

On VBR Video Teleconferencing over ATM Networks

Amy R. Reibman and Arthur W. Berger
AT&T Bell Laboratories
Holmdel, NJ 07733-3030

Abstract

This paper examines the problem of video transport over ATM networks using knowledge of both video system design and broadband networks. We find that the following trends hold for all sequences examined. First, increasing the delay in the video system decreases the necessary peak rate and significantly increases the number of calls that can be carried by the network. Second, as an operational traffic descriptor for video, the leaky-bucket algorithm appears to be superior to the sliding-window algorithm. And finally, with a delay in the video system, the statistical multiplexing gain from VBR over CBR video is upper bounded by roughly a factor of three, and to obtain a gain of 1.5 to 2.0 can require the operational traffic descriptor to have a window or bucket size on the order of a thousand cells. We briefly discuss how increasing the complexity of the video system may enable the size of the bucket or window to be reduced.

1 Introduction

We envision the following scenario for carrying a video call in a Broadband Integrated Services Digital Network (B-ISDN) that is based on Asynchronous Transfer Mode (ATM). When the user decides to initiate a call, the video terminal contacts the network and uses a traffic descriptor to characterize the traffic it intends submit to the network. The network accepts the call if it can provide the desired quality of service to a call conforming to that traffic descriptor. If the network does not have available resources, it may simply reject the call, or the network and terminal may negotiate alternative parameters for the traffic descriptor and/or the quality of service. If the call is established, the network monitors the submitted traffic to ensure that it does comply with the negotiated traffic descriptor.

We examine the impact that the presence of a network traffic monitor (or Usage Parameter Control (UPC)) has on the video system. We argue that this monitor will force the video system to apply rate control to ensure its submitted traffic indeed meets the negotiated traffic descriptors. Thus, we are interested in the traffic descriptor for a video call in a B-ISDN using ATM. We consider *operational* (a.k.a. *algorithmic*) traffic descriptors, which have important advantages over *statistical* ones; see [1] for a detailed discussion.

Currently, CCITT has only specified a traffic descriptor for the peak rate, though additional parameters are expected to be specified in the future [2]. The definition of the peak rate is operational, though the specifics of the definition are not important here.

Herein, we begin the investigation of suitable traffic descriptors (TDs) for teleconference calls by comparing two potential TDs: One consisting of a peak rate and a sliding-window algorithm, and the other consisting of a peak rate and a leaky-bucket algorithm. (These algorithms are reviewed in Section 5.) Note that herein the sliding-window and leaky-bucket algorithms are not attempting to approximate the peak rate but rather are an additional element of the TD. Also, the algorithms do not attempt to approximate the average rate of the call (number of bits transmitted divided by the duration of the call). The algorithms do determine, though, a rate at which the source could continuously submit compliant traffic; we call this rate the “negotiated average rate.”

For sample teleconference sequences, we determine the parameter values of the leaky bucket and sliding window so that the given video sequence is compliant with the traffic descriptor. Our results not only illustrate the magnitude of the parameter values that would be needed, but also provide a comparison of the two algorithms.

Section 2 briefly describes the video used for the experimental results in this paper. It also describes some important characteristics of a generic packet video system. In section 3, we describe the notions of traffic descriptors and Usage Parameter Control used in this paper, and illustrate their importance for video systems. In section 4, we discuss comparisons between CBR and VBR video. Section 5 presents traffic descriptor parameter values for video. Section 6 briefly discusses call admission and possible multiplexing gains given the actual video data. Section 7 concludes the paper with some discussions about how this information can be applied to the design of real systems. Throughout this paper, we describe the problem using the frame period as the discrete time unit; however, smaller sampling intervals are possible.

2 Experimental video sequences

The video used in these examples was recorded at an actual meeting. The video was coded using a one-layer codec that has syntax compatible with H.261 [3]. Three sequences are

used here. Sequence A is 10 minutes long and consists of a person listening and interspersing occasional comments and questions. Sequences B and C are both 5 minutes long, and each contain one active participant. In sequence B, the subject is constantly moving, while in sequence C, the subject moves only occasionally.

2.1 A generic packet video system

The system we consider is shown in Figure 1. A video signal is applied to the video encoder, which produces an encoded video bit-stream. The number of encoded bits produced by frame i is E_i . The encoded bit-stream is stored in the encoder buffer before being transmitted via the channel interface to the network. The rate control device selects R_i , the number of bits transmitted on the channel during frame period i , such that no video buffer constraints will be violated and no traffic that exceeds the negotiated TD parameters is submitted to the network [4]. In addition, the rate control device also selects the quantizer step-size used by the encoder. After being transmitted across the network, the video bit-stream is stored in the decoder buffer. It is then input to the video decoder, which outputs a video signal. For a CBR video system, everything is identical except the traffic descriptor specifies a constant bit rate.

A delay is necessary in the video system to guarantee that the decoder will have access to all the bits corresponding to frame i by the time that frame needs to be displayed. The delay is defined by the interval between the time the decoder receives the first bit at the start of the call until the time the decoder begins to decode. Once decoding commences, a frame must then be displayed every $T = 1/30$ seconds. Thus, the value of the delay, given by LT seconds where L is a non-negative real number, must be known *a priori*, at both the encoder and the decoder. During the course of the call, the delay in each of the encoder buffer, the network, and the decoder buffer are variable. However, the per-frame delay from the encoder buffer input to the decoder buffer output is constant at LT .

3 VBR video and network policing

Variable bit-rate (VBR) video is expected to be advantageous both for the network and for the user. Through statistical multiplexing, the network should be able to carry more VBR video calls than constant bit-rate (CBR) video calls. Alternatively, the user is expected to obtain better quality with VBR video than CBR video, even when both systems have the same average rate.

However, if all streams have completely unconstrained bit-rate, the network either could not ensure a quality of service for established connections, or would be under-utilized due to a very conservative call admission policy. Hence, some form of service contract between the network and the user is necessary, along with traffic and congestion controls by end-system and network (see for example [5]). The user understands two things from the service contract. First, the

network *will* transport (with agreed-to cell-loss rate (CLR)) any traffic the user submits that is compliant with the negotiated traffic descriptor. Second, if excess traffic is submitted, the network has the option of not transporting it. For video, some information is vital to the decoder and should not be dropped; therefore, the video terminal *must not* submit excess traffic as high priority.

The network enforces the negotiated traffic descriptor via the usage parameter control (UPC) function, informally known as the policing function. The UPC serves to protect the network and other users from malicious users or malfunctioning terminals. The UPC includes a monitoring algorithm for the incoming traffic and a control action that is applied to the excessive traffic. The control action could be either to immediately drop excessive traffic, or to mark excessive traffic as low-priority provided excessive traffic is within certain limits. In this paper, the control action to be performed is immaterial, since neither the immediate loss of a high-priority cell nor its possible later loss can be allowed to happen. The video system must therefore ensure that all high-priority traffic submitted to the network is compliant with the negotiated traffic descriptor. Thus, the importance of the UPC for the present work is that its existence causes the video system to control the (high priority) traffic to be compliant with the negotiated traffic descriptor and that this control necessitates buffering in the video system or adjustment of quantizer step-size, or both.

4 VBR video versus CBR video

In this section, we set a level playing field between VBR and CBR video codecs by equating both quality and delay. We assume the network delay is identical for both CBR and VBR video. Therefore, we discuss only the delay induced by the buffers in the video system (see figure 1). Once the quality and the delay are identical, VBR will be advantageous only if the network can carry additional calls. This in turn could allow the network to offer lower prices for VBR service.

The compressed video produced by most codecs is by nature VBR. To transmit across today's circuit switched channels, it must be converted into CBR using an encoder-decoder buffer pair with feedback. The bit-rate is varied by changing the quantizer step-size in response to the fullness of the encoder buffer.

The buffers induce a delay within the video system, and the feedback produces variable quality. In general, using a larger delay produces more constant video quality, since variations in the bit-rate can be smoothed over a larger interval.

On the surface, VBR video should be simple to obtain; simply remove the encoder buffer and disconnect the feedback loop that controls the bit-rate by varying the quantizer step-size. However, as we saw in section 3, the high-priority bit-rate must never exceed the negotiated TD, so the VBR output of a video codec can not be completely

unconstrained. A feedback loop will still be necessary, although it will not be exercised as frequently as for CBR.

A typical method to compare CBR and VBR video is to generate an unconstrained VBR video bit-stream and assume that the CBR bit-stream with comparable quality must have a peak rate (which is the same as its average rate) that is identical to the peak rate of the VBR bit-stream. Then, the CBR bit-stream fills in the valleys between VBR peaks by reducing the quantizer step-size and thus increasing the instantaneous bit-rate.

It is important in this comparison, however, not to ignore the delay in the CBR system. Typically, the VBR output is assumed to be unconstrained and hence to have zero delay, which produces quite a large CBR average rate. However, by assuming the VBR system has a delay identical to the CBR system, the VBR peaks can be smoothed using the buffers. So to compare the CBR and VBR video, we should assign the CBR average rate to be the peak of the VBR output with delay, not without delay. The playing field will then be level, with the CBR and VBR video systems having identical buffer sizes and identical delays.

5 Traffic descriptor parameters

As mentioned in the introduction, we compare two potential traffic descriptors for a video call: One consisting of a peak rate and a sliding-window algorithm, and the other consisting of a peak rate and a leaky-bucket algorithm. We define the peak rate to be $R_{max} = \max_i R_i$.

We assume that within a frame period, cells are evenly spaced. Furthermore, for the simplicity of the mathematical descriptions below, we assume the size units are measured in bits and the time-interval units are measured in frame periods. These can be easily converted into size units of bytes or cells, and time units of seconds.

5.1 Sliding window

A sliding window specifies that no more than a given number of bits (or cells) can be emitted in a time interval of a specified length, where the time interval can begin at any epoch. The sliding window can be described by two parameters, the time duration S_{win} , and the maximum number of bits that can be transmitted in that time window, W_{max} . An alternate description could use W_{max} and the negotiated average bit-rate, $\bar{R} = W_{max}/S_{win}$. Mathematically, the constraint on the channel rate that is imposed by the sliding window is $\sum_{j=k+1}^{k+S_{win}} R_j \leq W_{max}$, for all k .

For a given sequence of R_i , it is a simple matter to determine the size of the sliding window parameters necessary to pass a given bit-stream without violation by computing the maximum number of bits in any window for each window length of interest.

5.2 Leaky bucket

A leaky bucket is a counter that increments by one for each cell emitted, up to a maximum value, and decrements at a given rate to as low as zero – a cell can be emitted if the counter is less than the maximum value minus one.

The leaky bucket can be considered as an imaginary FIFO buffer of size N_{max} bits with constant drain rate \bar{R} bits per frame period. Let N_i be the bucket fullness (in bits) at the end of frame period i . Since R_i bits arrive in frame period i , $N_i = \max\{0, N_{i-1} + R_i - \bar{R}\}$, if we ignore the finite capacity of the bucket. Thus, the traffic would be in compliance if we choose the bucket capacity N_{max} such that $N_i \leq N_{max} \forall i$.

5.3 Examples

We present examples both with and without delay in the video system. Without delay, the number of transmitted bits per frame period is equal to the number of encoded bits in that frame period, $R_i = E_i$. With delay, the R_i are obtained using the smoothing algorithm in [6].

Table 1 shows the peak rate R_{max} and mean rate for 3 teleconferencing sequences. In the last column, C denotes the capacity of the slowest link in the connection, which is 127.155Mb/sec¹. The notation $\lfloor x \rfloor$ in the last column means the largest integer less than or equal to x . Thus, $\lfloor \frac{C}{\bar{R}_{max}} \rfloor$ equals both the number of VBR video connections that can be carried on the link assuming a peak rate allocation, and the number of CBR video connections that can be carried with rate R_{max} .

Table 1 shows how increasing the delay in the video codec decreases the peak rate and increases the number of video connections that can be carried. For example, for sequence A, if the delay is increased from zero to three frames, then the number of CBR connections that can be carried on a link increases three fold from 49 to 150.

Figures 2–4 show the parameter values of the sliding window and the leaky bucket that guarantee the sequences are compliant to the traffic descriptor. The negotiated average rate \bar{R} is plotted as a function of the “size of the TD.” Herein, we use the phrase the “size of the TD” to refer to W_{max} for the sliding window algorithm and N_{max} for the leaky bucket algorithm. In these figures and throughout the paper unless otherwise specified, “with delay” refers to a video delay of $L = 3$ frames. In each figure, the left vertical dotted line indicates the actual average rate. The right dotted line indicates twice the average rate.

Four observations can be made from these figures. First, the general behavior is identical for each sequence, although there is some significant variation between the sequences.

Second, for both traffic descriptors, the size is generally quite large when the necessary negotiated average rate approaches even twice the actual average rate. Table 2 shows

¹A physical layer of 155.52Mb/sec provides 149.760Mb/sec to the ATM layer. If the video ATM Adaptation Layer uses three bytes and the ATM header uses five bytes in a 53 byte cell, the bit rate available for the video information, C , is $149.760 \times 45/53 = 127.155$ Mb/sec.

Seq.	Mean rate (kb/s)	Video delay (frames)	Peak rate R_{max} (kb/s)	Peak to mean ratio	$\lfloor \frac{C}{R_{max}} \rfloor$
A	239.6	0	2562	10.69	49
		3	846	3.53	150
B	849.9	0	3376	3.97	37
		3	1561	1.84	81
C	351.1	0	2340	6.66	54
		3	897	2.55	141

Table 1: Statistics for three sequences

Sequence	Leaky bucket without delay	Sliding Window without delay
A	1,217 (3,381)	9,566 (26,572)
B	59 (164)	1,924 (5,344)
C	751 (2,086)	6,784 (18,844)
	Leaky bucket with delay	Sliding window with delay
A	1,169 (3,247)	9,520 (26,444)
B	- (-)	- (-)
C	693 (1,925)	6,719 (18,664)

Table 2: TD size in kilobits and, in parentheses, in number of ATM cells, when the negotiated average rate, \bar{R} , is twice the actual average rate. (For Sequence B with delay, R_{max} , is less than twice the actual average rate. Thus, the TD size can default to its minimum value.)

the leaky bucket and sliding window sizes when the negotiated average rate, \bar{R} , is twice the actual average rate of the sequence. Note that the TD size can be thousands of cells.

The third observation is that video delay can have a significant impact on the selection of the TD. As shown in Table 1, the presence of video delay in the video system significantly reduces the peak rate. For the three video sequences, the peak rate reduces by one half to one third when the codec delay goes from zero to three frames.

The fourth observation is that for a given negotiated average rate, the leaky-bucket TD requires a bucket size that is significantly less than the window size required by the sliding-window TD. Although the size parameters are not directly comparable, they do strongly influence the duration a source could burst at the peak rate and still be compliant with the TD. For the leaky-bucket descriptor, the duration of such a burst (a.k.a. the ON period) is given by $N_{max}/(R_{max} - \bar{R})$ and for the sliding-window descriptor is given by W_{max}/R_{max} , where we view the source as a continuous flow of bits. For the sample video sequences, we found that the burst duration allowed by the SW is significantly longer (typically 2 to 10 times longer) than that allowed by the LB. For example, for sequences A and C,

the ratio of the ON period for the SW divided by the ON period for the LB is 3.5 and 2.1 respectively, when \bar{R} is twice the actual average rate and the video delay is three frames. (For sequence B with this video delay, the peak rate is less than twice the actual average rate, though for this sequence with zero video delay, the above ratio of ON periods is 17.2) Therefore, although the SW and LB algorithms describe the same video source, a network that admits calls based on the worst case ON-OFF source that is compliant with the TD would be able to admit more calls if the LB algorithm is used as opposed to the SW algorithm.

6 Call admission using TDs

Given that the traffic descriptor for the video call consists of a peak rate and of a leaky-bucket or sliding-window algorithm, we compute a conservative estimate for the number of VBR video calls that can be admitted to a link. We compare this number to the number of CBR video calls that could be admitted.

Given that the network operator is only told the peak rate and the parameters of a leaky bucket or sliding window algorithm, we suppose that for call admission the network operator makes the conservative assumption of periodic ON-OFF sources that stress the limits of the traffic descriptor. We assume each call has the same traffic descriptor and a random onset time which it maintains relative to the other calls. We assume the network buffers are small relative to the burst size. Therefore, if the instantaneous rate is greater than the link capacity, we assume cell loss will occur. We compute the cell loss ratio (CLR) using a binomial approximation, see e.g. [7].

We compute the statistical multiplexing gain (SMG) to be the ratio of the number of VBR calls the link can accept with $CLR \leq 10^{-9}$ to the number of CBR calls the link can accept without loss. It is important to realize that to compute the number of CBR calls the link can accept, we set the CBR rate to the VBR peak rate given a delay of three frames, as described in Section 4 and Table 1.

The statistical multiplexing gain is shown in Figures 5-7 for the three sequences, as a function of the TD size. Our calculation for the number of calls that can be admitted does not directly use the TD size, but only the peak and negotiated average rates. However, to show the impact of the particular algorithms of leaky bucket and sliding window, we pick the independent variable to be the TD size (N_{max} or W_{max} respectively); the associated negotiated average rate that allows the traffic to be compliant for the particular descriptor can be found in Figures 2-4.

The first observation about these figures is that, for the VBR sources without delay, the SMG can be less than 1. In this case, the peak rate of the VBR source is large enough to offset any gains that could be obtained by its smaller mean. (Recall that the peak rate of the VBR without delay is significantly larger than the rate of CBR, which is the peak rate of the VBR *with* delay.) However, the SMG of

the VBR sources with delay is always greater than or equal to 1. Therefore, using delays in the VBR video system is important for obtaining any SMG advantage over CBR.

Second, the statistical multiplexing gain increases as the TD size increases. The maximum statistical gain occurs when the negotiated average rate is equal to the actual average rate of the sequence. Therefore, for the network to carry the maximum number of calls, the TD size must be quite large.

Third, the statistical gains are fairly small. However, the values shown in Figures 5-7 are pessimistic because they use an ON-OFF source (not a video source) and a simple approximation for CLR. We expect the true SMG values to be higher, although they are upper bounded by the peak-to-mean ratio of 3.53, 1.84, and 2.55 respectively for the sequences A, B and C, as shown in Table 1 for $L = 3$. Furthermore, we expect the previous two observations to still be valid. That is, we expect the SMG to improve both as the video codec delay and the TD size increases.

7 Conclusions

We have presented a comparison between VBR and CBR video. It sets a level playing field between the two by equating both the quality and the delay. We presented the parameter values for a sliding-window and a leaky-bucket traffic descriptor that are necessary to ensure sample video teleconferencing sequences are completely compliant, and we examined expected multiplexing gains.

We have four primary conclusions. First, the presence of delay in the video system can reduce the necessary peak rate, and can allow significantly more calls to be carried by the network, whether coded as VBR or CBR.

Second, for the sample teleconference sequences to be compliant with the leaky-bucket or sliding-window traffic descriptor, the size of the bucket or the size of the window may need to be large (on the order of thousands of cells), even when the negotiated average rate of the traffic descriptor is twice the true average rate. (Large bucket or window sizes have the disadvantage that a longer burst of cells at the peak rate could be submitted and still be compliant with the traffic descriptor.)

While we have chosen these parameters to ensure the sequences are compliant, in a real implementation, there is nothing to stop the user/network from agreeing to smaller parameters. However, this will imply that for the video system to obtain a compliant bit-stream, it will have to not only shape its source through the buffering, but also reduce its quality to decrease the overall bit-rate. In this case, the rate-control algorithm in the video system must choose the rate onto the network to conform to leaky bucket (or sliding window) constraints as well as encoder and decoder buffer constraints [4]. Voeten et. al. [8] have also considered a preventive policing mechanism that enables a video terminal to emit only compliant traffic onto the network.

Third, comparing the two operational traffic descriptors,

the leaky bucket is superior to the sliding window. For leaky-bucket and sliding-window parameters with a common negotiated average rate and chosen so that a sample teleconference sequence is compliant, the worst case burst that could be admitted is several times smaller with the leaky bucket than with the sliding window.

Fourth, moderate multiplexing gains are only possible when a delay is present in the video system and when the size of TD is large. With a delay in the video system, the statistical multiplexing gain from VBR over CBR video is upper bounded by roughly a factor of three, and to obtain a gain of 1.5 to 2.0 can require the operational traffic descriptor to have a window or bucket size of thousands of cells, given constant quantizer step-size.

While our results seem to indicate that VBR may not provide large SMG, we expect that potential for multiplexing still exists. However, further research will be necessary to extract the full potential.

Acknowledgements

The authors would like to thank A. Eckberg, B. Haskell and D. Lubinsky for their assistance and advice in the work of this paper.

References

- [1] A. W. Berger and A. E. Eckberg. A B-ISDN/ATM traffic descriptor, and its use in traffic and congestion controls. In *Proc. GLOBECOM'91*, pages 266-270, December 1991.
- [2] Traffic control and congestion control in B-ISDN. CCITT SG.XVIII Draft Recommendation I.371, June 1992.
- [3] Recommendations of the H-series. CCITT SG.XV, August 1990.
- [4] A. R. Reibman and B. G. Haskell. Constraints on variable bit-rate video for ATM networks. *to appear in IEEE Transactions on Circuits and Systems for Video Technology*.
- [5] A. E. Eckberg, B. T. Doshi, and R. Zoccolillo. Controlling congestion in B-ISDN/ATM: issues and strategies. *IEEE Communications Magazine*, 29(9):64-70, September 1991.
- [6] A. R. Reibman and A. W. Berger. Traffic descriptors for VBR video teleconferencing over ATM networks. *submitted for publication*.
- [7] J. W. Roberts, ed. Performance evaluation and design of multiservice networks: Final report of COST 224. Commission of the European Communities, 1992.
- [8] B. Voeten, F. Van der Putten, and M. Lamote. Preventive policing in video codecs for ATM networks. In *Fourth International Workshop on Packet Video*, pages G1.1-G1.6, 1991.

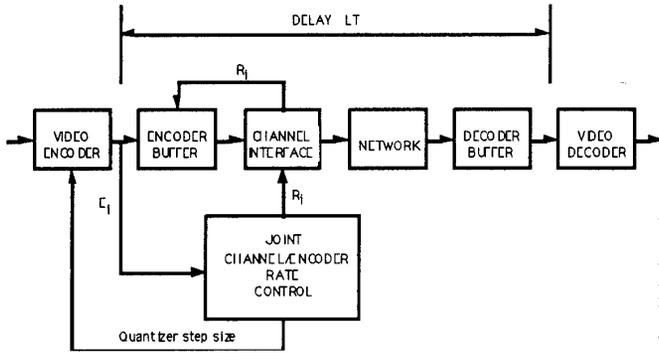


Figure 1: A generic packet video system

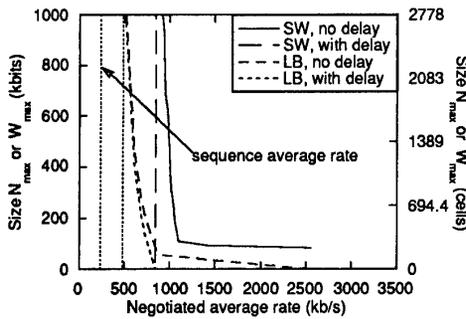


Figure 2: TD parameters, sequence A

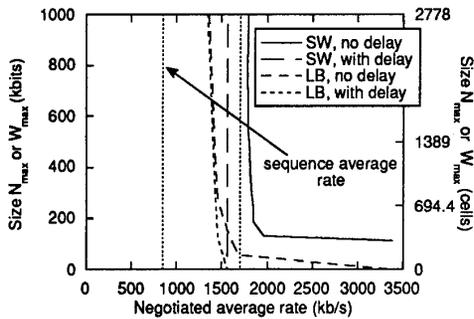


Figure 3: TD parameters, sequence B

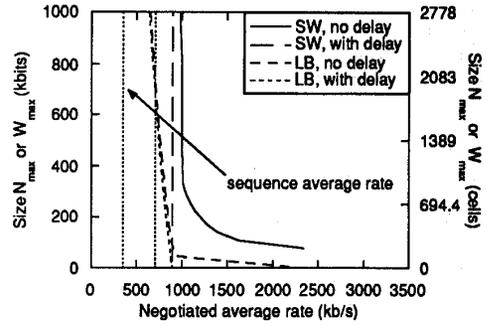


Figure 4: TD parameters, sequence C

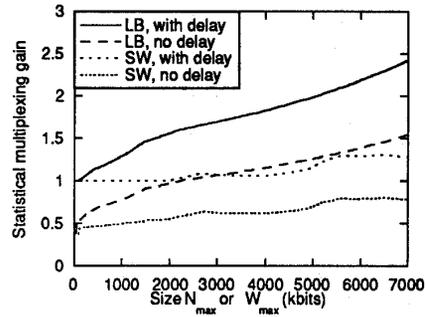


Figure 5: Statistical multiplexing gain, sequence A

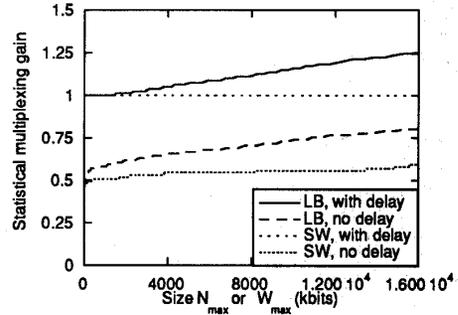


Figure 6: Statistical multiplexing gain, sequence B

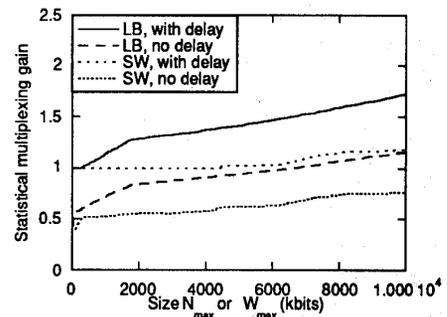


Figure 7: Statistical multiplexing gain, sequence C