A Fast Non-linear Adaptive Algorithm for Video Traffic Prediction

Hong Zhao, Nirwan Ansari and Yun Q. Shi
Advanced Networking laboratory
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
University Heights
Newark, NJ, 07102, USA
Email: nirwan.ansari@njit.edu

Abstract†

To guarantee quality of service (QoS), the requirements for video transmission such as delay and cell loss rate (CLR) are very stringent. These constraints are difficult to meet if high network utilization is desired. Dynamic bandwidth allocation is thus needed. Video traffic prediction can play an important role in dynamic bandwidth allocation and traffic management in high-speed networks. In this paper, we propose a fast convergent algorithm to predict the variation of I frames, based on which the bandwidth is assigned. The new algorithm can achieve fast convergence and small prediction error.

1. Introduction

Variable Bit Rate (VBR) video is one of the major applications to be supported by broadband packet switched networks. Video is inherently dynamic, and MPEG video coding results in VBR. If the bandwidth is allocated according to the peak rate of the video traffic, no packet loss occurs, but a substantial amount of the bandwidth is wasted during most of the transmission. On the other hand, if the bandwidth is not allocated close to the peak rate, large delays and excessive packet loss may be experienced. In transporting the VBR video traffic, achieving effective use of the network resource while providing QoS guarantees is not trivial due to the bursty characteristics of VBR traffic. Researchers have found the existence of correlation in video trace generated from an MPEG encoder; this phenomenon can be used for traffic prediction. The prediction, when combined with dynamic bandwidth allocation, can provision both network efficiency and QoS guarantees. Earlier work in this area includes frequency-domain prediction and time-domain prediction.

Chong *et al.* [1] approached the problem in the frequency domain. They proposed a method to dynamically allocate the bandwidth based on predicting the low frequency part of the video rate input sequence. The low frequency part of the signal represents the slow time variations of the VBR rate and it is used to determine the allocated bandwidth.

Wang *et al.* [2] proposed an adaptive wavelet prediction method for VBR video traffic. This method uses the wavelet transform to transform the video sequence into the wavelet domain. Though it can improve the prediction performance, the computational complexity is rather high.

A adaptive linear prediction scheme was proposed by Adas [3]. This scheme does not require any prior knowledge of the video statistics nor does it assume stationary, and is thus suitable for on-line real time prediction. However, when there are scene changes, the bit rate variation is so high that the prediction error can be large. Xu and Qureshi [4] proposed a composite MPEG traffic prediction scheme which smoothes the predicted data based on predicting relative changes of frame sizes between adjacent GOPs. Since I, P and B possess different statistical characteristic, this method is not effective in guaranteeing the cell loss rate (CLR) and needs renegotiating for every frame, a big burden to network management.

Owing to the above drawbacks, we propose a dynamic bandwidth allocation algorithm based on the predicted relative size change of I frames. This not only smoothes the predicted data but also reduces the renegotiation frequency, and the prediction error is much smaller than the composite MPEG traffic prediction scheme, but one problem associated with this LMS algorithm is its slow convergence. In VBR video traffic characterized by frequent scene changes, the LMS algorithm may result in an extended period of intractability and thus experience excessive cell loss during scene changes. In this paper, we propose a fast

† This work has been supported in part by the New Jersey Commission on Science and Technology via the New Jersey Center for Wireless Telecommunications and the New Jersey Commission on Higher Education via the NJI-TOWER project.

0-7695-1506-1/02 $17.00 © 2002 IEEE

convergent nonlinear adaptive algorithm to predict the variation of I frames. This new algorithm converges faster and hence, tracks scene changes better. The rest of the paper is organized as follows. In Section 2, characteristics of the MPEG video traffic are examined. In Section 3, the fast convergent nonlinear adaptive algorithm is proposed. Section 4 presents the performance of our proposed scheme. Finally, concluding remarks are drawn in Section 5.

2. Characteristics of MPEG Videos

An MPEG encoder that compresses a video signal at a constant picture rate (e.g., 30 pictures/s) produces a coded bit stream with a highly variable bit rate, thus called Variable Bit Rate (VBR). The changes in the output rate of an MPEG encoder are attributed to the following three aspects:

1. The encoding of one block to the next within a picture.
2. From one picture to the next within the video sequence being encoded.
3. From one scene to the next within the video sequence.

The rate fluctuations from one picture to the next are the most troublesome for network management. If we can predict the frame size more accurately, network utilization can be improved and QoS can be guaranteed.

An MPEG video stream is divided into units called group of pictures (GOP). A GOP consists of an I frame and an arrangement of B and P frames. Video traffic is correlated and its autocorrelation has a heavy tail, because MPEG uses intra-frame techniques (exploiting the spatial redundancy within a picture) as well as inter-frame techniques (exploiting the temporal redundancy present in a video sequence). A highly correlated input process with a heavy tail, if served at a fixed rate not close to the peak rate, causes large queues, large delays and excessive cell loss [5].

The frame size trace from the output of the MPEG encoder contains all statistical information about the encoded video. The frame by frame correlation depends on the patterns of the GOP and in principle always looks like Figure 1 if the same GOP pattern is used for the whole sequence. For this example, the GOP pattern is: IBBPBBPBBPBBI…

Figure 1 shows the ACF of the MPEG coded *star war*. In Figure 1, the large positive peaks stem from I frames, the smaller positive ones from P frames, and the negative ones from B frames. A large I frame is followed by two small B frames, then a middle size P frame is followed by two small B frames again. The pattern between two I frame peaks is repeated with slowly decaying amplitude of the peaks.

From this figure, we can see that the MPEG video is highly correlated. If it is not served at a rate close to the peak rate, large queues, large delays and excessive cell loss will result, but if we reserve a bandwidth at least equal to the predicted value, we only need to buffer the error caused by the prediction. If the error resembles white noise or at most short memory, only small buffers will suffice, and high utilization and small delays can be achieved.

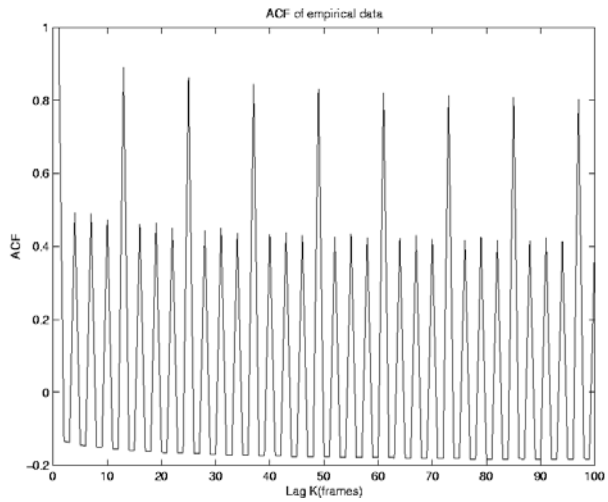

Fig. 1. ACF of MPEG video

3. The Fast Algorithm

Thorough analysis of MPEG video traces indicates that, within a GOP, the I frame is often the largest, and B frames are the smallest. Most of the time, when the size of the I frame changes significantly, so do those of P and B frames, implying that the increase or decrease of I frame size often indicates the increase or decease of P and B frame sizes, respectively, and therefore bandwidth should be allocated based on the size of I frames. Thus, the primary goal of this paper is to accurately and promptly predict the I frame size. Let $I(k)$ be the size of the I frame of the kth GOP and $I(k\text{-}1)$ be the size of the (k-1)th GOP, then the relative size change of I frame, denoted by $s(k)$, is defined as:

$$s(k) = \frac{I(k) - I(k-1)}{I(k-1)} \quad . \qquad (1)$$

The sequence $s(k)$ is much smoother than the sequence $I(k)$. Thus the linear adaptive prediction will perform better if we predict the sequence $s(k)$ instead of the sequence $I(k)$; the I frame size can then be retrieved by

$$I(k) = s(k)I(k-1) + I(k-1) \ . \qquad (2)$$

A one-step linear predictor can be used to predict the *s(k)* sequence, i.e., prediction of *s(k+1)* using a linear combination of the current and previous values of *s(k*). The number of the current and previous values of *s(k)* used to predict *s(k+1)* is called the order of the linear predictor. The *p*th-order linear predictor has the following form:

$$\hat{s}(k+1) = \sum_{l=0}^{p-1} w(l)s(k-l) = W^T S(k), \qquad (3)$$

where p is the order of the linear predictor, and $w(l)$, for $l = 0,1,\ldots, p-1$, are the prediction filter coefficients. The prediction error is

$$e(k) = s(k+1) - \hat{s}(k+1) \cdot \qquad (4)$$

The LMS predictor minimizes the mean square error by adaptively adjusting the coefficient vector *W*. In normalized LMS algorithm [3], if we use the one-step linear predictor, *W* is updated by

$$W(k+1) = W(k) + \frac{\mu e(k) s(k)}{\|s(k)\|^2} \ . \qquad (5)$$

Since at time *k* the value of *s(k*+1*)* is not available to compute *e(k), e(k*-1) is used instead. In the standard LMS algorithm [6], μ is a constant; we refer to this algorithm as the fixed step size algorithm (FSA). Since video traffic is bursty, if we increase the step size μ , we can achieve fast convergence at the cost of a larger prediction error. On the other hand, the prediction error can be made small by decreasing the step size μ at the cost of the convergence rate. The choice of the step size reflects a trade off between misadjustment and the speed of adaptation. The slow convergence of LMS may cause an extended period of intractability and excessive cell loss during scene changes.

Kwong and Johnston [7] proposed a variable step size algorithm for adjusting the step size μ_k :

$$\mu'_{k+1} = \alpha\mu_k + \gamma\varepsilon_k^2 , \qquad (6)$$

with $0 < \alpha < 1, \qquad \gamma > 0$,

and
$$\mu_{k+1} = \begin{cases} \mu_{\max} & \text{if } \mu'_{k+1} > \mu_{\max} \\ \mu_{\min} & \text{if } \mu'_{k+1} < \mu_{\min} \\ \mu'_{k+1} & \text{otherwise.} \end{cases} \qquad (7)$$

$$W_{k+1} = W_k + \mu_k e_k S_k ,$$

where $0 < \mu_{\min} < \mu_{\max}$. The initial step size μ_0 is usually taken to be a little larger, although the algorithm is not sensitive to the choice. As can be seen from (6), the step size is always positive and is controlled by the size of the prediction error and the parameters α and γ. Intuitively, a large prediction error increases the step size to provide faster tracking. If the prediction error decreases, the step size will be decreased to reduce the misadjustment. The constant $\mu_{\max}$ is chosen to ensure that the mean-square error (MSE) of the algorithm remains bounded. Usually, $\mu_{\min}$ is chosen to be close to the value that has been chosen for the fixed step size algorithm. We propose to modify Equation (6) to the following:

$$\mu'_{k+1} = \alpha\mu_k + \gamma(q_1 e_k^2 + q_2 e_{k-1}^2) , \qquad (8)$$

to accommodate the video traffic characteristics. We refer to this algorithm as the fast convergent variable step size algorithm (VSA) in this paper. Here, e_k and e_{k-1} are the current and previous prediction errors, respectively, and q_1 and q_2 are their respective weights. From numerous simulations, we found that α=0.98 and γ=0.015 work well in our real video trace simulations. Thus, we empirically set α=0.98 and γ=0.015 .

The Akaike information criterion (AIC) [8] is used to chose the best order not greater than 12. The AIC criterion associates a cost function with the order of the filter. It was found by numerous simulations that the autocorrelation function of the prediction error *e(n)* is close to that of white noise with *p*=12 [3]. We have observed the prediction error of video traffic is a rather "uncorrelated" process resembling white noise. Thus, we use one step, 12-order adaptive linear predictor for both VSA and FSA.

4. Simulation Results

Simulations on four half-hour long empirical VBR video traffic data sets were conducted. These data sets correspond to frame-size traces. Since frame size traces from the output of the MPEG encoder contain all statistical information about the encoded video, we can reserve bandwidth at least equal to the predicted value, and thus only the prediction error needs to be buffered. If the error resembles white noise or at most short memory, only small buffers are needed and high utilization and small delays can be achieved.

For performance comparison between VSA and FSA, we use $SNR^{-1} = \Sigma e^2(n) / \Sigma s^2(n)$ as a metric, which is the ratio of the sum of squares of prediction error and the sum of squares of input data. For a fair comparison, both FSA and VSA use the same 12-order and one step ahead prediction. The results are shown in Table 1.

From Table 1, VSA incurs smaller prediction error than FSA in all the four tested sequences. Figure 2 shows the convergence properties of FSA ($\mu = 0.009$) and VSA. Note that VSA converges much faster than FSA. If we increase the step size to $\mu = 0.3$ for the FSA, the convergence is fasten as shown in Figure 3 (note that the MSE is expressed in dB), but the prediction error is increased greatly; in this case, $\sum e^2(n)/\sum s^2(n)$=0.0191 for FSA, $\sum e^2(n)/\sum s^2(n)$=0.0032 for VSA.

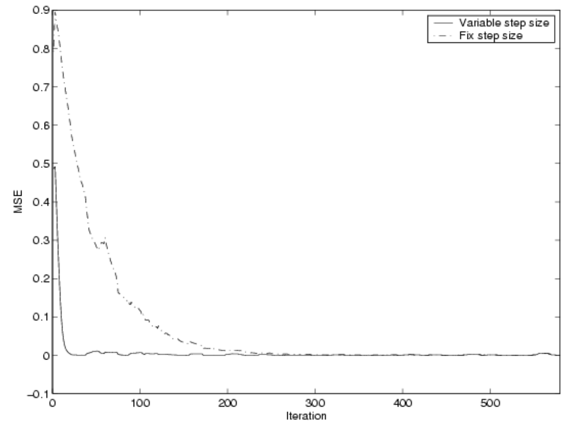

Fig. 2. Comparison of convergence properties of FSA ($\mu = 0.009$) and VSA on CD122 trace

5. Conclusions

We have proposed a variable step size predictor for VBR video traffic, where the step size adjustment is controlled by the squares of the prediction error to reduce the trade off between misadjustment and tracking ability of the fixed step size LMS algorithm. Our simulations show that VSA not only incurs small prediction errors but more importantly also achieves faster convergence. Video traffic prediction can play an important role in dynamic bandwidth allocation. When employed for dynamic bandwidth allocation, VSA can significantly reduce CLR. This scheme does not require any prior knowledge of the video statistics nor does it assume stationary, and is thus very suitable for on-line real time prediction. It can also track scene changes better than FSA. This algorithm can be used for on-line real-time video transmission.

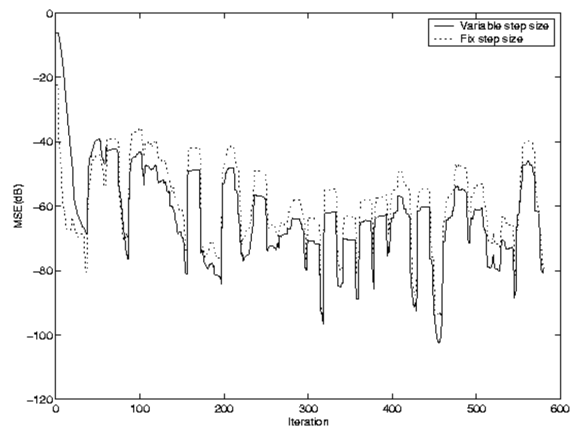

Fig. 3. Comparison of MSE (dB) of VSA and FSA ($\mu = 0.3$) on CD122 trace

Acknowledgements

The authors would like to thank Mark Garrett of Telecordia and Wu-Chi Feng of the Ohio State University for the traffic traces provided by them.

Table 1. Performance comparison of FSA and VSA predictors on relative size changes of I frames

(Use $\sum e^2(n)/\sum s^2(n)$ as a metric)

Sequence	FSA	VSA	Improvement (%)
CD122	0.0040	0.0032	20
Talk2	0.0078	0.0069	12
News	0.0247	0.0210	15
SoccerWM	0.0512	0.0404	21

References

[1] S. Chong, S. Li, and J. Ghosh, "Efficient Transport of Real Time VBR Video over ATM via Dynamic Bandwidth Allocation," *IEEE J. Select. Areas Communication*, vol. 13, Jan., 1995, pp. 12-23.

[2] X. Wang, S. Jung and J. Meditch, "Dynamic Bandwidth Allocation for VBR Video Traffic Using Adaptive Wavelet Prediction," *Proc. IEEE Intl. Conf. on Communication,* 1998, vol. 1, pp. 549-553.

[3] A. Adas, "Using Adaptive Linear Prediction to Support Real-Time VBR Video Under RCBR Network Service Model," *IEEE/ACM Trans. on Networking,* vol. 6, NO.5, Oct. 1998, pp. 635-644.

[4] W. Xu and A. G. Qureshi, "Adaptive Linear Prediction of MPEG Video Traffic," *5th International Symposium on Signal Processing and it's Application,* 1999, pp. 67-70.

[5] M. Hayes, *Statistical Digital Signal Processing and Modeling.* New York: Wiley, 1996.
[6] M. Livny, B. Melamed and A. K. Tsiolis, “The Impact of Autocorrelation on Queuing System,” *Manage. Sci*., Mar. 1993, pp. 329-339.
[7] R.H. Kwong and E.W. Johnston, “A Variable Step Size LMS Algorithm,” *IEEE Trans. Signal Processing,* vol. 40, Jul. 1992, pp. 1633-1642.
[8] A. Adas, “On Resource Management and QoS Guarantees for Long Range Dependent Traffic,” *Proc. IEEE INFOCOM*, Apr. 1995, pp. 779 –787.
[9] O. Rose, “Statistical Properties of MPEG Video Traffic and their Impact on Traffic Modeling in ATM System,” *Proc. 20th Conf. on Local Computer Networks*, 1995, pp. 397-406.
[10] O. Rose, “Queue Response to Input Correlation Functions: Discrete Spectral Analysis,” *IEEE/ACM Trans. on Networking,* vol.1, Oct. 1993, pp. 522-533.
[11] H. Liu, N. Ansari and Y.Q. Shi, “Dynamic Bandwidth Allocation for VBR Video Transmission,” *Proc. IEEE Intl. Conf. on Information Technology, Coding and Computing,* 2001, pp. 284-288.