# Adaptive VBR Video Traffic Management for Higher Utilization of ATM Networks

Girish Chiruvolu[1], Ravi Sankar[2] and N. Ranganathan[1]

1. Dept of Comp Sci & Eng. 2. Electrical Eng.

Univ. of South Florida, Tampa, Florida-33620

email: {girish, sankar, nrangana}@sunflash.eng.usf.edu

**Abstract**

The VBR video traffic exhibits high burstiness and correlation properties that are quite complex to be captured by a single traffic model. Efficient resource management based on few parameters of the source traffic is highly desirable. The real-time VBR video traffic has stringent quality of service (QoS) requirements such as delay (few milliseconds) and cell loss (1 in $10^{-5}$) that are difficult to achieve with good utilization ($> 0.6$) by static bandwidth allocation schemes. In order to satisfy such QoS constraints with good utilization, proper adaptive mechanisms have to be devised. This paper presents a dynamic bandwidth allocation scheme for VBR video traffic based on buffer monitoring and a simple LMS (least mean square) traffic prediction system. The goal is to reduce the frequency of the bandwidth reallocations and at the same time reduce the Cell-loss Ratio (CLR) with increased utilization. Simulation results indicate that utilization up to 0.8 can be achieved by the proposed scheme even under high source alignment [26] for bursty VBR video traffic. It is found that the proposed adaptive scheme outperforms the static FCFS allocation scheme with lower buffer requirements and fewer ($< 5\%$) bandwidth reallocations.

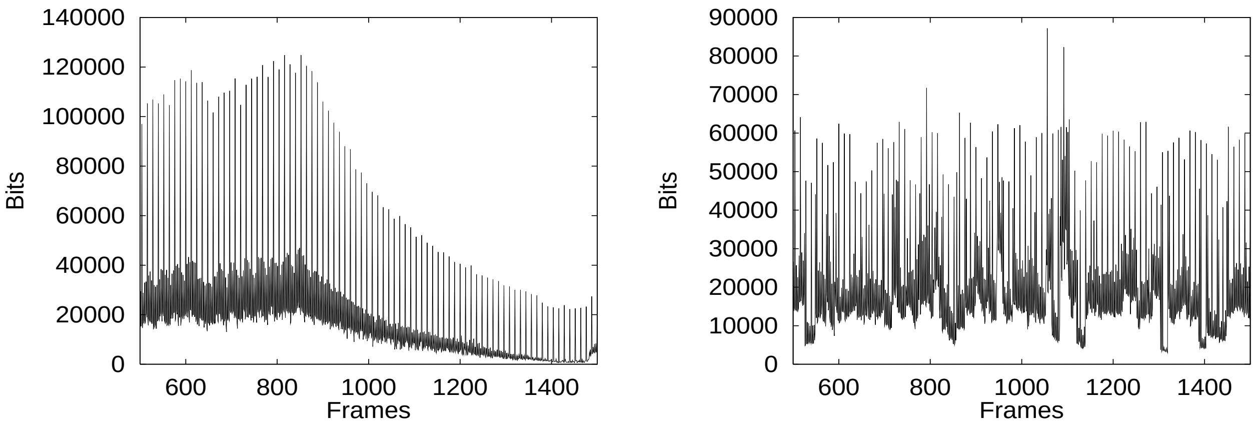
**Keywords**

VBR video traffic prediction, Dynamic bandwidth allocation, ATM network.

## I. Introduction

Digital video traffic is expected to be a major component of the ATM based B-ISDN networks.The video traffic has stringent real-time constraints on parameters such as cell loss ratio (CLR) and delay that determine the quality of service (QoS). ATM offers the advantage of statistical multiplexing gain (SMG) to efficiently utilize the network resources. The SMG has to be chosen such that the QoS constraints are satisfied simultaneously, leading to a better bandwidth utilization. The video packets may be transmitted using Constant Bit Rate (CBR) or in Variable Bit Rate (VBR) modes. Efficient video codecs invariably generate bursty VBR traffic with a peak-to-mean ratio of 10. CBR approach has simple peak rate bandwidth allocation scheme that often leads to low utilization. One can easily meet the desired CLR and delay requirements at low utilization factors. In ATM network, Available Bit Rate (ABR) data services utilize the left-over bandwidth by the VBR services. Currently, research efforts are focussed on developing algorithms to efficiently assess the available bandwidth and share the left-over bandwidth among the ABR sessions based on some fairness criteria such as *max-min* [3], [16], [19]. Nevertheless, in order to find the left-over bandwidth, one needs to accurately estimate the bandwidth requirements of VBR traffic that satisfy QoS constraints with high utilization.

Novel dynamic bandwidth allocation schemes are necessary in order to efficiently utilize the network resources (e.g., bandwidth, buffers) and maximize the number of video sessions that can be supported with existing resources. While the dynamic bandwidth allocation schemes improve the utilization through proper adaptation to the VBR video traffic rates, frequent adaptation is undesirable. Most of the techniques in the literature focus on a session-based renegotiation schemes that involve overhead of fast renegotiating protocols and round-trip delay for the resource management. The prediction based adaptation schemes are shown to be very effective, especially in the context of highly correlated video traffic [1,4-7,21]. The statistical multiplexing gain across the video sources can be exploited to increase the utilization. Nevertheless, with constant service rate, at utilization factors beyond 0.6, large buffers are necessary to avoid a large CLR. This is mainly due to the large impact of high correlation and burstiness properties of the video traffic on the queueing system. Constant monitoring of buffer is necessary to avoid buffer overflows (cell loss) at higher utilization factors. A few backlog (buffered cells) based dynamic bandwidth allocation schemes have been proposed in the literature [15],[20],[23]. Most of these approaches are based on actual measurement of the cell loss and need a large time-window in order to assess the CLR of order $10^{-5}$. This restricts the reactive control to minimize the CLR that is very much essential for correlated bursty video traffic. It is difficult to estimate from smaller measurement intervals, a CLR of order $10^{-5}$. Nevertheless, they can be useful in providing a quick reactive control. Intuitively, a higher utilization with lower cell loss can be achieved by lowering the utilization during the backlog build-up period and maintaining a higher utilization during other times. This paper explores one such method through traffic prediction and buffer monitoring. The proposed scheme is based on traffic prediction and is a variant of backlog based dynamic bandwidth allocation [20].




(a) Starwars (b)MTV traces

Fig. 1. The VBR video traces

The paper is organized as follows. Section 2 provides a brief discussion on the VBR video traffic (MPEG) characteristics. Issues that are relevant to VBR video traffic management are also examined in section 2. Section 3 proposes a new dynamic bandwidth estimation scheme with low complexity that is suitable for implementation in high-speed networks. Section 4 evaluates the scheme with real traces and some of the results are presented. A summary and directions for future research are given in section 5.

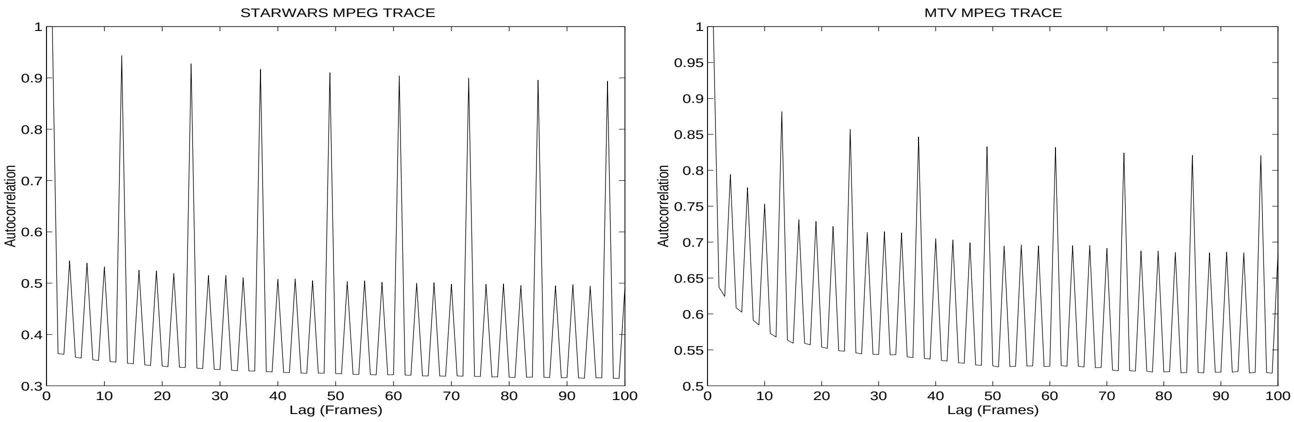
## II. Characteristics of Video Traffic and their implications

This section provides a brief description of the characteristics of VBR video MPEG traces that are used in this study. Encoding of the digitized video signal was performed using the UC Berkeley MPEG encoder at 24 frames/sec. The encoded frames are of three types: i) Intra (I) ii) Predictive (P) and iii) Bidirectional (B). Among these, the I frames have higher bandwidth requirements as the temporal relationships between the forward and backward frames are not exploited. The B frames have the least bandwidth requirements. After encoding, the frames are arranged as "IBBPBBPBBPBB". Figure 1 depicts the traffic generation per frame for *Starwars* and *MTV* MPEG sequences [9], [25].

Figure 2 shows the typical autocorrelation characteristics of the VBR video traffic. It is shown that the mean queue length of correlated input traffic is much higher (by an order of 10) than uncorrelated input traffic [18], [22], [29]. Video traffic, which is traditionally modeled using short-range dependent (SRD) Markov-Modulated Poisson Processes [28], is shown to exhibit self-similar and long-range dependent (LRD) characteristics [2],[9]. The LRD is defined in the sense $\sum_k \rho(k) = \infty$, where $\rho(k)$ is the correlation for lag $k$. The

SRD processes, in contrast to LRD, have summable correlations, i.e., $\sum_k \rho(k) < \infty$.

The LRD processes are typically characterized by Hurst (H) parameter [2], [9]. The correlations decay very slowly (hyperbolic rather than exponential) with lag (frames). The effect due to correlated input traffic on a queueing system raises the question: On what time scale, the correlation (up to what time-lag and the magnitude of correlation) of the input real-time traffic that needs to be taken into account in a finite buffer queueing system?




(a) Starwars (b)MTV traces

Fig. 2. Correlation properties of VBR video traces

Recently, there have been a number of studies focusing on the relevance of the LRD (in terms of correlation) properties of the VBR video traffic with respect to cell loss and delay for the case of finite buffer queueing systems [10], [14], [27]. One of the main results of these studies is the correlations of the input traffic have a critical impact up to a finite time-scale, Critical Time Scale (CTS), as far as the performance measures such as cell loss and delay are concerned. The CTS is shown to attain a small value that depends on the finite buffer of a queueing system. Quantitatively, the CTS is estimated to be proportional to the buffer size and inversely proportional to the standard deviation of the marginal distribution of input traffic [10]. On one side, for a zero buffer queueing system, the correlation of the input traffic has no effect at all on queue lengths and delays. On the other side, the LRD properties have a dominant impact on the queueing system with infinite buffer [8]. Due to the delay constraints of the real-time video traffic, large buffers (equivalent to few hundred milliseconds or greater) are unnecessary and thus it is concluded that SRD rather than LRD correlation properties should be considered for performance evaluation [10], [27]. Thus, the time scale and the buffer sizes play an important role in the queueing performance evaluation.

The video-conferencing, that typically does not have frequent scene changes, generates less bursty traffic than a high-activity (frequent scene changes) broadcast video such as *Starwars* [9] and *MTV* [25]. The VBR video-conference traffic has been successfully modeled with the help of discrete autoregressive process [12]. A simple unified traffic model for broadcast video traffic that captures the complex nature of the VBR video traffic with various tail distributions that are observed in traces is yet to be designed [13]. So, call admission control schemes may have to rely on approximate estimates of the traffic source parameters. This may lead to conservative utilization of network resources. Thus online traffic measurement based dynamic bandwidth allocation schemes are very essential to increase utilization.

The contributions of the research reported in this paper are as follows. A traffic prediction-based dynamic bandwidth allocation scheme that achieves higher target utilization with less frequent reallocations has been developed. We also studied buffer timescales at which the cell loss and delay can be optimized for any backlog-based adaptive schemes.

## III. The proposed dynamic bandwidth allocation scheme

The VBR video services typically require a CLR of the order $10^{-5}$ or lower [25]. Furthermore, random cell loss may be tolerated by some MPEG decoders, but experimental results show that due to strong correlation properties exhibited by the video traffic, the cell loss occurs in bursts rather than in isolation [26]. The SMG across the sessions can be exploited to reduce the cell loss due to the traffic bursts to some extent. Nevertheless, there can be instances with finite probability at which the aggregate traffic may well exceed the server capacity leading to buffer overflows. Once the backlog builds up, buffer overflows are inevitable unless appropriate preventive actions are taken such as increasing the bandwidth, especially in the context of bursty video traffic. As mentioned before, call admission control schemes are based on inaccuracies of the source traffic parameters, as accurate models that describe the various VBR video sources are difficult to be specified at call admission time. Hence online traffic measurement based adaptive techniques are very essential to properly assess and satisfy the QoS requirements.

An LMS adaptive filter is employed to predict the traffic at regular intervals. The LMS is well known for its simplicity and yet delivers effective output in an adaptive filtering environment. The LMS adaptive scheme is as follows.

$$\hat{x}_{n+1} = W_n^T X_{n-1} \tag{1}$$

$$\epsilon_n = x_n - \hat{x_n} \tag{2}$$

$$W_{n+1} = W_n + 2\mu\epsilon_n X_n \tag{3}$$

where $\hat{x_n}$ is the predicted value and $W_n$ is the vector of adaptive filter coefficients at the $n$th update. $X^T_{n-1} = [x_{n-1}, ..., x_{n-N}]$ is the actual measured signal (input traffic), in cells/sec and $T$ denotes transpose of the corresponding vector. $\epsilon_n$ is the prediction error and $\mu$ is the adaptation gain constant that regulates the speed and stability of adaptation algorithm.

We introduce a *Buffer_Threshold (BT)* parameter that takes a value in the interval (0, 1] for buffer monitoring. Let $BUF$ be the total available buffer size in cells. A backlog greater than $BT$*$BUF$ cells triggers an adaptive bandwidth reallocation. One of our goals is to vary $BT$ to study the CLR and the frequency of adaptation for various buffer sizes. Let the target utilization be $\rho_{target}$, and the long-term average input traffic rate be $\lambda_{long\_term}$ (assumed to be specified during call admission or can be obtained from online traffic measurements). From the above two, the server capacity $\mu_{target}$ can be found and this is the server capacity for VBR video traffic during low buffer-occupancy time (i.e, less than $BT * BUF$). The LMS adaptive scheme is used for traffic prediction.

The proposed scheme works as follows. Let $\hat{x_n}$ be the traffic predicted for the $nth$ interval. Let $b_{n-1}$ be the backlog adaptation interval. The required server capacity $c_n$ for the aggregate VBR video traffic in the $nth$ interval is estimated as

$$c_n = \hat{x_n} + b_{n-1} \tag{4}$$

If the current backlog $b_{n-1}$ is less that $BT * BUF$, then no bandwidth reallocation takes place and the server capacity is maintained at $a_n = \mu_{target}$, where $a_n$ denotes allocated capacity. Otherwise, we allocate the maximum of $\{ c_n, a_{n-1} \}$ for the $n$th interval (i.e., $a_n = \max\{c_n, a_{n-1} \}$). It is assumed that the estimated bandwidth by the proposed scheme is always available.

The server can be modeled as having two states. One state represents the safe state where in the backlog is less than $BT * BUF$ and the system serves the cells at $\mu_{target}$ rate. The other state represents the unsafe state, wherein the backlog is greater than $BT * BUF$ and a bandwidth reallocation is triggered for the future adaptation intervals. It is trivial to observe that the system must enter an unsafe state before any cell loss can occur. We also define the predictor to be in two states. One state being the lag-phase wherein the predictor underestimates the bandwidth ($\epsilon_n < 0$). The other state is lead-phase wherein the predictor overestimates the bandwidth ($\epsilon_n > 0$) or accurately estimates the bandwidth ($\epsilon_n = 0$).

**Proposition:** The system must be in an unsafe state for cell loss to occur. If there exists an unsafe period of $\tau$ consecutive adaptation intervals, then the predictor is continuously in lag-phase with estimation error greater than $BT * BUF$ for the $\tau$ intervals. The cell loss occurs, if at all, during those unsafe and lag-phases only.

Let the above mentioned unsafe period last for a duration of $\tau$ consecutive adaptation intervals starting from an arbitrary time $n$. Let $(x)^+$ denote the maximum of zero and x.

Case 1: The predictor cannot be in lead-phase during the $\tau$ consecutive adaptation intervals.

Proof is trivial for the case $\tau = 1$. For the cases $\tau > 1$, the proof is by contradiction. Suppose, in contrary to the proposition, the predictor is in lead-phase for any $k$th adaptation interval ( $n < k + 1 < n + \tau$) in which the system is in an unsafe state. Let $b_k$ be the backlog at the end of $k$th interval. Now the total estimated capacity for the $k$th interval is $c_k = \hat{x_k} + b_{k-1}$ and the actual allocated capacity is $a_k = \max\{c_k, a_{k-1}\}$. If $c_k$ is allocated, then the backlog at the start of $k$th adaptation interval is given by $b_k = (b_{k-1} + x_k - c_k)^+$. Hence $b_k = (x_k - \hat{x_k})^+$ which is zero as the predictor is in lead-phase, by assumption. Thus $b_k$ is less than $BT * BUF$ and the server moves to a safe state at the end of the $k$th adaptation interval. Since $k$ is less than $n + \tau$, we arrive at a contradiction that the unsafe period lasted for $\tau - 1$ consecutive intervals. If $a_k$ is selected from $\max\{c_{k+1}, a_k\}$, then $a_k > c_{k+1}$, thus by the same argument above, the system has to transit to safe state at the end of $k + 1$th interval making the unsafe period last for a duration less than $\tau$. Thus, the predictor cannot be in lead-phase during unsafe period. Note that vice-versa, predictor cannot lag during safe periods, is not necessarily true.

Case 2: Estimation error has to be greater than $BT * BUF$ during lag-phase for unsafe duration of $\tau$ consecutive intervals.

Here the predictor is in lag phase. Proof is again by contradiction. Assume that the estimation error be $\epsilon_k = \hat{x_k} - x_{k-1} < BT * BUF$. Now, $c_k = \hat{x_k} + b_{k-1}$. According to the algorithm, the allocated capacity for $k$th interval is $a_k = \max\{c_k, a_{k-1}\}$. If $c_k$ is allocated, then it is obvious that $b_k = (b_{k-1} + x_k - c_k)^+ = \epsilon_k$ will be less than $BT * BUF$. Thus the system moves into a safe state at the end of $k$th interval that is less than $n + \tau$, a contradiction that the unsafe state lasted for $\tau - 1$ consecutive adaptation intervals. If, $a_{k-1}$ is chosen from $\max\{c_k, a_{k-1}\}$ then, again a similar above argument leads to contradiction. Note that $a_k$ cannot be greater than $b_{k-1} + x_k$ for the system to continue in an unsafe state. Thus, the estimation error has to be greater than $BT * BUF$ is a necessary but not a sufficient condition for system to continue in an unsafe state.

**Lemma:** Not all unsafe, lag-phases lead to cell loss.
This is based on simple observation that if $BT * BUF < \epsilon_k < BUF$, for all $k < \tau$, then no cell loss occurs during $\tau$. Here $\epsilon_k = \hat{x_k} - x_k$ is the estimation error and $\tau$ is the length of unsafe lag-phase period.

The above two properties indicate that if the estimation error is bounded by at least $BT * BUF$ or better, then it gives rise to less frequent reallocations and cell loss. Note that the burstiness of the input traffic in addition to the correlation properties, determines the rate of transition of the server from safe to unsafe states. This is further explored in the next section through simulations.

## IV. Simulation results

In this section we analyze the proposed scheme via simulations using real video traces [9], [25]. The ATM switch is modeled at a rate of 622 Mbits/sec. The adaptation interval is fixed at 40 ms (video frames rate). Eight independent video streams are multiplexed with FCFS scheduling. In order to assess the worst case scenario, a high source alignment is maintained (i.e., I frames of various sessions arrive simultaneously).

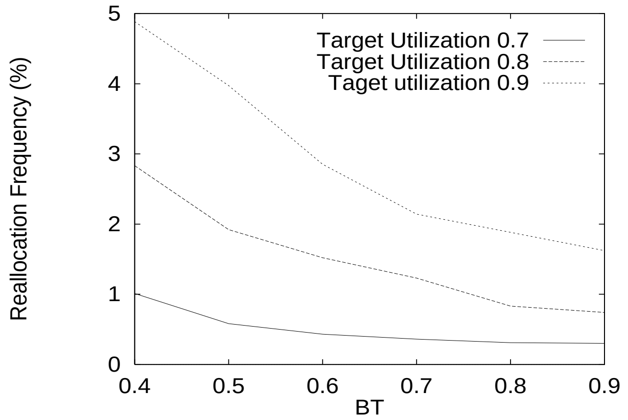



Fig. 3. BT vs. Reallocation Frequency (%) for buffer size of 600 cells

Figure 3 shows the frequency of reallocations as $BT$ is varied by keeping the buffer size fixed at 600 cells. As expected, the frequency of reallocations decreased with increase in $BT$ value, As described earlier, $BT$ is the backlog threshold beyond which a reallocation of bandwidth takes place for VBR video traffic. Since our goal is to reduce the frequency of bandwidth reallocations, we studied the system for $BT > 0.4$. For values of $BT$ less than 0.4, a better queueing response can be achieved at the cost of more frequent adaptations. Figure 4 depicts the CLR with $BT$ varied from 0.4 to 0.9 for target utilizations of 0.7 and 0.8.

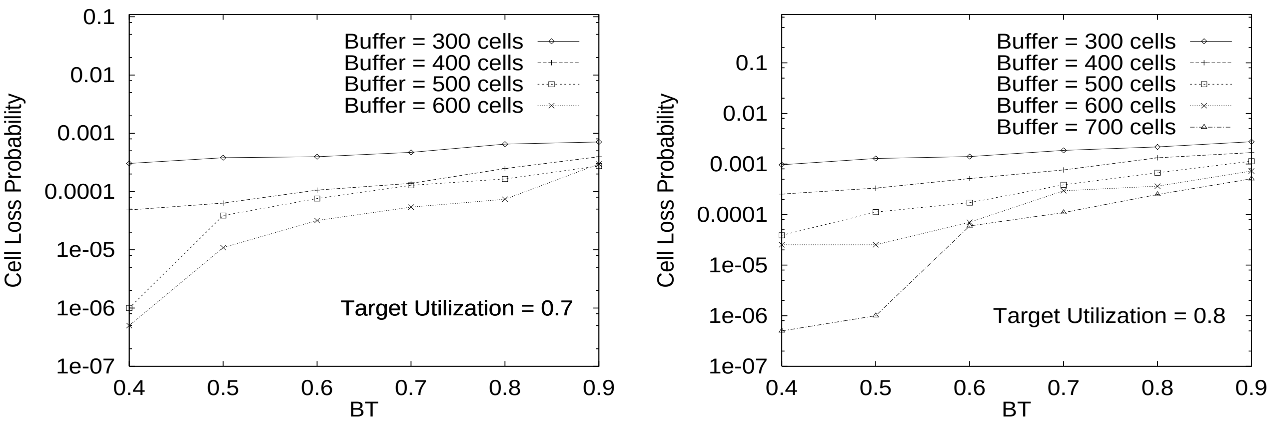



Fig. 4. CLR vs. Buffer Threshold

We define minimal buffer size for a given utilization as the minimum buffer with which the adaptive scheme can achieve a desired CLR of say $10^{-5}$. For example the minimal buffer size for an utilization of 0.7 is 600 cells and for 0.8 utilization, it is 700 cells when a CLR of $10^{-5}$ is considered. The proposed scheme is very effective in maintaining the CLR below $10^{-5}$ for values of $BT$ up to 0.55 with a corresponding minimal buffer size. Thus *half-buffer time scale* for minimal buffer size assumes importance especially in the context of correlated and bursty video traffic.

As $BT$ is increased towards 1, the CLR increased considerably and assumed values of order $10^{-3}$ for a given buffer size ($< 600$ cells). This can be attributed to the lack of proper adaptation. The late reactive control and the high likelihood of error in the predictor estimates during traffic bursts result in considerable cell loss. Such a reactive control can only be of limited use when burstiness of the input traffic is high (e.g., VBR video traffic). In order to maintain the CLR of $10^{-5}$, the buffer has to be appropriately increased with increase in utilization factor. As the utilization factor is increased from 0.7

TABLE I

PERFORMANCE FOR TARGET UTILIZATION OF 0.7 AND BUFFER SIZE 600 CELLS

| $BT$ | Mean delay (secs) | Reallocation Frequency (as %) | Achieved Utilization |
|---|---|---|---|
| 0.4 | 0.000273 | 1.0112 | 0.683 |
| 0.5 | 0.000332 | 0.5807 | 0.686 |
| 0.6 | 0.000369 | 0.4305 | 0.6885 |
| 0.7 | 0.000424 | 0.3604 | 0.6897 |
| 0.8 | 0.000437 | 0.3104 | 0.6903 |
| 0.9 | 0.000460 | 0.3004 | 0.696 |

TABLE II

PERFORMANCE FOR TARGET UTILIZATION OF 0.8 AND BUFFER SIZE 700 CELLS

| $BT$ | Mean delay (secs) | Reallocation Frequency (as %) | Achieved Utilization |
|---|---|---|---|
| 0.4 | 0.000876 | 2.0324 | 0.783 |
| 0.5 | 0.001158 | 1.3216 | 0.792 |
| 0.6 | 0.001500 | 1.0212 | 0.7915 |
| 0.7 | 0.001801 | 0.8310 | 0.7947 |
| 0.8 | 0.002140 | 0.7309 | 0.7963 |
| 0.9 | 0.002500 | 0.6408 | 0.798 |

to 0.8, the buffer has to be increased approximately by 15% in order to maintain the CLR at $10^{-5}$ (see Figure 4).

Even though the CLR is minimized with increase in buffer size, the mean delay, the frequency of bandwdith reallocation, and the achieved utilization have to be considered in order to fully evaluate the proposed scheme. Tables 1-3 summarize the relevant statistics for the utilization factors of 0.7, 0.8 and 0.9, respectively. It is observed that for the target utilization of 0.9, there is a substantial increase in mean delay when compared with those in the cases of utilization factors of 0.8 and 0.7 due to the large buffer size and lower bandwidth. As shown in Table 3, the target utilization of 0.9 could not be achieved. This is attributed to the fact that for higher utilization factors, the delay and cell loss requirements cannot be satisfied simultaneously for correlated bursty VBR video traffic. Thus a tradeoff between the utilization, delay and cell loss requirements for VBR video traffic is required.

TABLE III

PERFORMANCE FOR TARGET UTILIZATION OF 0.9 AND BUFFER SIZE 800 CELLS

| $BT$ | Mean delay (secs) | Reallocation Frequency (as %) | Achieved Utilization |
|---|---|---|---|
| 0.4 | 0.002417 | 3.8847 | 0.857060 |
| 0.5 | 0.003295 | 2.9736 | 0.862469 |
| 0.6 | 0.004324 | 2.3528 | 0.867961 |
| 0.7 | 0.005722 | 1.9423 | 0.872029 |
| 0.8 | 0.006911 | 1.7821 | 0.872226 |
| 0.9 | 0.008287 | 1.5218 | 0.875535 |

TABLE IV

PERFORMANCE COMPARISON BETWEEN ADAPTIVE SCHEME AND STATIC FCFS ALLOCATION SCHEME FOR $\rho_{target} = 0.8$

| The proposed adaptive scheme | | | Static allocation scheme | |
|---|---|---|---|---|
| Buffer Size (cells) | Mean delay (secs) | CLR | Mean delay (secs) | CLR |
| 300 | 0.000528 | 1.285780e-03 | 0.00168965 | 1.069e-02 |
| 400 | 0.000631 | 3.331310e-04 | 0.0025223 | 0.894717e-02 |
| 500 | 0.000788 | 1.121860e-04 | 0.00342902 | 8.02662e-03 |
| 600 | 0.000961 | 2.524190e-05 | 0.0043494 | 7.30364e-03 |
| 700 | 0.001158 | 1.503000e-06 | 0.00530169 | 6.74645e-03 |

The equivalent bandwidth allocation that is based on Gaussian approximation [11] is given by

$$C = \mu_{aggregate} + \sigma_{aggregate}\sqrt{-2ln(p) - ln(2\pi)} \tag{5}$$

where $\mu_{aggregate} = \Sigma\mu_i$, $\sigma_{aggregate} = \sqrt{\Sigma\sigma^2{}_i}$, and $\mu_i$, $\sigma_i$ are the average and standard deviation of cells/sec for the $i$th session. The probability of overflow is approximated as $p = P[X_{aggregate} > C]$ from the above equation (5). Based on the above equivalent bandwidth approximation, for a CLR of $10^{-5}$, the utilization that can be achieved with the same traces used in the above simulations, is estimated to be around a low value of 0.5. This can be explained based on the fact that the approximation is mostly based on tail distribution of the input traffic. The proposed adaptive scheme can achieve an utilization of as high as 0.8 even under source alignment.

Finally, Table 4 compares the performance of the proposed scheme for $BT = 0.5$ and target utilization of 0.8 with a static FCFS allocation scheme. The adaptive scheme outperforms the static allocation scheme for the given utilization of 0.8 both in terms of mean delay and CLR. Unlike the adaptive scheme, the static allocation scheme could not decrease the CLR by considerably increasing the buffer size. It is found that even for a large buffer size of 3000 cells, the static allocation scheme achieved a CLR of $10^{-3}$ for an utilization of 0.8. In comparison, the proposed adaptive scheme achieved a CLR of $10^{-5}$ with less than 3% bandwidth reallocation frequency for a buffer size of 800 cells. This shows that increasing the buffer size with static allocation schemes is only of very limited help. Similar conclusions are reported in [26]. The adaptive scheme for VBR video traffic enhances the utilization factor considerably satisfying the QoS requirements of delay and CLR simultaneously. Thus, we conclude that the proposed adaptive scheme drastically improves the queueing performance with a low bandwidth reallocation frequency that is suitable for implementing in high-speed networks.

## V. Conclusions and future research

The issues and approaches relevant to VBR video traffic management can be summarized as follows. The statistical characterization of VBR video traffic is difficult mainly due to the fact that video streams exhibit diverse characteristics. The aggregation of video streams poses much more challenge in characterizing the statistical gain that can be obtained across the sessions in order to utilize the resources effectively.

These factors lead to the following:

- The call admission control algorithms have to make decisions based on prior statistical characterization of VBR video traffic that may not be accurate enough and have to rely on prior approximate bandwidth estimates.
- Online adaptation to the changing video traffic rates is essential due to the delay constraints imposed by real-time VBR video. Dynamic bandwidth allocation schemes lead to better utilization of network resources.
- Session-wise frequent bandwidth renegotiations are undesirable due to the overhead and the delay involved in the process of renegotiation. The proposed adaptive scheme can be effectively employed to avoid such session-wise frequent renegotiations.
- Half-buffer time scale can be effectively used for buffer monitoring in order to achieve low CLR.

The proposed scheme is simple and based on online traffic measurements coupled with buffer monitoring that can achieve good utilization factor. This scheme can be efficiently implemented in hardware and one such VLSI implementation can be found in [24]. The proposed scheme can mitigate the ill-effects of the inaccuracies of the source models upon which the call admission control relies, through online traffic estimation. Nevertheless, in order to achieve realistic performance bounds, aggregate traffic models have to be devised for VBR video traffic so that utilization, buffer size, CLR, mean delay can be optimized for the proposed scheme.

The traces that were employed in this study have a typical peak-to-mean ratio around 5. However, due to the strong correlational properties, the VBR video traffic is predictable. The proposed model is equally applicable to other VBR video traffic having a higher peak-to-mean ratio. Thus, the continous adaptation to the traffic rates and the exploitation of the statistical gain across various multiplexed VBR video sources thereby decreasing the

CLR and delay is applicable to all VBR video sources irrespective of the burstiness factor. Higher the burstiness, the effectiveness of the proposed scheme in increasing the utilization is more when compared to static and other dynamic bandwidth allocation schemes. This is due to the fact that a higher bandwidth reallocation is triggered only when the system is in the unsafe state. The traffic rates are continually tracked and adapted to, by the prediction system. However, the scheme should be evaluated with other VBR video sources and this is an issue for future research.

Another practical issue that arises is the loss of a cell that carries header information for decoding purposes at the destination end. The evaluation of the proposed scheme has been done based on the assumption that no header information is lost for a given CLR ($10^{-5}$). The cells containing the header information for the decoder at the destination end can be given higher priority when compared to that of the cells carrying compressed video data. Future work shall address this issue.


## Acknowledgments

The authors sicnerely thank the anonymous reviewer for the critical comments and M. W. Garrett of Bellcore and O. Rose, Univ. of Wurzburg for their video traces that were employed in this work.


## References


[1] A. Adas, *Supporting Real Time VBR Video Using Dynamic reservation Based on Linear Prediction*, Proc. IEEE INFOCOM, Vol 3, pp. 1476-1483, 1996.

[2] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, *Long-Range Dependence in Variable-Bit-Rate Video Traffic*, IEEE transactions on communications, Vol 43, pp. 1566-1579, 1995.

[3] P. D. Bertsekas and R. G. Gallager, *Data Networks* , Prentice-Hall, Englewood Cliffs, N.J., 1987.

[4] G. Chiruvolu, K. J. Christensen and R. Sankar, *Short-Term Resource Management for Better ATM QoS (extended abstract)* , Sixth International Workshop on Network Operating System Support for Digital Audio and Video, pp. 31-34, 1996.

[5] G. Chiruvolu and R. Sankar, *An Approach Towards Resource Management and Transportation of VBR Video Traffic*, Proc. IEEE ICC, pp. 550-554, 1997.

[6] S. Chong, San-qi Li, and J. Ghosh *Predictive Dynamic Bandwidth Allocation for Efficient Transport of Real-Time VBR Video over ATM*, IEEE JSAC, Vol 13, pp. 12-23, Jan. 1995.

[7] S. El-Henaoui, R. Coelho, S. Tohme, *Bandwidth allocation protocol for MPEG VBR traffic in ATM networks*, Proc IEEE Infocom, pp. 1100-07, 1996.

[8] A. Erramilli, O. Narayan and W. Willinger, *Experimental Queueing Analysis with Long-Range Dependent Packet Traffic*, IEEE/ACM Transactions On Networking, pp. 209-223, Apr. 1996.

[9] M. W. Garrett and W. Willinger, *Analysis, Modeling and Generation of Self-Similar VBR Video Traffic*, Proc. of ACM Sigcomm, pp. 269-280, Sep. 1994.

[10] M. Grossglauser and J. Bolot, *On the Relevance of Long-Range Dependence in Network Traffic*, Proc. of ACM Sigcomm, 1996.

[11] R. Guerin, H. Ahmadi, M. Nagshineh Equivalent Capacity and Its Application in High-Speed Networks, IEEE JSAC, Vol 9, pp. 968-981, Sep. 1991.

[12] D. P. Heyman, A. Tabatabai and T. V. Lakshman, *Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks*, IEEE Transactions on Circuits. Systems. Video Technology, pp. 43-49, Mar. 1992.

[13] D. P. Heyman and T. V. Lakshman *Source Models for VBR Broadcast-Video Traffic*, IEEE/ACM Transactions on Networking, Vol 4, pp. 40-48, Feb. 1996.

[14] D. P. Heyman and T. V. Lakshman *What Are The Implications of Long-Range Dependence For VBR-Video Traffic Engineering?*, IEEE/ACM Transactions on Networking, Vol 4, pp. 301-317, Jun. 1996.

[15] I. Hsu and J. Walrand, *Dynamic bandwidth allocation for ATM Switches*, Jour. Applied Probability, vol. 33, pp. 758-771, 1996.

[16] R. Jain, S. Kalyanaraman, R. Viswanathan, *The OSU Scheme for Congestion Avoidance Using Explicit Rate Indication*, ATM Forum, AF-TM 94-0883, Sep. 1994.

[17] H. Kanakia, P.P. Mishra and A. Reibman *Adaptive Congestion Control Scheme for Real-Time Packet Video Transport*, IEEE/ACM Transactions on Networking, Vol 3, pp. 671-682, Dec 1995.

[18] M. Linvy, B. Melamed, A. K. Tsiolis, *The Impact of Autocorrelation on Queueing Systems*, Management Science, vol. 39, pp. 332-339, 1993.

[19] P. P. Mishra and S. K. Tripathi, *Dynamic bandwidth allocation in high speed integrated service networks* , Proc. IEEE Infocom, pp. 1298-1305, 1993.

[20] K. Ma, R. Sankar, and K. Christensen, A New Explicit Rate-Based Congestion Control Scheme for ABR Services, Proc. 22nd IEEE Annual Conference on Local Computer Networks, pp. 195-201, 1997.

[21] P. Pancha and M. E. Zarki *Bandwidth Allocation Schemes for Variable-Bit-Rate MPEG Sources in ATM Networks*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 3, pp. 190-198, Jun. 1993.

[22] B. E. Patuwo, R. L. Disney, D. C. Mcnickle, *The Effect of Correlated Arrivals on Queues*, IIE Transactions, Vol. 25, pp. 105-110, May 1993.

[23] S. Rampal, D.S Reeves and I. Viniotis, *Dynamic Resource allocation based on measures QoS* , Technical report TR 96-2, NCSU, 1996.

[24] N. Ranganathan, R. Anand and G. Chiruvolu, *A VLSI ATM Switch Architecture for VBR traffic* , Proc. IEEE International Conference on VLSI design, 1998.

[25] O. Rose, *Statistical Properties of MPEG video traffic and their impact on traffic modeling in ATM systems*, Technical Report No. 101, Univ. Of Wurzburg, Institute of computer science, 1995.

[26] O. Rose and M. R. Frater, *Impact of MPEG Video Traffic on an ATM Multiplexer* , High Performance Networking VI, Chapman and Hall, pp. 157-168, 1995.

[27] B. K. Ryu and A. Elwalid, *The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities*, Proc. of ACM Sigcomm, 1996.

[28] P. Sen, B. Maglaris, N.E. Rikkli and D. Anastassiou, *Models for Packet-Switching of Variable-Bit-Rate Video Sources*, IEEE JSAC, Vol 7, pp. 761-770, 1989.

[29] P. Tin, *A Queueing System with Markov Dependent arrivals*, Jour. Applied Probability, vol. 22, pp. 668-677, 1985.