Comparison of VBR and CBR service classes for MPEG-2 video multiplexing - Is statistical multiplexing really useful for bursty video transmission? -

Naoki Wakamiya, Masayuki Murata, Hideo Miyahara
Graduate School of Engineering Science, Osaka University 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

SUMMARY

It has been believed that a statistical multiplexing technique can bring an effective bandwidth usage. When it is applied to MPEG-2 video, it has been pointed out that two to several times multiplexing gain can be attained due to highly bursty nature of video traffic. However, most of past researches do not take into account the implementation aspects of this technique. In this paper, we thoroughly investigate the actual multiplexing gain by comparing ATM CBR and VBR service classes (deterministic and statistical service classes), and point out that the statistical multiplexing gain is far from one may expect.

Keywords: MPEG-2, ATM, VBR service, CBR service, statistical multiplexing gain

I. INTRODUCTION

There is an increasing interest in the network support of the effective and high-quality video transfer. Because it is very expensive to transfer un-compressed videos, several coding and compression algorithms are developed and used, such as motion JPEG [1], H.261 [2], MPEG-1 [3] and MPEG-2 [4]. These coding algorithms are different from each other, but generated coded video traffics have the same characteristics called "burstiness". Burstiness means that the traffic rate fluctuates depending on time, and is usually measured by peak to average ratio of the traffic rate. For example, the peak rate becomes two, three, or, sometimes several times higher than the average rate when a video stream is compressed with the MPEG-2 coding algorithm. This highly bursty nature of coded video traffic seems to be well suited to the statistical multiplexing strategy of the high speed networks such as ATM (Asynchronous Transfer Mode) networks. However, multimedia applications employing the video data for an impressive presentation require QoS (Quality of Service) guarantees; continuous and in-time data delivery and a little data losses. There have been a lot of researches on how to guarantee QoS in networks (see, e.g., [5]), but it is still a controversial issue.

As known widely, ATM is designed and developed to be capable of supporting multimedia traffic with various QoS requirements. For this purpose, several service classes are standardized [6]–[9]. Those are UBR (Unspecified Bit Rate), ABR (Available Bit Rate), VBR (Variable Bit Rate) and CBR (Constant Bit Rate) service classes. Each service class differs from another in QoS they can provide. In ITU-T and the ATM Forum, the UBR service class is defined for transferring traditional computer data traffic which requires no QoS guarantee. In the ABR service class, on the other hand, the rate based congestion control is employed to guarantee QoS in terms of CLR (Cell Loss Ratio)[10]. Neither UBR nor ABR service class is therefore appropriate for video data transfer because typical required QoS are CDV (Cell Delay Variation) and maximum cell transfer delay (CTD) as well as CLR.

To employ the CBR service class is the easiest way of transferring video data over ATM networks. Furthermore, it is a realistic solution, which we will show in this paper. The CBR service class is for CBR (Constant Bit Rate) traffic and bandwidth allocation is performed at a peak rate basis [11]. The sender begins connection setup with a description of a peak cell rate (PCR). If enough bandwidth is available in the network, then the network allocates a fixed bandwidth, equal to PCR, to the connection. The sender adjusts the cell emission rate according to PCR, and there are no cell loss and a little cell delay variation as long as the cell emission rate conforms to PCR. These features of the CBR service class are attractive to real-time video transfer applications. However, there are some problems in this strategy. First, as stated above, the sender must estimate an appropriate PCR value of video traffic prior to the connection setup even when no information is available about the characteristics of coded video traffic. Such difficulty arises especially when the video coding is performed in a real-time fashion, but it can be relaxed to some extent by introducing appropriate control mechanisms, such as a play-out buffer and/or error concealing techniques [12], [13]. Play-out buffer can regulate cell emission rate by holding excessive traffic for a while. The error concealing technique, along with intentionally discarding a part of video traffic, can keep cell emission rate under the allocated bandwidth without introducing unpreferable buffering delay. Another problem in video transferring over the CBR service class is that an effective bandwidth usage is never expected when the allocated bandwidth is equal to the peak rate (i.e. maximum frame size) of VBR (bursty) traffic. If enough bandwidth is available and allocated to a connection, one can receive a video stream with little transfer delay and no cell loss. However, in MPEG-2 coded video, almost two thirds of the allocated bandwidth would be unused and wasted as will be shown in Section III..

The VBR service class is then introduced to effectively utilize the network capacity in transferring VBR traffic over ATM networks [8]. For this purpose, the VBR service class employs the statistical multiplexing technique. When a large number of VBR traffic are multiplexed onto a link, total required capacity for multiplexed connections becomes lower than the sum of PCRs. Compared with the peak rate based bandwidth allocation mechanism adopted in the CBR service class, the number of connections simultaneously multiplexed is expected to be increased. However, QoS such as CLR and/or cell delays are guaranteed only statistically, not in a deterministic way.

To perform statistical multiplexing, the sender must first declares an appropriate set of traffic parameters such as SCR (Sustainable Cell Rate) and BT (Burst Tolerance) as well as PCR, and required QoS parameters in terms of CLR, CDV and the maximum cell transfer delay (CTD) (see, e.g., [14]). After CAC (Call Admission Control) procedure, the new connection is accepted only if the required QoS of the new connection is guaranteed and the newly added connection does not affect guaranteed QoS of other existing connections. As long as the emitted cell flow conforms to the negotiated traffic descriptors, the connection is provided with guaranteed QoS. The conformance test is performed through a UPC (Usage Parameter Control) mechanism, such as sliding window, jumping window or leaky bucket [15], [16].

Many researches have been devoted to investigating the effectiveness of VBR video transfer over the VBR service class (see, e.g., [17]–[19]). Those papers mainly focused on the degree of statistical multiplexing gain (SMG) to show the superiority of the VBR service class over the CBR service class. For example, in [17], the authors show that the number of multiplexed connections in the VBR service class is more than three times larger than that of the CBR service class. This result sounds good for network providers and carriers who face the increase of bandwidth demand and the shortage of network capacity. However, those researches miss or ignore several drawbacks of the VBR service class.

First, traffic characteristics of coded video is assumed to be known in advance in [18], [20], [21]. In an actual situation, especially when the video coding is performed in a real-time fashion, it is absolutely hard to predict the required parameters such as traffic parameters for CAC (PCR, SCR and BT) and UPC parameters (the bucket size, drain rate and buffer size, in case the leaky bucket mechanism is employed) at the call setup time [22]. Even when the video data have already been coded and stored, no appropriate method is provided for users to determine the traffic parameters from the actual video traffic characteristics. When these parameters are inappropriately estimated and described, the coded video traffic will not conform to the negotiated parameters. To avoid the violation of a contract, those excess cells are discarded, tagged or buffered at UPC, which results in the unacceptable degradation of perceived video quality at destinations. More recently, several authors have been engaged in mathematical modeling of actual MPEG video traffic [23]–[26] to accurately estimate and determine traffic and UPC parameters. In those papers, authors conclude that the VBR video traffic has a heavy-tailed distribution and a long-range dependence. This conclusion implies that a traditional Markovian source model is not applicable to VBR video traffic any longer. However, the difficulty still remains in determining parameters for video source model.

Second, statistical QoS guarantees in the VBR service class implies that there is unavoidable cell loss at the intermediate nodes even if the appropriate and precise parameter setting is performed at the call setup time [27]. The QoS requirements are satisfied only if they are monitored and averaged over the connection time, while the perceived QoS degradation must be sensitive to shorter term fluctuation, e.g., in the magnitude of seconds. Unpredictable changes in cell delay variation are also caused by statistical multiplexing in the VBR service class [28]. To avoid those QoS degradations, rate control mechanisms must be employed at the sender [29], [30]. On the other hand, the CBR service class provides the deterministic QoS guarantees. No cell loss is observed as long as the enough bandwidth is allocated and the emitted cell rate is kept under the allocated bandwidth. Further, the cell transmission delay can be kept under some deterministic bound.

Last, it is sometimes assumed in comparison that only a peak rate based bandwidth allocation is performed in the CBR service class [17], [18] while the burstiness of VBR traffic is reduced by introducing the buffer at the sender in the VBR service class. For example, in [18], it is shown that an achievable SMG is larger than three. However, to achieve such a highly effective statistical multiplexing, a huge amount of buffer is required at the sender. It is apparently an unfair comparison. The CBR service class can also introduce the buffer to smooth VBR traffic, and the smoothing can reduce the required bandwidth and increase the number of multiplexed connections. That is, if we compare CBR and VBR service classes under the same condition (i.e., the same amount of smoothing buffer), SMG is again decreased.

In summary, there are several remaining obstacles to achieving high SMG in the VBR service class. Those are the complicated CAC mechanism, the difficulty in appropriate parameter setting of traffic characteristics, a large amount of buffer for UPC and switches, and degradation of QoS (cell loss and delay jitter) caused by statistical guarantees. Nevertheless, a lot of researches on the statistical multiplexing of VBR videos are still being pursued and published [31]–[37].

In this paper, we investigate a real effectiveness of the VBR service class by employing actual MPEG-2 video traffic. We try to fairly compare VBR and CBR service classes by considering the problems described above. Through experiments, we will show that the achievable SMG by statistical multiplexing of the VBR service class is unexpectedly low; at most two and sometimes less than one. In obtaining these results, we will assume that traffic parameters are known a priori to employ the CAC algorithm which is described in the published literature. It is also assumed that UPC parameters can be determined appropriately based on the traffic characteristics. Our intention is that even with such favorable assumptions to the VBR service class, SMG is lower than that one may expect.

Instead of pursuing the VBR service class, we consider the encoding algorithm suitable to the CBR service class. In the CBR service class, a fixed bandwidth is allocated to each connection according to the negotiated PCR. However, negotiated PCR may be smaller than the actual peak cell rate of coded video because of the difficulty in PCR determination. To overcome such a problem, the sender can employ the ap-

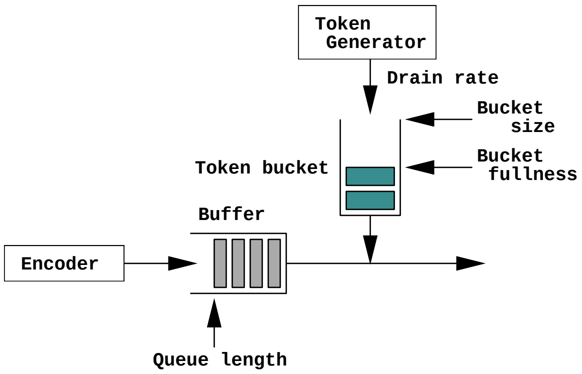

Fig. 1. Leaky bucket model

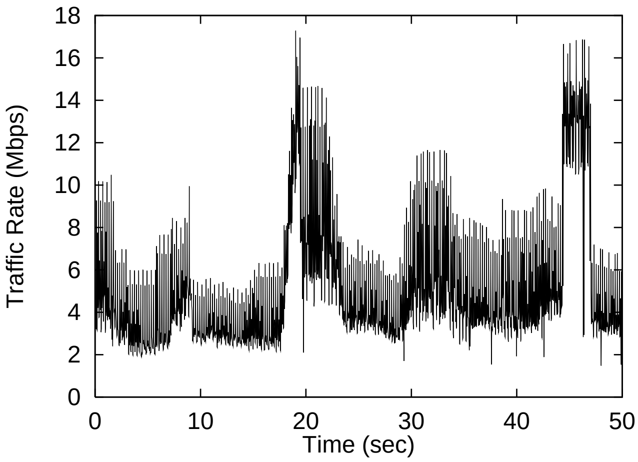

Fig. 2. MPEG-2 video traffic (*Scenery*)

propriate mechanisms to control the coding rate [12], [38], which is called MPEG-2 Test Model 5 [38]. MPEG-2 Test Model 5 ("TM5" in short) regulates the degree of quantization to keep the averaged traffic rate around the allocated bandwidth. In this paper, we will show the effectiveness and applicability of MPEG-2 Test Model 5 to the CBR service class.

Last, we note here that we focus on MPEG video and the CBR and VBR service classes of ATM networks in the current paper. However, we believe that results obtained from comparisons can also be applied to other packet networks where a bandwidth reservation mechanism is implemented [39]–[41].

This paper is organized as follows. In Section II., we first introduce test scenarios for comparing CBR and VBR service classes fairly. Then, in Section III., we compare the effectiveness of multiplexing mechanisms of VBR and CBR service classes for MPEG-2 video transfer. The applicability of TM5 is also discussed briefly. Finally, we conclude our study in Section IV..

II. System models

In this section, we introduce our test scenarios to fairly compare the VBR and CBR service classes by means of the statistical multiplexing gain. We first describe assumptions and model of the CBR and VBR service classes and define SMG.

In the CBR service class, a sender first specifies a PCR (Peak Cell Rate) value at the connection setup to reserve enough bandwidth [11]. The network accepts the request if remain-

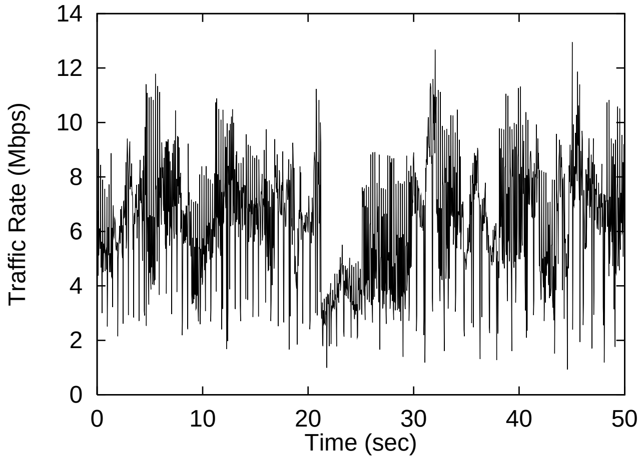

Fig. 3. MPEG-2 video traffic (*Starwars*)

ing capacity is larger than PCR on all links which the connection traverses. In our setting, all connections are assumed to transmit the same video data. Therefore, the maximum number of connections simultaneously acceptable on the link becomes C/R_{max} where C [bps] denotes the link capacity and R_{max} [bps] is the PCR value.

Once the connection is accepted and established in the CBR service class, the user should conform its cell flow to negotiated PCR. The mechanism to perform the conformance enforcement is called UPC (Usage Parameter Control), which may be implemented by means of the leaky bucket mechanism [42]; that is a kind of a credit-based flow control mechanism. The leaky bucket mechanism works as follows. As shown in Fig.1, a token is generated from the token generator at a fixed rate called "drain rate". Then, generated tokens are pooled in the token bucket of finite capacity. Two parameters controls the behavior of the leaky bucket. They are the drain rate of tokens and the size of the token bucket, denoted as (ρ [bits], σ [bps]), respectively [43].

In our model, video data stream produced by an MPEG-2 encoder is segmented into cells and stored in a buffer. It means that cells which do not conform to the PCR value are not discarded nor tagged, but buffered at UPC. Stored cells in the buffer are injected into the network as long as enough number of tokens is pooled in the token bucket. The token in the token bucket is then removed when the cell is injected into the network. This mechanism is called "traffic shaping" and the leaky bucket works as a "traffic shaper". In the case of the CBR service class, a drain rate ρ is usually set at the negotiated PCR and the token bucket size σ is zero. That is, the leaky bucket mechanism just behaves as a simple "spacer" in the CBR service class. The token bucket is used in the VBR service class to allow bursty cell emission as described later. The interval of cells emitted into the network is kept no less than 1/PCR. The buffer is used when the negotiated PCR is less than the actual peak rate of the original video. The buffer size should be determined by an allowable maximum buffering delay, which should be dependent on the application.

In the case of the VBR service class, the conformance test and traffic shaping for SCR (Sustainable Cell Rate) are also required. They can also be implemented by means of the leaky bucket mechanism. The drain rate of the leaky bucket (ρ) is set for the negotiated SCR, and the token bucket is of finite size σ which is determined from the relation $R_{peak} = \frac{\sigma}{\text{time unit}} + \rho$ [19]. R_{peak} [bits per frame time (1/30 msec)] is the maximum rate of coded video traffic. The purpose of the token bucket is to allow the instantaneous bursty cell emission even if its rate is larger than SCR. When there are enough tokens in the bucket, the cells are immediately injected into the network. Otherwise, they are just stored and delayed in the buffer, not discarded nor tagged in our model. With this traffic shaping mechanism, the emitted cell rate averaged over long duration can be kept under or equal to the negotiated SCR. CAC (Call Admission Control) becomes complicated in the VBR service class. According to the standard [7], the sender must declare the SCR (Sustainable Cell Rate) and BT (Burst Tolerance) in addition to PCR. Further, we need take into account UPC to decide the call admission, since the cell flow to the network is affected by the UPC mechanism. In our case, we assume the CAC procedure works as follows. For given parameters such as the number of connections N, a maximum cell rate from leaky bucket R_{max} [bits per frame time], link capacity C [bits per frame time], and buffer size at the intermediate node B [bits], the cell loss ratio of a connection is determined as [18]

$$\begin{aligned} CLR &= \frac{\text{Expected number of lost bits}}{\text{Expected total number of bits}} \\ &= \frac{\left[\sum_{m=n_0+1}^{N} [mR_{max} - C - B]\binom{N}{m} p^m (1-p)^{N-m} \right]}{pNR_{max}} \end{aligned} \quad (1)$$

where $n_0 = \lfloor C/R_{peak} \rfloor$ is the maximum number of connections when the peak rate bandwidth allocation is performed. As in [18], we consider CLR is equal to the bit loss ratio. We should note that an intermediate node is modeled as a simple FIFO queue and cells are stored in the buffer only when the sum of incoming traffic rate exceeds the link capacity ($mR_{max} - C$). In Eq.(1), a cell flow departing from the leaky bucket is modeled as the conservative periodic on-off source. The peak rate R_{max} during the on period corresponds to a sum of the drain rate ρ and the token bucket size σ, and is PCR in the VBR service class. Conformance test on PCR is performed at the second leaky bucket in the case of a dual leaky bucket UPC mechanism (not shown in Fig. 1). The probability p that the video source is on is given as $p = \rho/R_{max}$. Then, from Eq.(1), we can easily derive the maximum number of connections N_{VBR} that the network can accept while satisfying the specified CLR. We note here that in an actual situation, detailed characteristics of coded video traffic are not known in advance while it is not an easy task to estimate the actual peak cell rate, the sustainable cell rate and the burst tolerance. However, a CAC procedure utilizing Eq.(1) (and any other estimation method) cannot be applied without such knowledge, which is one of main problems of the VBR service class as we have mentioned in Section I.. We further notice that we determine σ and ρ from R_{max} (an actual peak rate) of test sequences. Those are often inappropriate since those values are too large, because R_{max} is determined from an instantaneous peak rate. When the application

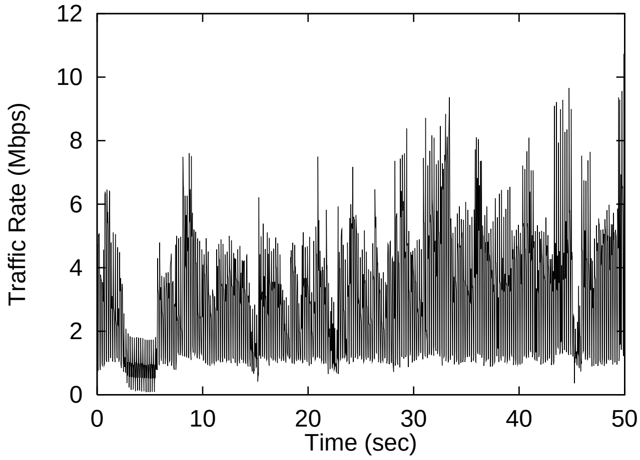

Fig. 4. MPEG-2 video traffic (*Live*)

TABLE I
Traffic characteristics of test videos

Name		Mean	Maximum	Burstiness
Scenery	Size	170 Kbits	576 Kbits	3.40
	Rate	5.10 Mbps	17.28 Mbps	
Starwars	Size	205 Kbits	432 Kbits	2.11
	Rate	6.15 Mbps	12.96 Mbps	
Live	Size	99 Kbits	358 Kbits	3.62
	Rate	2.97 Mbps	10.74 Mbps	

can tolerate the extra delay, we can choose the smaller R_{max} and the degree of statistical multiplexing can be increased as in the case of the CBR service class. In Subsection B., we will provide further discussions for such a case.

Finally, the statistical multiplexing gain (SMG) is defined as the ratio of the number of multiplexed connections in the VBR service class to that of the CBR service class,

$$SMG = N_{VBR}/N_{CBR} \quad (2)$$

where N_{VBR} and N_{CBR} are the maximum allowable numbers of connections in VBR and CBR service classes, respectively. In Section C., we will use this measure to investigate the statistical multiplexing effect of the VBR service class by comparing with the CBR service class. To perform fair comparison, we derive N_{VBR} and N_{CBR} under the same delay bound in such a way that the maximum buffering delays are identical in both VBR and CBR service class.

Throughout this paper, we employ three different traces of video traffic, "*Scenery*", "*Starwars*" and "*Live*". They are coded by an MPEG-2 coding algorithm. These videos are captured from laser disk softwares and are of 1500 frames long (50 seconds). Their traffic variations are shown in Figs.2 through 4 where the horizontal axis corresponds to time and the vertical axis shows the coded video's traffic rate per frame. Each video is coded with a fixed degree of quantization, and their traffic characteristics are summarized in Table I. For each video, the mean and maximum frame size and traffic rate are shown. We also show burstiness (defined as peak to average ratio) of video sequences. From the table, it is obvious that the bandwidth allocation to video traffic on peak rate basis results in waste of network resources. For example, when the peak rate bandwidth allocation is performed against "*Scenery*" video traffic, over two thirds of the allocated bandwidth is unused and wasted.

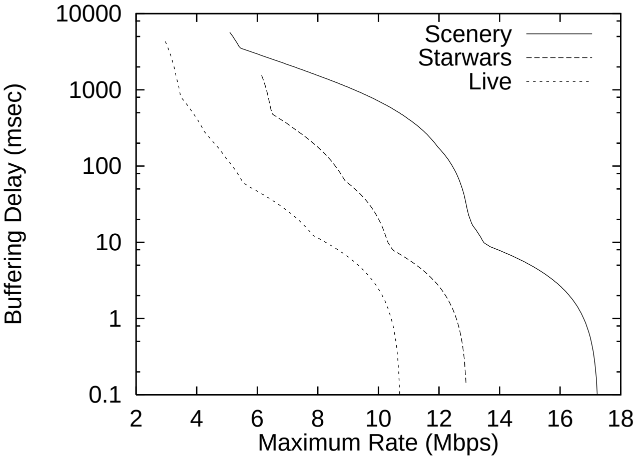

Fig. 5. Effect of buffering on maximum rate in CBR service class

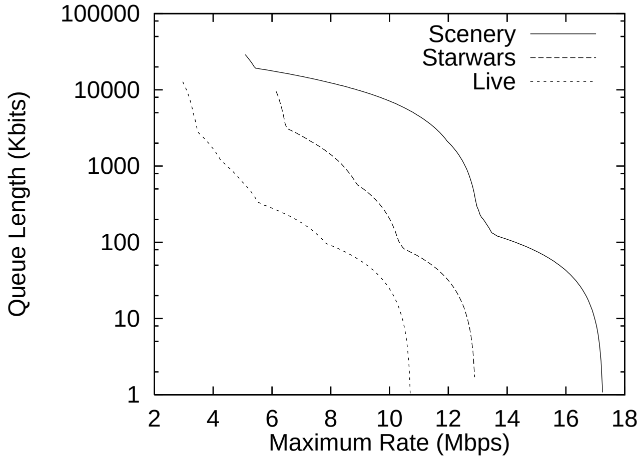

Fig. 6. Effect of buffering on maximum queue length in CBR service class

III. Fair comparisons of VBR and CBR service classes

In this section, we first examine the buffering effects in CBR and VBR service classes in Subsections A. and B., respectively. The buffer at the sender can decrease the burstiness of video traffic while introducing the buffering delay. We then investigate the statistical multiplexing gain of the VBR service class by comparing with the CBR service class in Subsection C.. To perform a fair comparison, buffering effects are taken into account in both service classes.

We assume that each connection behaves identically based on the same traced data. It is also assumed that all traffic parameters are known a priori. Those are unrealistic, but necessary assumptions to apply CAC algorithm described in the previous section when we consider the VBR service class. This is another problem of the VBR service class, and our intention is that even if we introduce such favorable assumptions to the VBR service class, the statistical multiplexing gain is not as high as expected, which will be demonstrated below. All results in this section are obtained arithmetically.

A. Effect of buffering on multiplexing gain in the CBR service class

We first investigate how effective multiplexing can be achieved with the buffer at the sender (Fig. 1) in the CBR service class. Without any buffer, the maximum number of connections which can be simultaneously multiplexed on the link of 150 Mbps capacity are 7, 10 and 12 for each videos (Table I). This comes from the fact that the maximum rate of each video traffic is very high. By introducing the buffer at the sender, their maximum rate can be decreased at the cost of extra buffering delay. This is shown in Fig. 5 where the relationship between the maximum buffering delay in the smoothing buffer and the allocated bandwidth (PCR) is depicted. In this case, video traffic is stored and smoothed by the buffer and the cell emission rate is regulated in order to fit the allocated bandwidth. In obtaining the figure, we assumed that the buffer has an infinite capacity. Maximum buffering delay in this case should be regarded as the maximum allowable delay for the application. For example, when the application can tolerate one second buffering delay additional to the network delay, the required amount of bandwidth for "*Scenery*" can be decreased to 9.16 Mbps from 17.28 Mbps. It means that two times *deterministic* multiplexing gain can be obtained if we allow one second delay at the sender. However, the required buffer size increases as shown in Fig. 6 where the maximum queue length is displayed on the vertical axis. For example, one second delay in "*Scenery*" corresponds to 9.31 Mbits buffer.

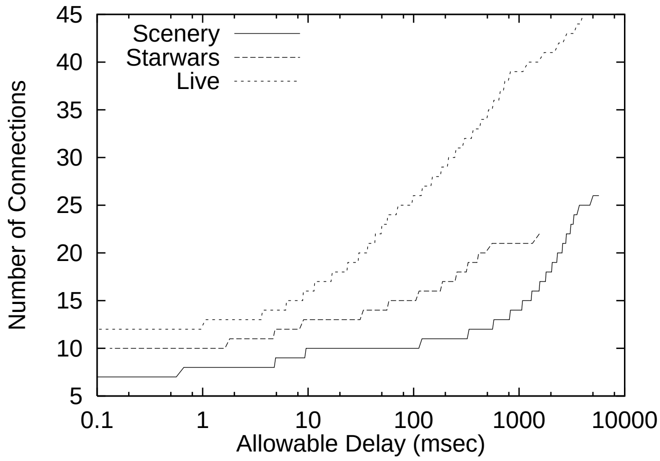

Fig. 7. Effect of allowable delay on number of connections in CBR service class

The gain of deterministic multiplexing is illustrated in Fig. 7 where the number of multiplexed connections on 150 Mbps link is plotted against the allowable buffering delay. As shown in the figure, the multiplexing gain can be obtained even when the maximum allowable delay is only ten or hundred milliseconds in the case of "*Live*", and real-time applications can tolerate such a small delay. On the other hand, the gain is not very high in the case of "*Scenery*". One second delay and 9.31 Mbits buffer are required to obtain the gain of two and it is applicable to only non real-time applications such as a one-way video distribution. We should note here that we do not take into account video coding delay and the propagation delay between sender and receiver. However, we do not need to consider the cell queueing delay at switches since the CBR service class is considered here.

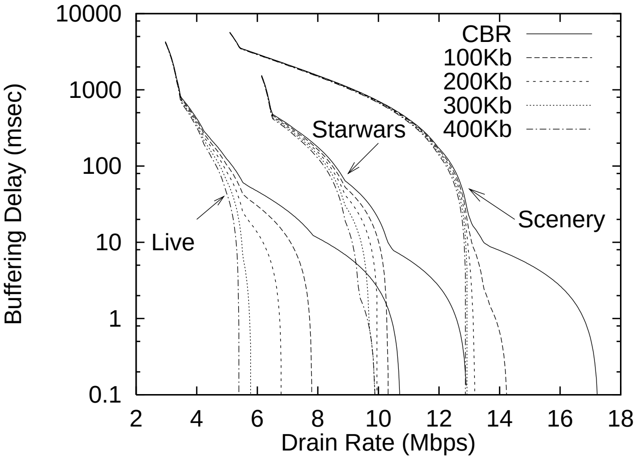

Fig. 8. Relationship among drain rate and buffering delay

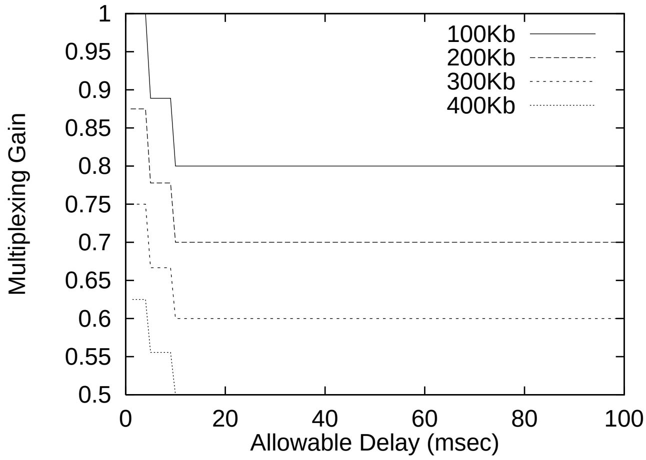

Fig. 9. Statistical multiplexing gain (*Scenery*, LinkCapacity=150 Mbps, CLR=10^{-6}, no-buffer)

We may conclude that even in the CBR service class, the multiplexing gain can be increased if we introduce the buffer at the sender to decrease the maximum cell emission rate. However, because the degree of multiplexing gain depends on maximum allowable delays and video contents, it is difficult to apply the buffer smoothing to real-time video communications. One possibility to overcome these difficulties is to introduce the rate control mechanism for video traffic, which will be described in Subsection D..

B. Buffering effect in the VBR service class

In this subsection, we investigate the relationship among drain rate ρ, token bucket size σ and cell buffering delay in the VBR service class. We first evaluate the relationship between the buffering delay and the drain rate ρ. The drain rate of the leaky bucket mechanism is the negotiated SCR. In Fig. 8, the relationship between token drain rate and buffering delay is shown for "*Scenery*", "*Starwars*" and "*Live*". In the case of the VBR service class, we have another parameter σ, the size of the leaky bucket, which determines the burstiness of cell flow injected into the network from the leaky bucket. In our results, the bucket size is varied from 100 Kbits to 400 Kbits to see the effect of burstiness on buffering delay. The results of the CBR service class are also shown in the same figure for comparison purpose. In the CBR service class, the drain rate corresponds to the peak cell emission rate (allocated bandwidth).

In the figure, we can observe that the drain rate required to satisfy a specific delay bound in the VBR service class can be smaller than that of the CBR service class. It is because the token bucket of finite size leads to the bursty cell emission, and its rate can easily go beyond SCR in the VBR service class. For example, when an application can tolerate 10 msec buffering delay, the required bandwidth in the CBR service class is 13.48 Mbps. On the contrary, the drain rate in the VBR service class with 100 Kbits buffer is 13.13 Mbps. However, the difference between CBR and VBR service classes in the figure does not directly correspond to the statistical multiplexing gain of the VBR service class. The number of multiplexed connections in the VBR service class is more than that of the CBR service class, because the drain rate to satisfy the buffering delay bound is smaller. However, the bursty transmission allowed in the VBR service class decreases SMG, which will be investigated in the next subsection.

The appropriate parameter setting of drain rate ρ and token bucket size σ to satisfy some delay bound in the VBR service class heavily depends on the characteristics of the video sequence. It implies that control parameters for the leaky bucket, i.e., the drain rate and the token bucket size, should be determined with careful consideration on video traffic characteristics. However, the characteristics of video traffic generated by a multimedia application cannot be accurately estimated in advance. This is another reason why the VBR service class is difficult to implement.

C. Comparisons of deterministic and statistical Multiplexing gains

In this subsection, we fairly compare the effectiveness of the CBR and VBR service classes in terms of the statistical multiplexing gain (SMG). The SMG is defined as the ratio of the maximum number of connections multiplexed in the VBR service class to that of the CBR service class where the allowable delay bounds are the same for the two service classes. We note that the adequate buffer sizes to satisfy a delay bound differ between two service classes, and the VBR service class required a large buffer size. The reason why the maximum buffering delay is adopted is that it directly corresponds to the application's requirements.

We assume that the link capacity is 150 Mbps, and do not take into account the buffering delays at intermediate switches. Assuming that all connections generate the same video sequence and require identical QoS in terms of cell loss ratio, CLR = 10^{-6}, in the VBR service class, the number of multiplexed connections is derived from Eq.(1). The result for "*Scenery*" is shown in Fig. 9 as a function of the allowable delay. This figure corresponds to Fig. 7 for the CBR service class and Fig. 8 and Eq.(1) for the VBR service class. This figure shows that SMG is no more than one!, i.e., the maximum number of multiplexed connections in the VBR service class is smaller than that of the CBR service class under the same de-

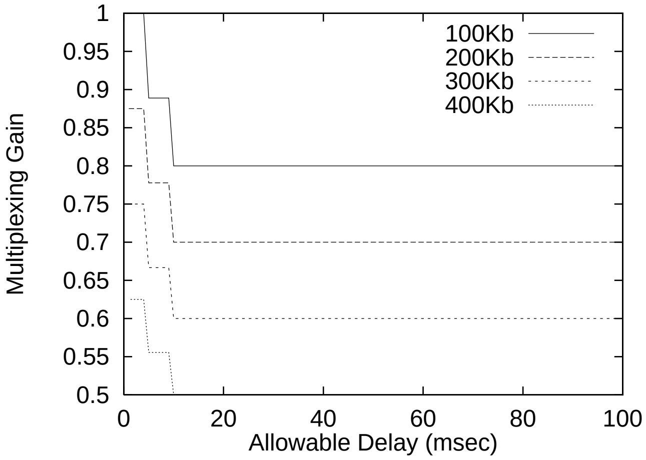

Fig. 10. Statistical multiplexing gain (*Scenery*, LinkCapacity=150 Mbps, CLR=10^{-3}, no-buffer)

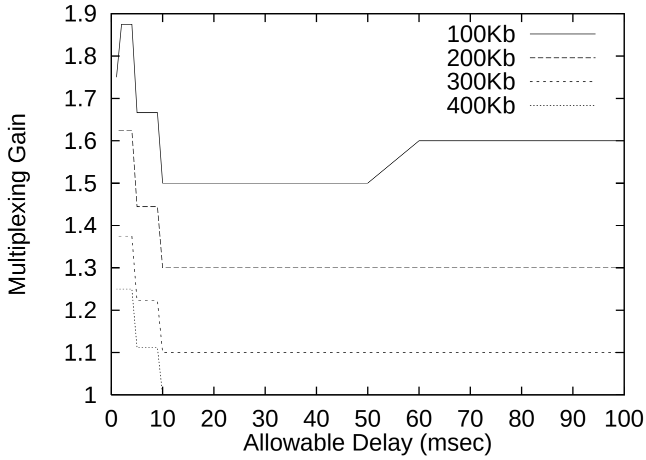

Fig. 12. Statistical multiplexing gain (*Scenery*, LinkCapacity=150 Mbps, CLR=10^{-6}, 10,000 cells buffer)

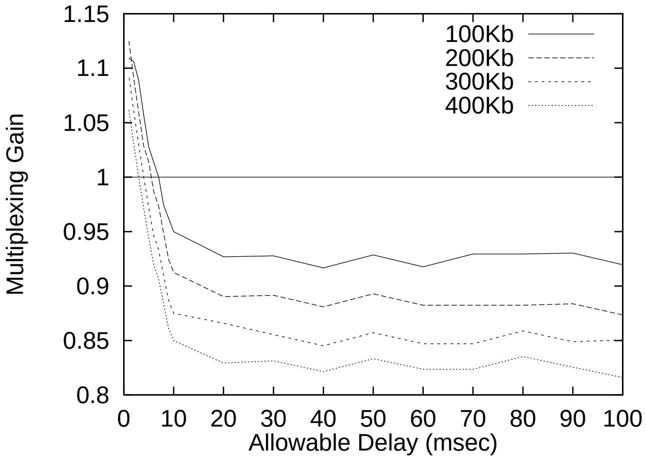

Fig. 11. Statistical multiplexing gain (*Scenery*, LinkCapacity=1.3 Gbps, CLR=10^{-3}, no-buffer)

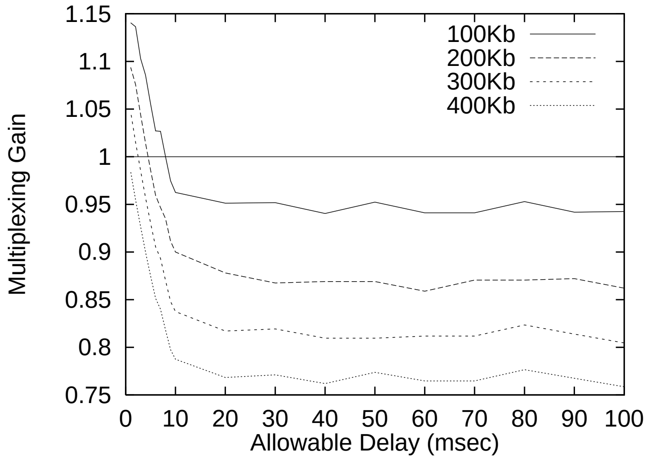

Fig. 13. Statistical multiplexing gain (*Scenery*, LinkCapacity=1.3 Gbps, CLR=10^{-6}, 10,000 cells buffer)

lay bound condition. This result arises from the fact that the maximum traffic rate (R_{max}) from the leaky bucket becomes larger than the allocated bandwidth when we take into account the token bucket of size σ in the VBR service class (see Section II.). For example, when the allowable delay is 10 msec, the required bandwidth in the CBR service class is 13.48 Mbps. On the other hand, the required drain rate becomes 13.13 Mbps in the VBR service class when the token bucket is 100 Kbits large. In the VBR service class, the drain rate itself is smaller than the required bandwidth in the CBR service class, but the maximum traffic rate instantaneously emitted to the network from the leaky bucket becomes 16.13 Mbps. Such a burstiness prevents the statistical multiplexing of the VBR service class from achieving a high SMG.

We investigate the way of increasing SMG. SMG is a function of several factors [14]. It can be increased by, e.g., (1) lower QoS (higher CLR in the current study), (2) larger link capacity, (3) larger buffer at intermediate switches, and (4) less bursty traffic. In what follows below, we will examine these cases.

First, we increase the acceptable CLR to see the effect of QoS on SMG. One may expect that larger SMG can be achieved when the acceptable CLR is increased (i.e., QoS requirement is relaxed). However, it is not true in our case. Improvement is very small as shown in Fig. 10 even if the acceptable CLR is increased from 10^{-6} to 10^{-3}. It is because the maximum video rate (17.28 Mbps) is comparatively too high to expect the large statistical multiplexing gain on the link of 150 Mbps. When the link capacity is set to be larger, SMG is actually improved. As shown in Fig. 11, the highest SMG becomes 1.13 with 200 Kbits token bucket. In this case, the link capacity is 1.3 Gbps and CLR is 10^{-3}. Of course, the perceived video quality is degraded due to cell losses. For example, the minimum SNR (Signal to Noise Ratio) of "*Scenery*" with no cell loss is 22.84, while it is degraded to 18.45 in the case of CLR = 10^{-3}. It means that one must make a sacrifice of video quality to achieve higher statistical multiplexing gain.

Another way to increase SMG is to introduce large cell buffers at intermediate switches. For example, when an intermediate switch is equipped with the buffer of 10,000 cells (although it is an extreme case and may be an unrealistic situation in the current memory technology), SMG of 1.875 can be achieved with 100 Kbits token bucket on 150 Mbps link as shown in Fig. 12. CLR is set at 10^{-6} in obtaining Fig. 12. Buffering delay in this figure includes both the buffering delay at the sender and the queueing delay at intermediate switches in

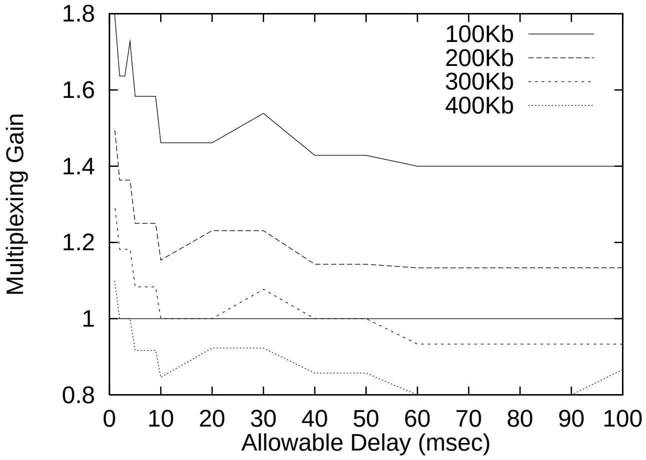

Fig. 14. Statistical multiplexing gain (*Starwars*, LinkCapacity=150 Mbps, CLR=10^{-6}, 10,000 cells buffer)

TABLE II
Statistical Multiplexing Gain (CLR=10^{-6})

Capacity	Buffer	*Scenery*	*Starwars*	*Live*
150Mbps	no-buffer	1.00 (100Kb)	1.00 (100Kb)	1.00 (100Kb)
	1300cells	1.13 (100Kb)	1.1 (100Kb)	1.17 (100Kb)
	10,000cells	1.75 (100Kb)	1.80 (100Kb)	1.92 (100Kb)
600Mbps	no-buffer	1.00 (100Kb)	0.95 (100Kb)	1.08 (300Kb)
	1300cells	1.03 (100Kb)	0.98 (100Kb)	1.12 (300Kb)
	10,000cells	1.22 (100Kb)	1.19 (100Kb)	1.39 (300Kb)
1.3Gbps	no-buffer	1.03 (100Kb)	1.02 (100Kb)	1.27 (300Kb)
	1300cells	1.05 (100Kb)	1.03 (100Kb)	1.29 (300Kb)
	10,000cells	1.14 (100Kb)	1.14 (100Kb)	1.43 (300Kb)

TABLE III
Statistical Multiplexing Gain (CLR=10^{-3})

Capacity	Buffer	*Scenery*	*Starwars*	*Live*
150Mbps	no-buffer	1.00 (100Kb)	1.00 (100Kb)	1.08 (100Kb)
	1300cells	1.13 (100Kb)	1.1 (100Kb)	1.25 (100Kb)
	10000cells	1.88 (100Kb)	2.00 (100Kb)	2.33 (300Kb)
600Mbps	no-buffer	1.06 (100Kb)	1.07 (100Kb)	1.41 (400Kb)
	1300cells	1.09 (100Kb)	0.98 (100Kb)	1.45 (400Kb)
	10000cells	1.31 (100Kb)	1.30 (100Kb)	1.78 (400Kb)
1.3Gbps	no-buffer	1.13 (200Kb)	1.12 (100Kb)	1.57 (400Kb)
	1300cells	1.14 (200Kb)	1.13 (100Kb)	1.59 (400Kb)
	10000cells	1.25 (200Kb)	1.24 (100Kb)	1.76 (400Kb)

the VBR service class. The effectiveness of a large amount of intermediate buffer decreases as the link capacity increases. As shown in Fig. 13, the largest SMG is only 1.14 even when the link capacity is changed to 1.3 Gbps. This degradation is due to the fact that SMG is the function of both the amount of buffer and the link capacity [see Eqs.(1) and (2)]. When an intermediate switch has no queueing buffer on 150 Mbps link, it can handle as much as 3 Mbits data (150 Mbps/30 fps) in a frame time interval (1/30 sec). When the 10,000 cells (3.84 Mbits) buffer is equipped with the intermediate switch, it can store and forward 2.2 times larger data (6.84 Mbits) on 150 Mbps link. On the other hand, the intermediate switch can handle 43.3 Mbits on 1.3 Gbps link without the network buffer and only 8.87 % increase is obtained with 10,000 cells buffer.

Another expectation which one may hold is that higher SMG can be achieved by applying less bursty traffic to the VBR service class. Recalling that "*Starwars*" is the least bursty traffic among three videos in Table I, we plot the results for "*Starwars*" video in Fig. 14. CLR is set to be 10^{-6} and the switch has 10,000 large buffer. The largest SMG then becomes 1.8 and higher than that of *Scenery*.

The largest values of SMGs under different conditions are summarized in Tables II and III. The conditions are varied according to combinations of the link capacity, 150 Mbps, 600 Mbps and 1.3 Gbps, the switch buffer size (no queueing buffer case, 1,300 cell buffer (about 500 Kbits) and 10,000 cell buffer). Leaky bucket sizes to achieve SMG are also shown in the tables within parentheses. With small drain rate and large switch buffer, the largest SMG can be achieved in all video sequences. However, the required amount of switch buffer is about 10,000 cells (about 3.84 Mbits). Further, to obtain the enough statistical multiplexing gain, the complicated CAC along with traffic descriptors and QoS parameters must be performed between the sender and networks at call setup time. UPC is also necessary after the connection is successfully established. Thus, UPC parameters should also be determined. It is especially difficult in the real-time video application to decide the control parameters appropriately for CAC and UPC. On the other hand, the CBR service class is easy to implement because only PCR is enough to perform CAC and UPC. Although it is also difficult to determine the PCR value exactly before actual encoding starts, it can be avoided to some extent by introducing the algorithm to control the encoding rate. One example of such an algorithm is MPEG-2 Test Model 5 [38], which we will briefly discuss in the next subsection.

D. Deterministic multiplexing of rate-controlled CBR videos in the CBR service class

As mentioned in the previous section, unpredictability of traffic characteristics in terms of PCR can be avoided by introducing rate-controlled video coding algorithms. With such algorithms, even if the sender fails in estimating adequate PCR and allocated bandwidth is insufficient, the encoder can adjust the coded video rate according to the allocated bandwidth. MPEG-2 Test Model 5 (abbreviated as TM5) is not only a video compression and coding algorithm, but also a rate control algorithm for video traffic. The rate control is accomplished by regulating the degree of quantization. The TM5 rate control algorithm can keep the average video traffic rate over a set of video frames to be lower than the allocated bandwidth. Thus, it makes possible for the sender to request the network to allocate the bandwidth without rigorous parameter estimation at call setup time. Then it regulates the video traffic rate according to the allocated bandwidth by the TM5 algorithm. However, as a result of quantization control, the coded video quality depends on the contents of picture, scene and sequence. And it is often claimed that the rate-regulated video quality is lower than the VBR video quality [44].

The relationship between the allocated bandwidth and the queueing delay in TM5 is shown in Fig. 15. As shown in the figure, the behavior of TM5 is absolutely different from the others. The queueing delay in TM5 is unchanged independently of the allocated bandwidth. Because TM5 regulates its traffic rate in order to avoid overflow and underflow of receiver's buffer, the

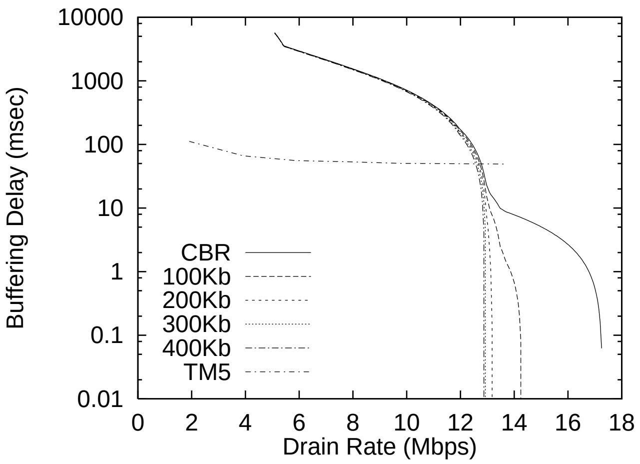

Fig. 15. Buffering delay in TM5 (*Scenery*)

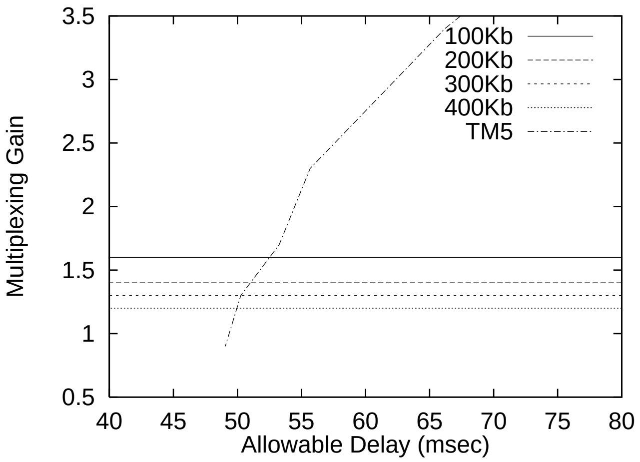

Fig. 16. Effectiveness of TM5 in multiplexing gain (*Scenery*)

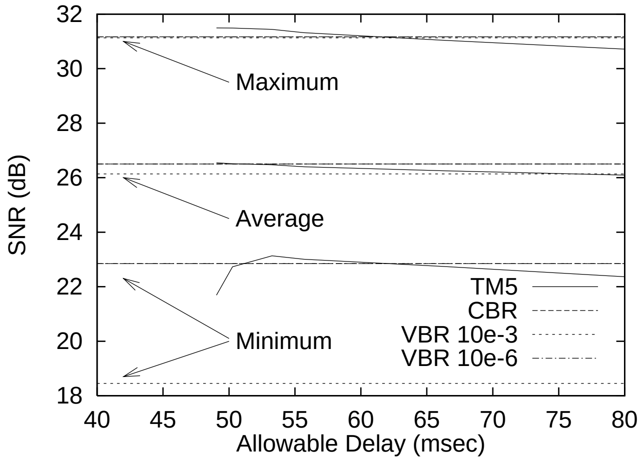

Fig. 17. Video quality comparison (*Scenery*)

buffering delay is kept almost unchanged. The smaller buffering delay in TM5 can be achieved by changing TM5 control parameter at the expense of perceived video quality. The achievable multiplexing gain on the 150 Mbps link is very high as shown in Fig. 16. SMGs of the VBR service class are also shown in Fig. 16 where the acceptable CLR is 10^{-3} and the queueing buffer at the intermediate node is 10,000 cells large for different token bucket sizes. As shown in Fig. 16, when the application can tolerate the delay of two frame time (66 msec), the number of connections which can be simultaneously multiplexed on the link with TM5 is 3.4 times larger than that of the CBR service class with buffering.

The number of connections multiplexed on a link is the network-related issue. In the case of multimedia application, we should take into account the quality of presentation, i.e., the perceived video quality at receivers. We compare SNR values of TM5 with the VBR and CBR service classes in Fig.17. The condition of comparison is the same as in the previous discussion. In the figure, the maximum, average and minimum values of SNR are shown against allowable delay bounds. For comparison purpose, the average SNR values of the CBR service class (labeled as "CBR"), the VBR service class where CLR are 10^{-3} ("VBR 10e-3") and 10^{-6} ("VBR 10e-6") are also shown in Fig.17. Lines of "VBR 10e-6" and "CBR" overlap each other. To achieve the larger multiplexing gain in TM5, the allocated bandwidth should be decreased. As a result, the perceived video quality is degraded, but the degradation is not striking. Since there is no cell loss in the CBR service class, the allowable delay does not affect the perceived video quality in terms of SNR. If we set CLR to be 10^{-6} in the VBR service class, there is almost no video quality degradation. When the allowable delay is small, i.e., the allocated bandwidth is large, the coded video quality is higher than that of the CBR service class as shown in Fig. 17.

From above observations, we can conclude that the achievable statistical multiplexing gain in the VBR service class is not very high as expected in spite of the fact that it is very hard to implement. Employing the CBR service class is easier and it is a more realistic way of achieving the effective video data transfer over ATM networks. Especially when MPEG-2 Test Model 5 algorithm is used, the number of multiplexed video traffic can be increased while the perceived video quality is kept high.

One obstacle in applying the VBR service class to video data transfer over ATM can be relaxed by introducing VBR+ service class where UPC parameters can be re-tuned during the call holding time. However, there still remains the implementation problem in the VBR+ service class, that is, how the characteristics of video traffic should be monitored, how UPC parameters are determined from characteristics of actual video traffic and how the new parameters are re-negotiated among the sender and networks.

IV. Conclusion

In this paper, we have investigated the effectiveness of statistical multiplexing in the VBR service class. For this purpose, we have used real MPEG-2 coded video data. We fairly compared the CBR and the VBR service classes while taking consideration on the buffering effect in the CBR service class. Through experiments, we have found that the highest SMG obtained with statistical multiplexing in the VBR service class is still lower than 2.5. This result means that SMG is not very high while there are many obstacles to perform the statistical multiplexing. One might assert that the deterministic bandwidth allocation in the CBR service class leads to waste of unused bandwidth because of the bursty nature of video traffic. However, the network can effectively fill the unused bandwidth with best-effort data traffic, as far as the guaranteed video traffic is given higher priority.

We conclude that the CBR service class is an easier and a more realistic way to achieve high multiplexing gain in transferring MPEG-2 video over ATM networks with high QoS guarantee. We have further examined the effectiveness and applicability of TM5 applicable in achieving higher deterministic multiplexing gain in the CBR service class. We revealed that TM5 considerably improves the effectiveness of multiplexing as long as the application can tolerate about 100 msec buffering delay and a little degradation of the perceived video quality.

References

[1] ISO/IEC IS 10918-1, 'Digital compression and coding of continuous-tone still images,' *ISO standard* (1993).

[2] ITU-T Recommendation H.261, *Video codec for audiovisual services at px64 kbits*, International Telecommunication Union (1993).

[3] ISO/IEC DIS 11172, 'Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s,' *ISO standard* (1991).

[4] ISO/IEC DIS 13818-2, 'MPEG-2 video,' *ISO standard* (1994).

[5] J. Kurose, 'Issues and challenges in providing Quality of Service guarantees in high speed networks,' *ACM Comp. Commun. Rev.*, **23**, 6–15, (January 1993).

[6] ITU-T Recommendation I.371, *Traffic control and congestion control in B-ISDN*, International Telecommunication Union (1992 revised in 1995).

[7] The ATM Forum, *ATM User Network Interface (UNI) Specification Version 3.1, 1/e*, Prentice Hall Professional Technical Reference (1995).

[8] M. W. Garrett, 'A service architecture for ATM: From applications to scheduling,' *IEEE Network*, **10**, 6–14 (May/June 1996).

[9] D. Mustill, 'Quality of service guarantees: A challenge for ATM networks,' in *Proc. IEEE GLOBECOM '96*, 420–423 (November 1996).

[10] Special Issue on, 'Flow and congestion control,' *IEEE Commun. Mag.*, **34** (November 1996).

[11] P. Newman, 'Traffic management for ATM local area networks,' *IEEE Commun. Mag.*, **32**, 44–51 (August 1994).

[12] N. Wakamiya, M. Murata, and H. Miyahara, 'Quality of MPEG transfer over ATM DBR service class with effective bandwidth usage,' *Proc. IEEE GLOBECOM'96* 1942–1948 (November 1996).

[13] S. Aign and K. Fazel, 'Temporal & sapatial error concealment techniques for hierarchical MPEG-2 video codec,' *Proc. IEEE ICC'95*, 1778–1783 (June 1995).

[14] R. O. Onvural, *Asynchronous Transfer Mode Networks: Performance Issues*, Artech House, Inc. (1994).

[15] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall International Editions (1992).

[16] E. P. Rathgeb, 'Policing of realistic VBR video traffic in an ATM network,' *Int. J. Digital and Analog Commun. Sys.*, **6**, 213–226 (1993).

[17] H. Heeke, 'Statistical multiplexing gain for variable bit rate video codecs in ATM networks,' *Int. J. Digital and Analog Commun. Sys.*, **4**, 261–268 (1991).

[18] A. R.Reibman and A. W.Berger, 'Traffic descriptors for VBR video teleconferencing over ATM networks,' *IEEE/ACM Trans. on Networking*, **3**, 329–339 (June 1995).

[19] E. W.Knightly, D. E.Wrege, J. Liebeherr, and H. Zhang, 'Fundamental limits and tradeoffs of providing deterministic guarantees to VBR video traffic,' *Proc. ACM SIGMETRICS '95*, 98–107 (May 1995).

[20] E. W.Knightly, 'H-BIND: A new approach to providing statistical performance guarantees to VBR traffic,' *Proc. IEEE INFOCOM '96*, 1091–1099 (March 1996).

[21] Z.-L. Zhang, J. Kurose, J. D.Salehi, and T. Towsley, 'Smoothing, statistical multiplexing, and call admission control for stored video,' *IEEE J. Select. Areas in Commun.*, **15**, 1148–1166 (August 1997).

[22] S. Singh and S. Chan, 'A multi-level approach to the transport of MPEG-coded video over ATM and some experiments,' *Proc. IEEE GLOBECOM'95* 1920–1924 (November 1995).

[23] M. W. Garrett and W. Willinger, 'Analysis, modeling, and generation of self-similar VBR video traffic,' *ACM SIGCOMM '94*, 269–280 (September 1994).

[24] J. Beran, R. Sherman, and M. S. Taqqu, 'Long-range dependence in variable-bit-rate video traffic,' *IEEE Trans. on Commun.*, **43**, 1566–1579 (February/March/April 1995).

[25] D. P. Heyman and T. V. Lakshman, 'What are the implications of long-range dependence for VBR-video traffic engineering?,' *IEEE/ACM Trans. on Networking*, **4**, 301–317 (June 1996).

[26] J. C. Wu, Y. Chen, and K. Jiang, 'Modeling and performance study of MPEG video sources over ATM networks,' *Proc. IEEE ICC'95*, 1747–1750 (June 1995).

[27] M. Krunz, R. Sass, and H. Hughes, 'Statistical characteristics and multiplexing of MPEG streams,' *Proc. IEEE INFOCOM'95*, 455–462 (April 1995).

[28] G. Mercankosk and Z. L. Budrikis, 'Establishing a real time VBR connection over an ATM network,' *Proc. IEEE GLOBECOM '96*, 1710–1714 (November 1996).

[29] W. Luo and M. E. Zarki, 'Quality control for VBR video over ATM networks,' *IEEE J. Select. Area in Commun.*, **15**, 1029–1039 (August 1997).

[30] M. Hamdi, J. W.Roberts, and P. Rolin, 'Rate control for VBR video coders in broad-band networks,' *IEEE J. Select. Areas in Commun.*, **15**, 1040–1051 (August 1997).

[31] W. Luo and M. E. Zarki, 'Transmitting scalable MPEG-2 video over ATM based netwoks,' *Proc. First Workshop on ATM Traffic Management WATM'95*, 149–156 (1995).

[32] D. Park and K. Kim, 'Buffered rate-distortion control of MPEG compressed video channel for DBS applications,' *Proc. IEEE ICC'95*, 1751–1755 (June 1995).

[33] O. L. Stokes and A. A. Nilsson, 'Analysis of MPEG-1 transmission through B-ISDN ATM networks,' *Proc. IEEE INFOCOM'96*, 545–552 (March 1996).

[34] H. Kanakia, P. P. Mishra, and A. R. Reibman, 'An adaptive congestion control scheme for real time packet video transport,' *IEEE/ACM Trans. on Networking*, **3** (December 1995).

[35] P. Pancha and M. E. Zarki, 'Bandwidth requirements of variable bit rate MPEG sources in ATM networks,' *Proc. IEEE INFOCOM '93*, **3**, 902–909 (March 1993).

[36] P. Pancha and M. E. Zarki, 'MPEG coding for variable bit rate video transmission,' *IEEE Commun. Mag.*, 54–66 (May 1994).

[37] L. Chou and C. Chang, 'Experiments of the theory of effective bandwidth for markov sources and video traces,' *Proc. IEEE INFOCOM'96*, 497–504 (March 1996).

[38] ISO/IEC/JTC1/SC29/WG11, 'Test model 5,' (April 1994).

[39] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, 'Resource reservation protocol (RSVP) – version 1 functional specification,' *Internet Draft* (November 1996).

[40] L. Delgrossi and L. Berger, 'Internet stream protocol version 2 (ST2) protocol specification - version ST2+,' *Internet Draft/RFC1819* (August 1995).

[41] S. Deering and R. Hinden, 'Internet protocol, version 6 (IPv6) specification,' *Internet Draft/RFC1883* (December 1995).

[42] M. Butto, E. Cavallero, and A. Tonietti, 'Effectiveness of the 'leaky bucket' policing mechanism in ATM networks,' *IEEE J. Select. Areas in Commun.*, **9**, 335–342 (April 1991).

[43] R. L. Cruz, 'A calculus for network delay, part I: Network elements in isolation,' *IEEE Trans. on Info. Theory*, **37**, 114–131 (January 1991).

[44] C.-Y. Hsu, A. Ortega, and A. R.Reibman, 'Joint selection of source and channel rate for VBR video transmission under ATM policing constraints,' *IEEE J. Select. Areas in Commun.*, **15**, 1016–1028 (August 1997).