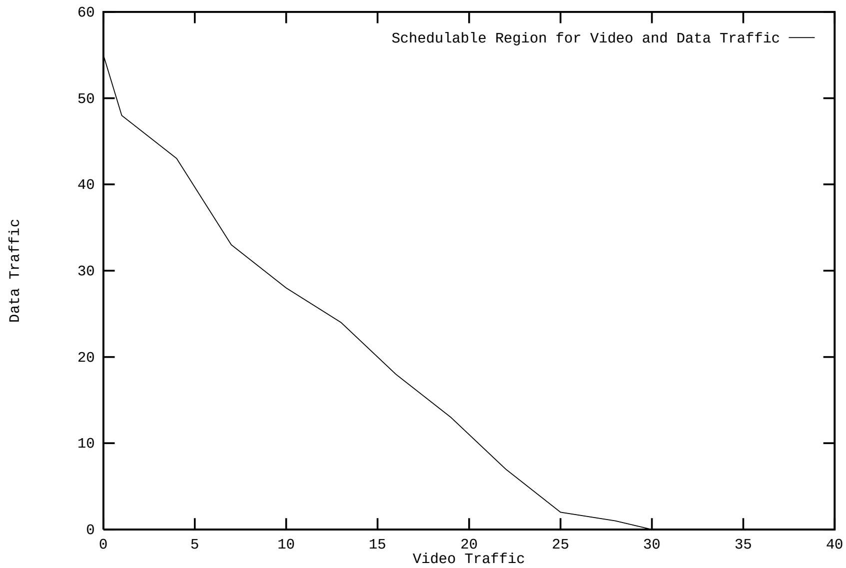


Figure A.1: Schedulable Region for Deadline Scheduling for the Video and Data Traffic Scenario

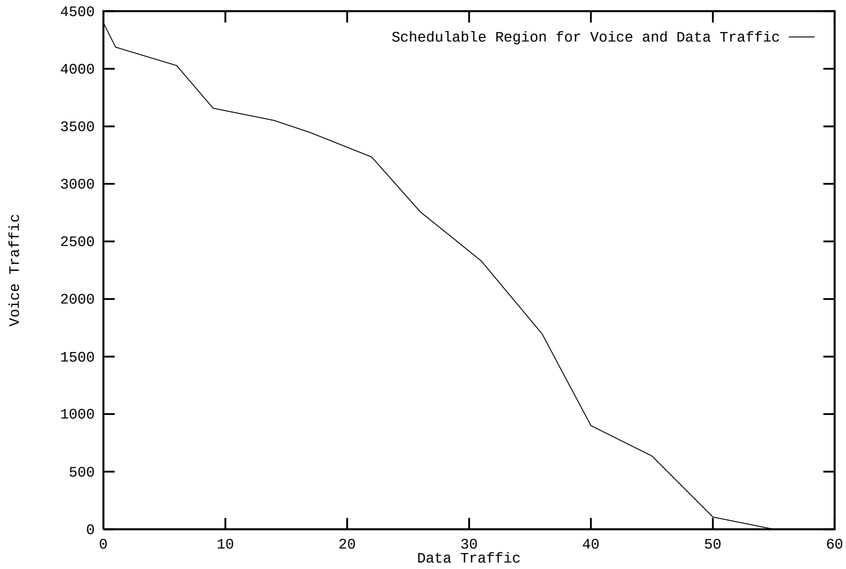


Figure A.2: Schedulable Region for Deadline Scheduling for the Data and Voice Traffic Scenario

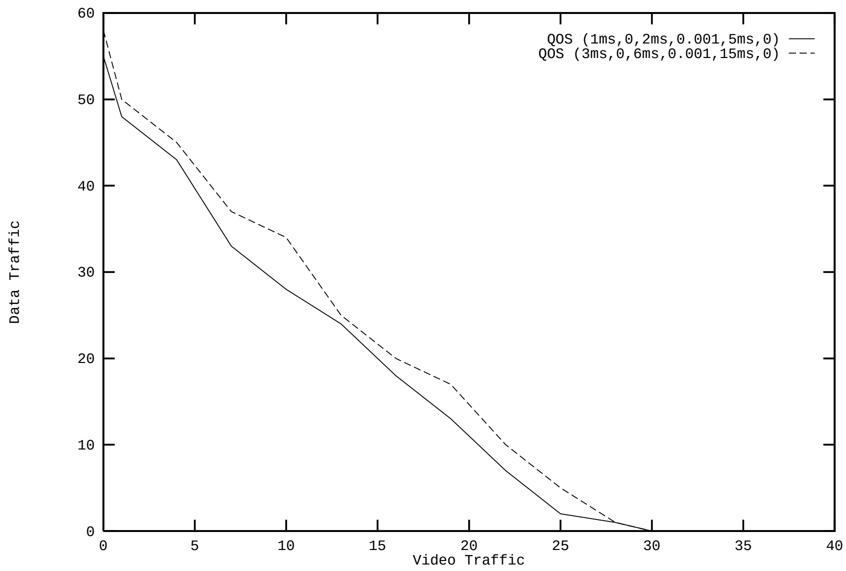


Figure A.3: The Comparison of Schedulable Regions Under $QOS\ SET_1$ and $QOS\ SET_2$ for Deadline Scheduling for the Video and Data Traffic Scenario

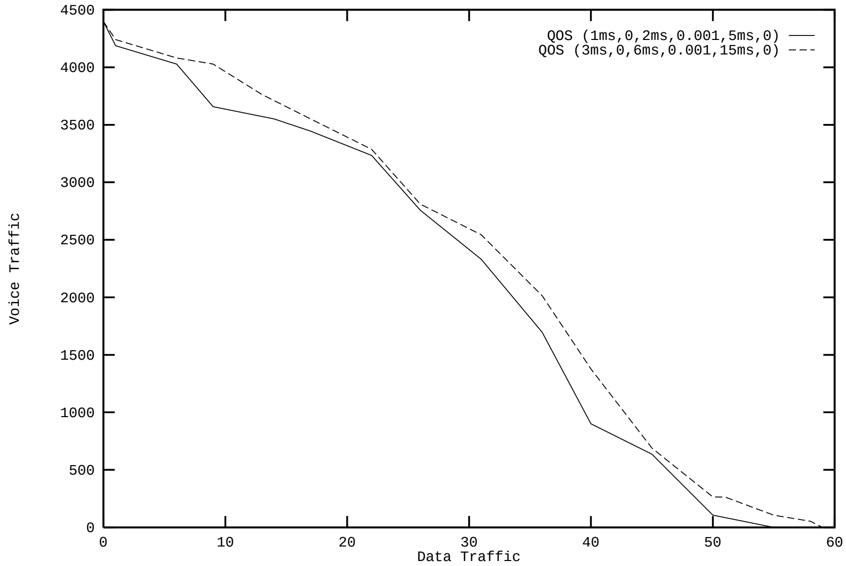


Figure A.4: The Comparison of Schedulable Regions Under $QOS\ SET_1$ and $QOS\ SET_2$ for Deadline Scheduling for the Data and Voice Traffic Scenario

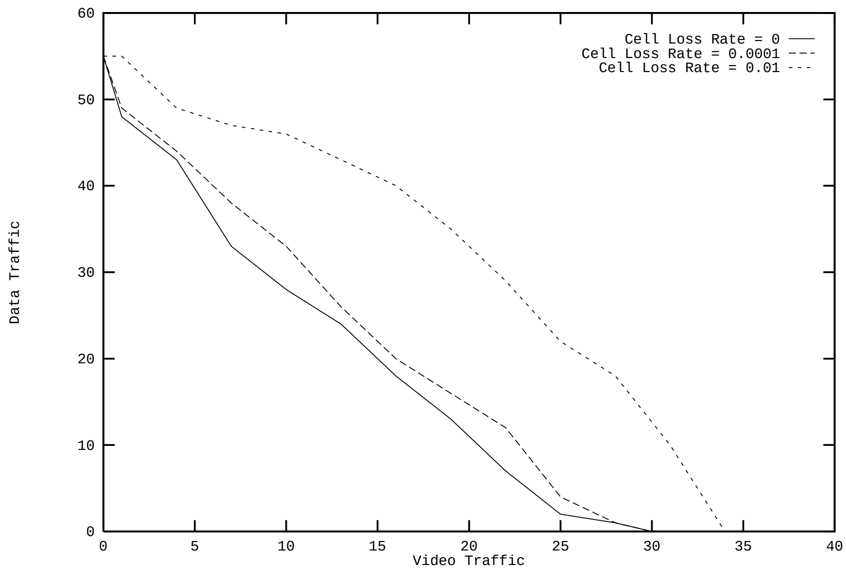


Figure A.5: The Comparison of Schedulable Regions With Relaxed Cell Loss Constraints for Deadline Scheduling for the Video and Data Traffic Scenario

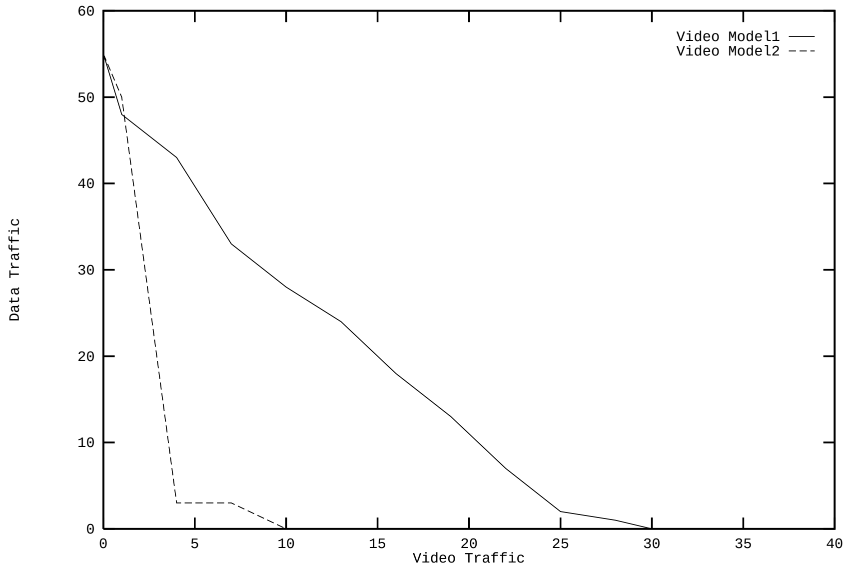


Figure A.6: The Comparison of Schedulable Regions With Different Traffic Characteristics for Deadline Scheduling for the Video and Data Traffic Scenario

Appendix B

Loading Region for the JSAC Policy

Appendix B includes the loading regions for the JSAC Policy for the two traffic scenarios of video and data, and data and voice; sensitivity of the loading region region to QOS constraints for the two traffic scenarios of video and the data, and data and voice; and the sensitivity of the loading region to the traffic characteristics for the video and data traffic scenario.

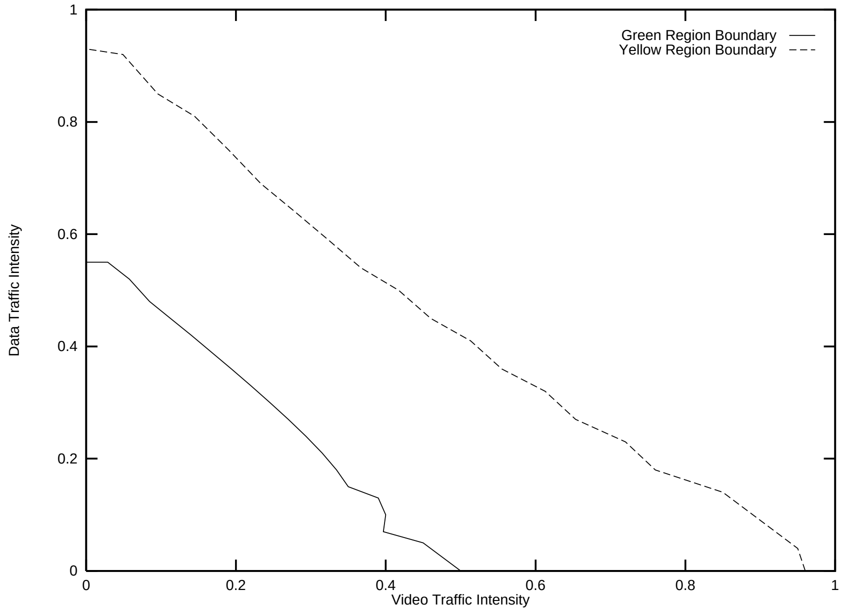


Figure B.1: The Loading Region for the JSAC Policy for the Video and Data Traffic Scenario

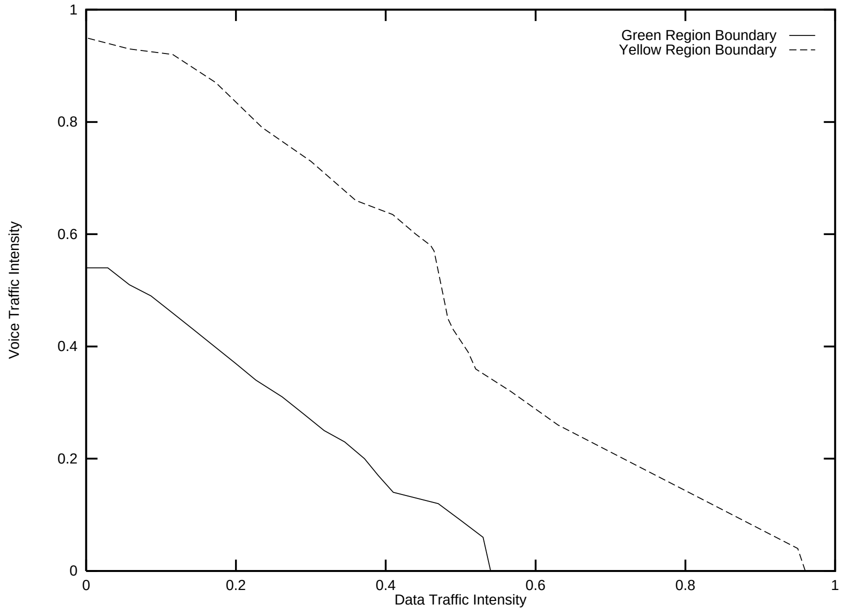


Figure B.2: The Loading Region for the JSAC Policy for the Data and Voice Traffic Scenario

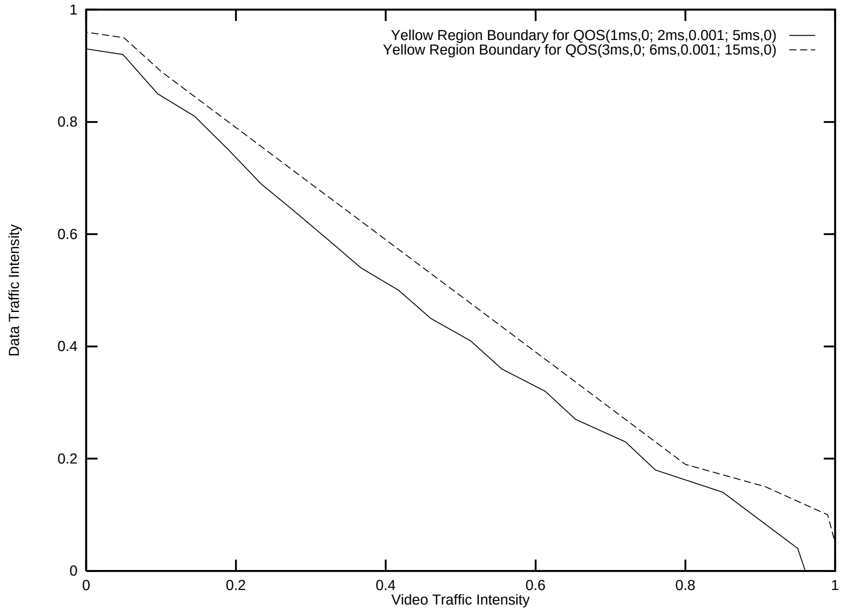


Figure B.3: The Sensitivity of the JSAC Policy to the Cell Delay Constraints for the Video and Data Traffic Scenario

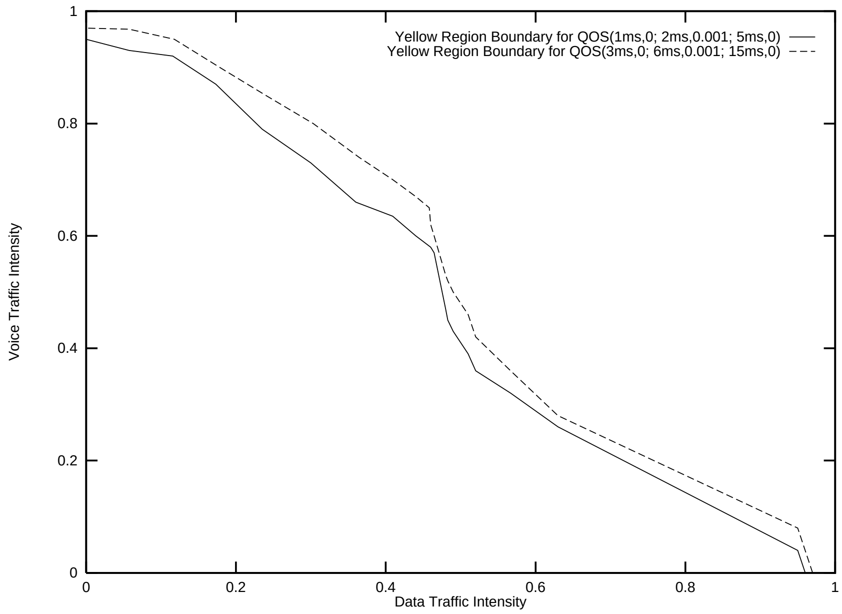


Figure B.4: The Sensitivity of the JSAC Policy to the Cell Delay Constraints for the Data and Voice Traffic Scenario

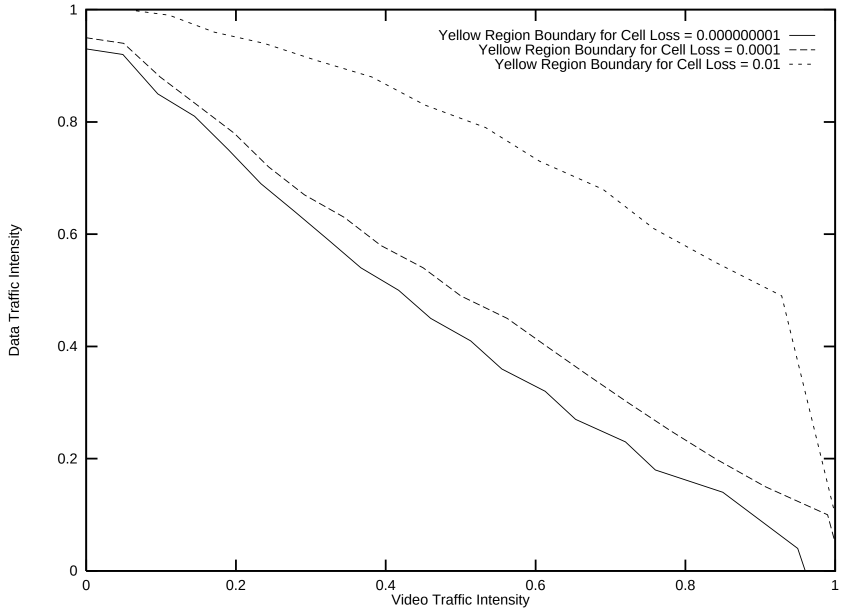


Figure B.5: The Sensitivity of The JSAC Policy to the Cell Loss Constraints for the Video and Data Traffic Scenario

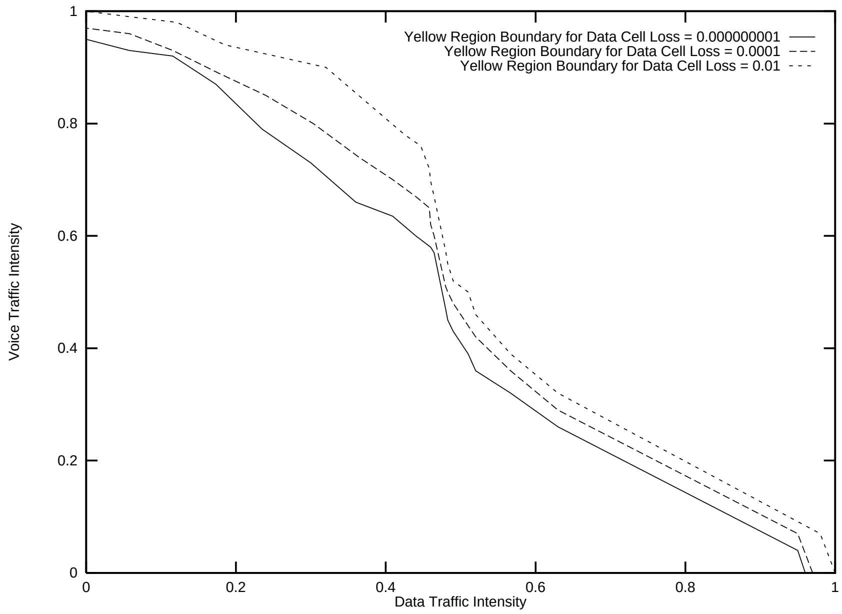


Figure B.6: The Sensitivity of the JSAC Policy to the Cell Loss Constraints for the Data and Voice Traffic Scenario

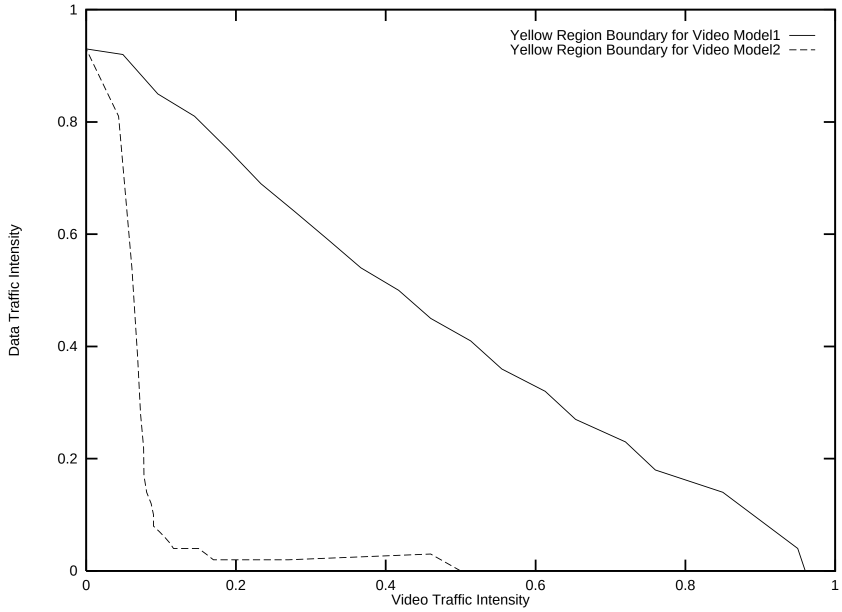


Figure B.7: The Sensitivity of the JSAC Policy to the Traffic Characteristics for the Video and Data Traffic Scenario

Appendix C

Loading Region for the Effective Bandwidth Policy

Appendix C includes the loading regions for the Effective Bandwidth Policy for the two traffic scenarios of video and data, and data and voice; sensitivity of the loading region region to QOS constraints for the two traffic scenarios of video and data, and data and voice; and the sensitivity of the loading region to the traffic characteristics for the traffic scenario of video and data.

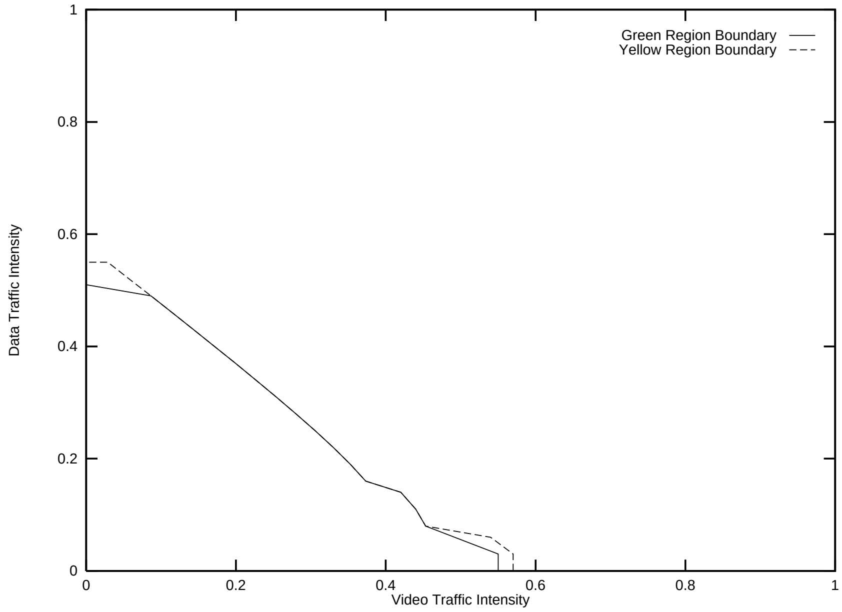


Figure C.1: The Loading Region for the Effective Bandwidth Policy for the Video and Data Traffic Scenario

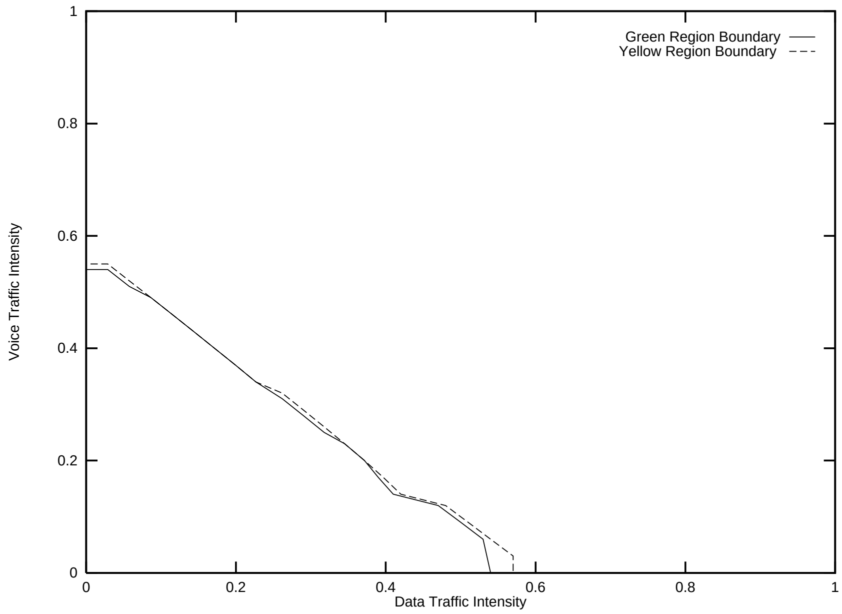


Figure C.2: The Loading Region for the Effective Bandwidth Policy for the Data and Voice Traffic Scenario

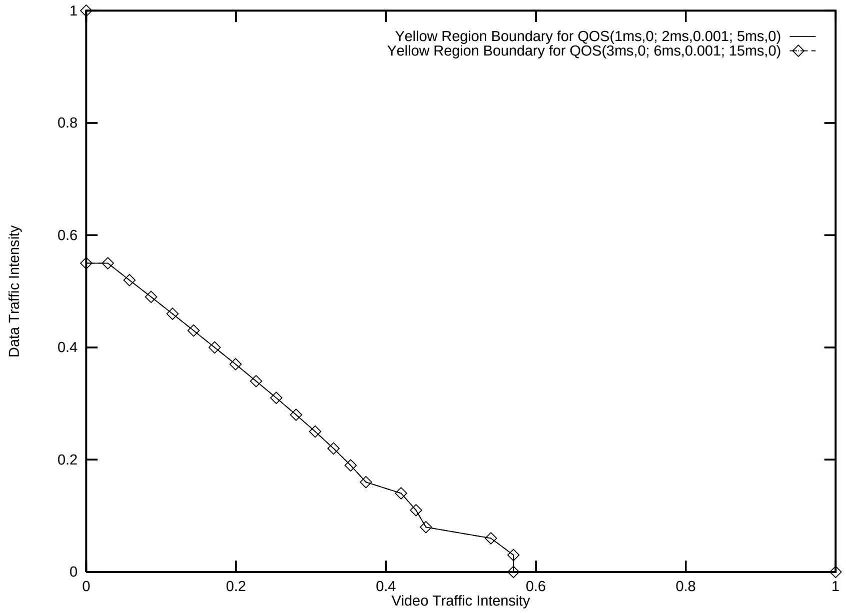


Figure C.3: The Sensitivity of the Effective Bandwidth Policy to the Cell Delay Constraints for the Video and Data Traffic Scenario

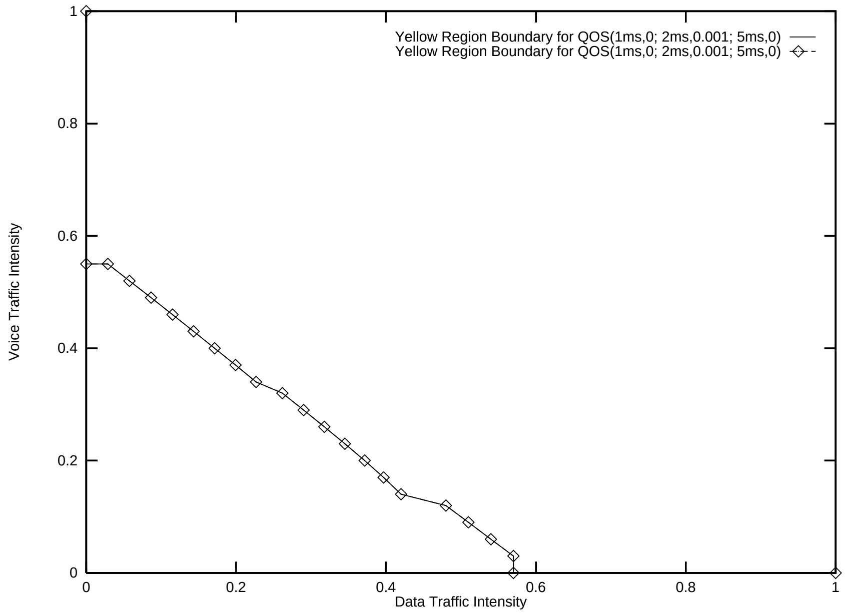


Figure C.4: The Sensitivity of the Effective Bandwidth Policy to the Cell Delay Constraints for the Data and Voice Traffic Scenario

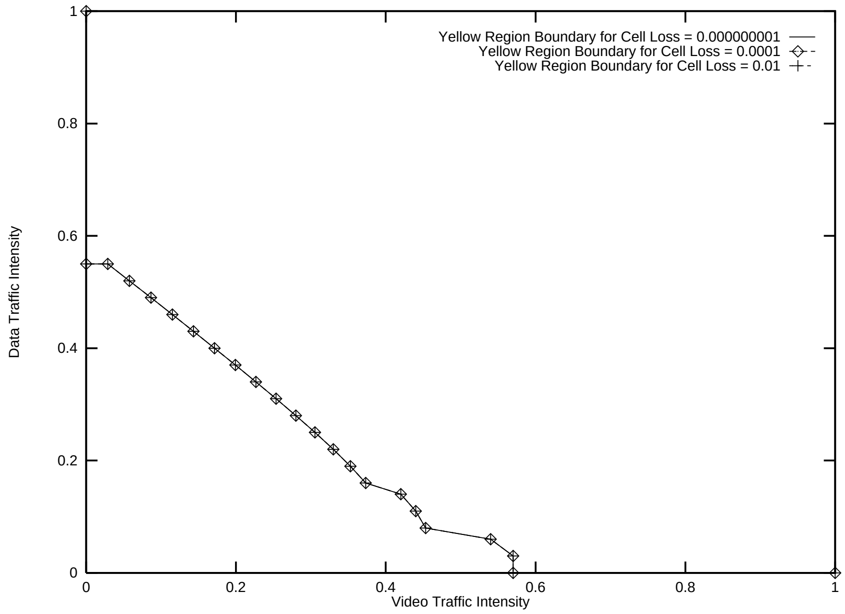


Figure C.5: The Sensitivity of the Effective Bandwidth Policy to the Cell Loss Constraints for the Video and Data Traffic Scenario

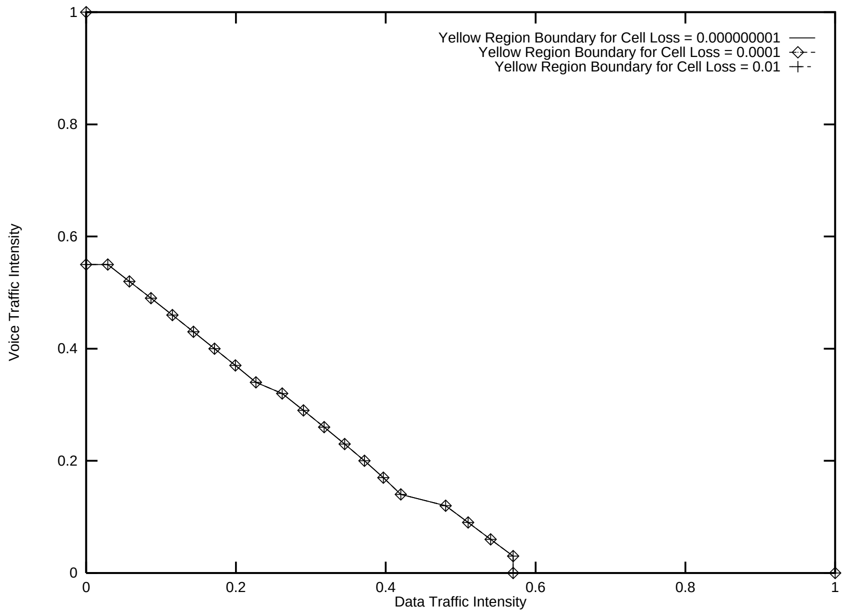


Figure C.6: The Sensitivity of the Effective Bandwidth Policy to the Cell Loss Constraints for the Data and Voice Traffic Scenario

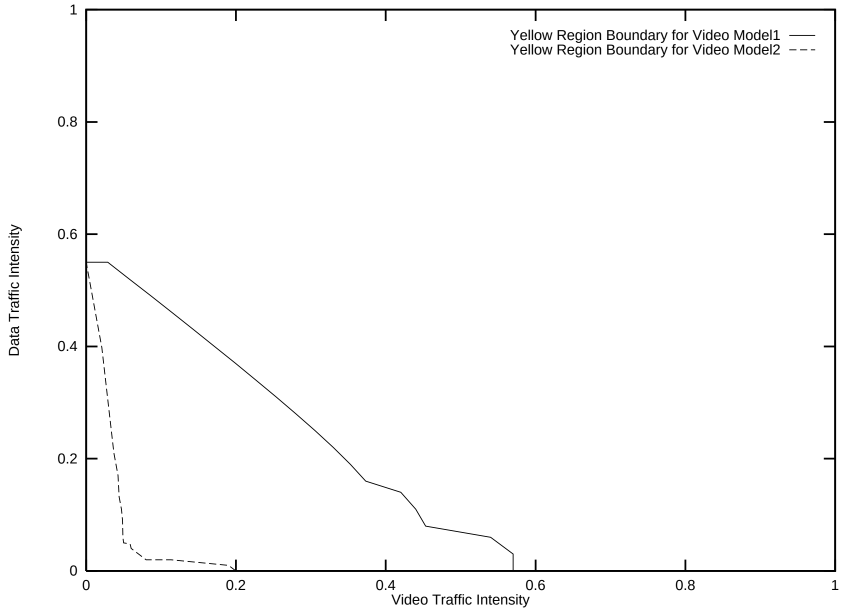


Figure C.7: The Sensitivity of the Effective Bandwidth Policy to the Traffic Characteristics for the Video and Data Traffic Scenario

Appendix D

Loading Region for the Virtual Bandwidth Policy

Appendix D includes the loading regions for the Virtual Bandwidth Policy for the two traffic scenarios of video and data, and data and voice; sensitivity of the loading region region to QOS constraints for the two traffic scenarios of video and data, and data and voice; and the sensitivity of the loading region to the traffic characteristics for the traffic scenario of video and data.

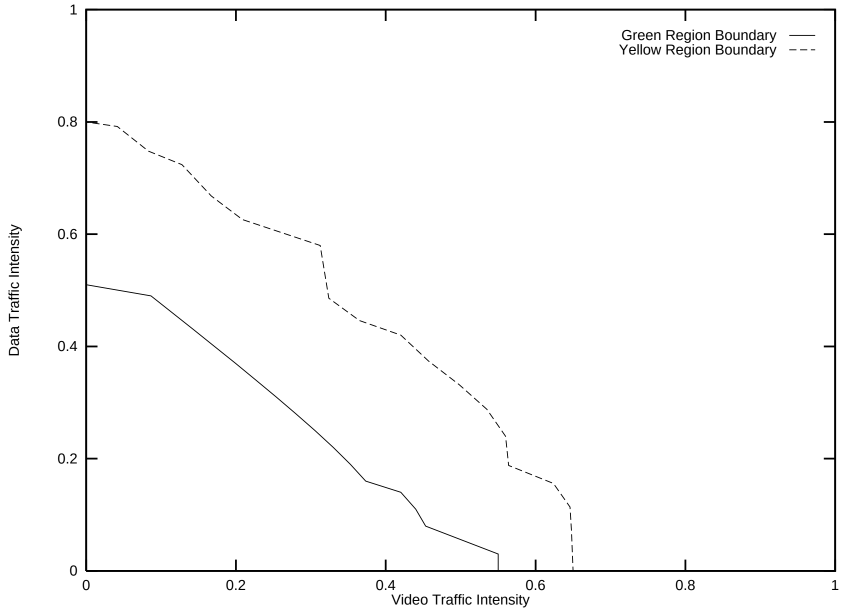


Figure D.1: The Loading Region for the Virtual Bandwidth Policy for the Video and Data Traffic Scenario

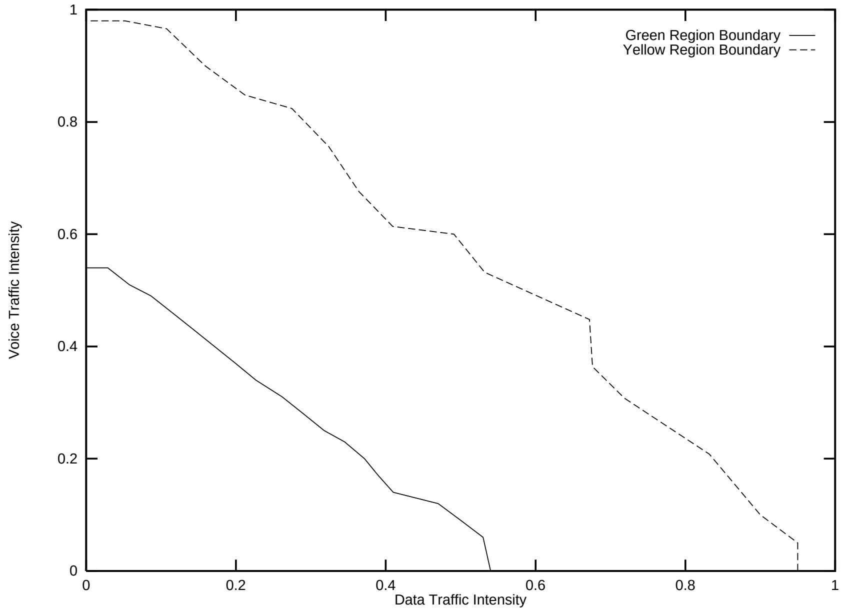


Figure D.2: The Loading Region for the Virtual Bandwidth Policy for the Data and Voice Traffic Scenario

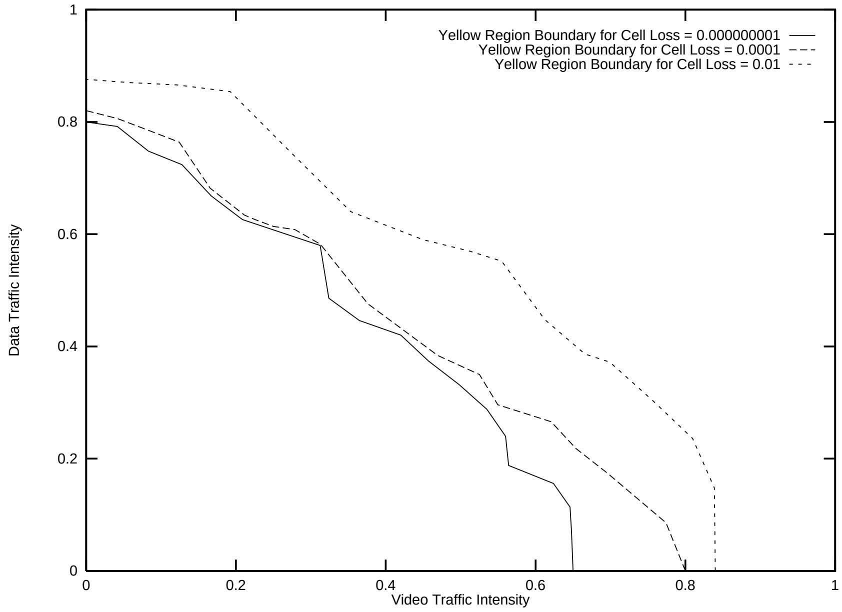


Figure D.3: The Sensitivity of the Virtual Bandwidth Policy to the Cell Loss Constraints for the Video and Data Traffic Scenario

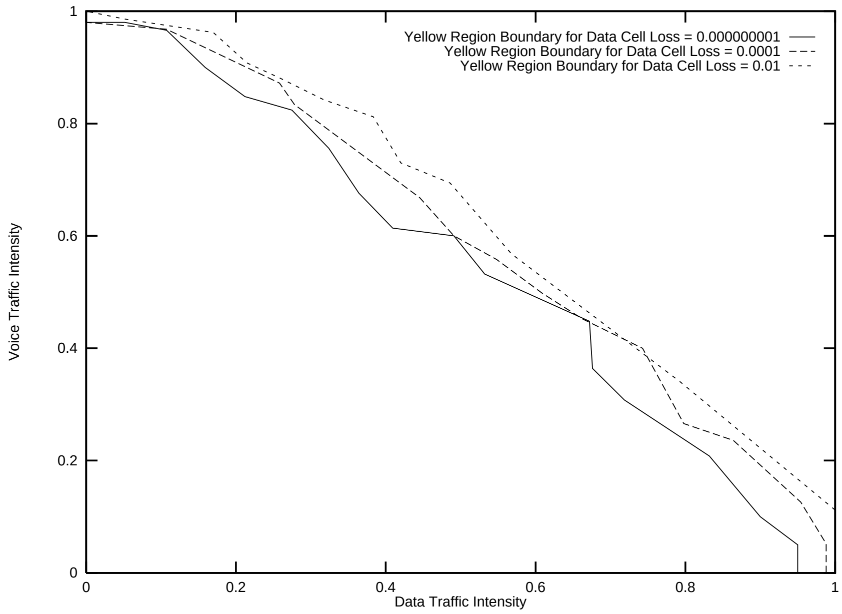


Figure D.4: The Sensitivity of the Virtual Bandwidth Policy to the Cell Loss Constraints for the Data and Voice Traffic Scenario

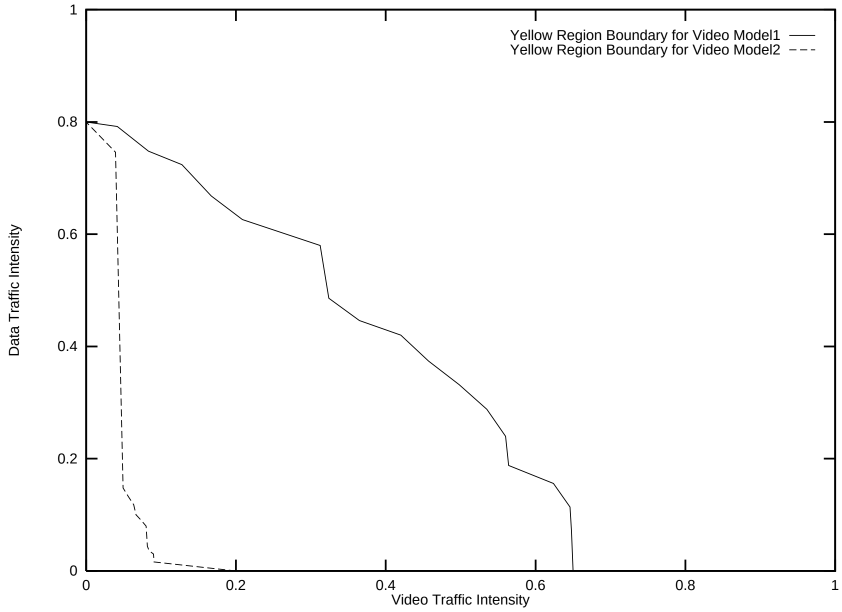


Figure D.5: The Sensitivity of the Virtual Bandwidth Policy to the Traffic Characteristics for the Video and Data Traffic Scenario