PAPER

Efficient Predictive Bandwidth Allocation for Real Time Videos

Hong ZHAO†, *Nonmember*, **Nirwan ANSARI**†, *Regular Member*, *and* **Yun Q. SHI**†, *Nonmember*

SUMMARY The Quality of Service (QoS) requirements such as delay and cell loss ratio (CLR) are very stringent for video transmission. These constraints are difficult to meet if high network utilization is desired. Dynamic bandwidth allocation in which video traffic prediction can play an important role is thus needed. In this paper, we suggest to predict the variation of I frames instead of the actual size of I frames, and propose an algorithm that can achieve fast convergence and small prediction error, thus imposing QoS and attaining high network utilization.

The performance of the scheme is studied using the renegotiated constant bit rate (RCBR) service model. The overall dynamic bandwidth allocation scheme based on our fast convergent algorithm is shown to be promising, and practically feasible for efficient transmission of real time videos.

key words: *Bandwidth prediction, QoS, LMS, VBR*

1. Introduction

Variable Bit Rate (VBR) is one of the major services to be supported by broadband packet switched networks. Video is inherently dynamic, and MPEG [1] 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 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. So in transporting the VBR video traffic, QoS guarantees provisioning is not trivial due to the bursty characteristics of the VBR traffic. However, existence of correlation in the video trace generated from an MPEG encoder can be exploited for traffic prediction, which, when combined with dynamic bandwidth allocation, can provision both network efficiency and QoS guarantees. Earlier work in this area includes the frequency domain and time domain prediction approach.

Chong et al. [2] 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 variation of the VBR rate and is used to determine the allocated bandwidth.

Wang et al. [3] predicted the VBR video traffic by transforming the video sequence into the wavelet domain. Though it can improve the prediction performance, the computational complexity is rather high.

An adaptive linear prediction scheme was proposed by Adas [4]. 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. However, when there are scene changes, the bit rate variation is so high that the prediction error can be large. Xu and Qureshi [5] proposed a composite MPEG traffic prediction scheme which smoothes the predicted data based on predicting relative changes of frame size between adjacent GOPs. Since I, P and B possess different statistical characteristics, this method is not effective in guaranteeing the CLR and needs renegotiations 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. The proposed algorithm not only smoothes the predicted data, reduces the renegotiation frequency, but also achieves much smaller prediction error than that by the composite MPEG traffic prediction scheme; one problem associated with this least mean squares (LMS) [6] algorithm is its slow convergence. In VBR video traffic characterized by the frequent scene changes, the LMS algorithm may result in an extended period of intractability, and thus experience excessive cell loss during scene changes. Thus, we also propose a fast convergent nonlinear adaptive algorithm to predict the relative size change of I frames. This new algorithm converges faster, and hence tracks scene changes better.

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 VBR. 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

Manuscript received March 15, 2002.

Manuscript revised July 19, 2002.

†The authors are with Center for Communications and Signal Processing Research, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA.

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 fluctuation from one picture to the next is the most troublesome for the network management. If the frame size can be predicted more accurately, network utilization can be improved and QoS can be guaranteed.

An MPEG video 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 [4].

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

Figure 1 shows the autocorrelation function (ACF) of the MPEG coded *Star Wars* (lag is expressed in terms of the frame number), in which the largest positive peaks stem from I frames, the larger positive ones from P frames, and the smallest 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.

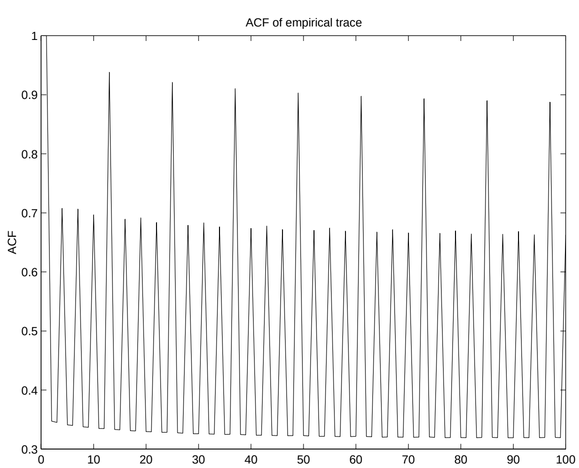

Fig. 1 ACF of an MPEG video.

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.

3. Predicting the relative size change of the I frame size

Through the analysis of the MPEG video trace we find that I frames often have large frame sizes, and B frames have small frame sizes. Most of the time, when the I frame size changes significantly, the P and B frame size also change significantly, implying that the increase or decrease of the I frame size often indicates the increase or decrease of the P and B frame sizes, and therefore we only need to predict the I frame size. Let I_k be the size of the I frame of the kth GOP and I_{k-1} be the size of the $(k-1)$th GOP, then the relative size change of I frame s_k is defined by

$$s_k = \frac{I_k - I_{k-1}}{I_{k-1}}. \tag{1}$$

The sequence s_k is much smoother than the sequence I_k. So 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

$$I_k = s_k I_{k-1} + I_{k-1}. \tag{2}$$

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 pth-order linear predictor has the following form:

$$\hat{s}_{k+1} = \sum_{l=0}^{p-1} w_l s_{k-l} = \mathbf{W}^T \mathbf{S}_k, \tag{3}$$

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

$$e_k = s_k - \hat{s}_k. \tag{4}$$

The LMS predictor minimizes the mean squares error by adaptively adjusting the coefficient vector $\mathbf{W}$. In normalized LMS algorithm [4], if we use the one-step linear predictor, $\mathbf{W}$ is updated by

$$\mathbf{W_{k+1}} = \mathbf{W_k} + \frac{\mu \mathbf{e_k S_k}}{\|\mathbf{S_k}\|^2}. \tag{5}$$

The Akaike information criterion (AIC) [7] is used to choose 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 of the prediction error e_k is close to that of the white noise. Thus we use one-step, 12-order adaptive linear predictor for both our algorithm and the composite prediction scheme [5].

The performance of our algorithm for the video trace CD122, justified by the inverse Signal to Noise Ratio (SNR^{-1}) is

$$SNR^{-1} = \frac{\sum e^2(n)}{\sum s^2(n)} = 0.0040.$$

Figure 2 shows the forecasted values appear close to the actual values except at sharp transition, which are most likely due to scene changes.

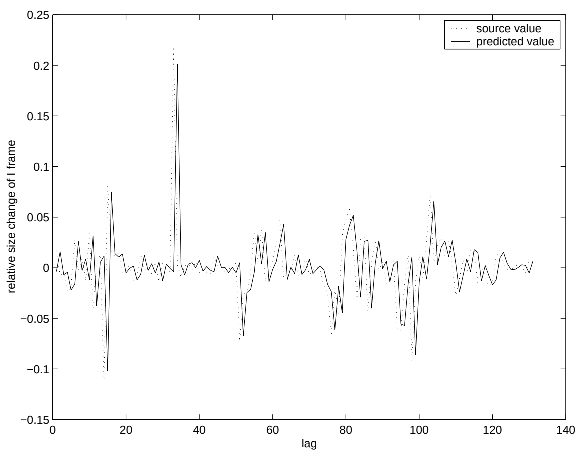

Fig. 2 Actual and forecasted I frame size variation for CD122; in this case, the lag implicitly represents the I-frame number.

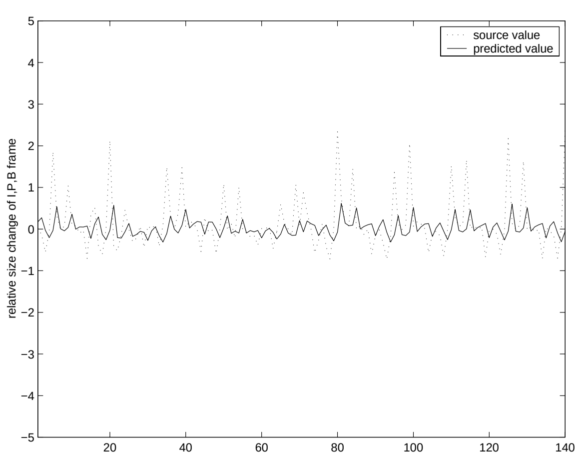

Fig. 3 Actual and forecasted frame size variation (I,P,B) for CD122.

Figure 3 shows the actual and forecasted frame size variation proposed by Xu and Quresh [5], which predicts the relative size change of every frame between two adjacent GOPs. The sequence is defined as follows:

$$f_k = \frac{F_k - F_{k-d}}{F_{k-d}}, \tag{6}$$

where F_k is the size of the kth frame in the MPEG encoded video sequence, d is the length of the GOP, and f_k is the relative change of frame sizes between adjacent GOPs. Since I, B, P frames are coded with different compression levels, they have different statistical properties. f_k is more fluctuated than s_k, and so the prediction error of f_k shown in Fig. 3 is larger than that of s_k shown in Fig. 2.

The performance of the composite prediction scheme for CD122 is

$$SNR^{-1} = \frac{\sum e^2(n)}{\sum s^2(n)} = 0.8471.$$

From Figs. 2 and 3 and their SNR^{-1}, our proposed scheme is more accurate than the composite prediction scheme [5]. On the other hand, it is not possible to allocate bandwidth based on every frame size, because the negotiation frequency is very high, thus imposing a big burden to the network. Hence, we should not predict every frame size.

4. The fast algorithm

The algorithm proposed in Section 3 not only smoothes the predicted data, reduces the renegotiation frequency, but also achieves much smaller prediction error than that of the composite MPEG traffic prediction scheme. The only drawback 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 may experience excessive cell loss during scene changes; hence, we propose a fast convergent nonlinear adaptive algorithm to predict the relative size changes of I frames. This new algorithm converges fast, and hence, tracks scene changes better.

In the standard LMS algorithm (3) $\sim$ (5), μ is a constant; we refer to this algorithm as the fixed step size algorithm (FSA). Since the video traffic is bursty, if we increase the step size μ we can achieve fast convergence at the cost of a large 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 the trade off between the misadjustment and the speed adaption.

Kwong and Johnston [8] proposed a variable step size algorithm for adjusting the step size μ_k:

$$\mu'_{k+1} = \alpha\mu_k + \gamma e_k^2, \tag{7}$$

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

$$\mu_{k+1} = \begin{cases} \mu_{max} & \text{if } \mu'_{k+1} > \mu_{max} \\ \mu_{min} & \text{if } \mu'_{k+1} < \mu_{min} \\ \mu'_{k+1} & \text{otherwise.} \end{cases} \quad (8)$$

The initial step size μ_0 is usually taken to be a little large, although the algorithm is not sensitive to the choice. As can be seen from Eq. (7), 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 μ_{max} is chosen to ensure that the mean-square error (MSE) of the algorithm remains bounded. Usually, μ_{min} is chosen to be close to the value that has been chosen for the fixed step size algorithm. We propose to modify Eq. (7) to the following:

$$\mu'_{k+1} = \alpha\mu_k + \gamma(q_1 e_k^2 + q_2 e_{k-1}^2), \quad (9)$$

where $q_1 + q_2 = 1$ to accommodate the video traffic characteristics. Since a video frame sequence consists of many scenes, the bit rate varies greatly among different scenes, while during a scene, the bit rate in the frame sequence has a strong auto-correlation, and thus better prediction performance is expected. Equation (9) includes an additional error term e_{k-1}, to smooth out the drastic change of μ_k (i.e., a spike) during the transition from one scene to another as shown in Fig. 4, thus easing the buffer management. We refer to this algorithm as the fast convergent variable step size algorithm (VSA), and Kong and Johnston's method as KVSA 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. We empirically derived from the video trace "Talk2" the parameters $\alpha = 0.98$, $\gamma = 0.015$, $q_1 = 0.7$ and $q_2 = 0.3$, and found that these parameters provide the best performance in all our real video trace simulations.

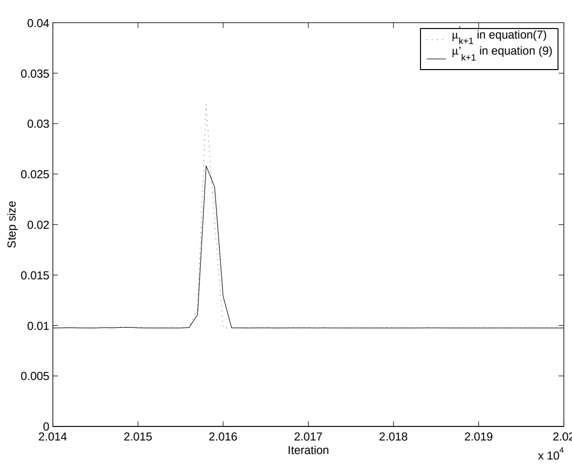

Fig. 4 Comparison of stepsize.

Simulations on four 1.5-hour long empirical VBR traffic data sets were conducted. These data sets correspond to the relative size change of I frames. For performance comparison among VSA, KVSA and FSA, we use $SNR^{-1} = \frac{\sum e^2(n)}{\sum s^2(n)}$ as a metric, which is the ratio of the sum of squares of prediction error and sum of squares of input data. For a fair comparison, VSA, KVSA and FSA use the same 12-order and one-step ahead prediction, and parameters α and γ are the same in both VSA and KVSA. The results are shown in Table 1.

Table 1 Performance comparison of FSA, KVSA, and VSA predictors on relative size change of I frames ($\frac{\sum e^2(n)}{\sum s^2(n)}$ is used as a metric)

Sequence	FSA	KVSA	VSA	Improvement(%)
CD122	0.0040	0.0035	0.0032	20
Talk2	0.0078	0.0071	0.0069	12
News	0.0247	0.0213	0.0210	15
SoccerWM	0.0512	0.0438	0.0404	21

From Table 1, VSA and KVSA incur smaller prediction error than FSA in all the four tested sequences. VSA further reduces the prediction errors as shown in Table 1. The performance has been improved greatly if we use VSA instead of FSA as shown in Table 1. The percentage improvement is respect to VSA over FSA. Figure 5 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 FSA, the convergence is faster as shown in Fig. 6 (note that the MSE is expressed in dB), but the prediction error is increased greatly; here the iteration represents the iteration index, in this case, $\frac{\sum e^2(n)}{\sum s^2(n)} = 0.0191$ for FSA, $\frac{\sum e^2(n)}{\sum s^2(n)} = 0.0032$ for VSA.

5. Dynamic bandwidth allocation

There are many ways to use the prediction for allocating bandwidth for the VBR video traffic. The bandwidth allocation mechanism can be activated by each prediction, or only those predictions whose results are larger than a certain threshold. The variation reflects the trade-off between the network utilization and the overhead for bandwidth negotiation. The choice for a network application depends on many factors, such as the network service model, latency, and implementation complexity. Since the increase or decrease of I frames often indicates the increase or decrease of P and B frames, we propose to allocate bandwidth based on the predicted I frame size using VSA to improve the QoS and network utilization.

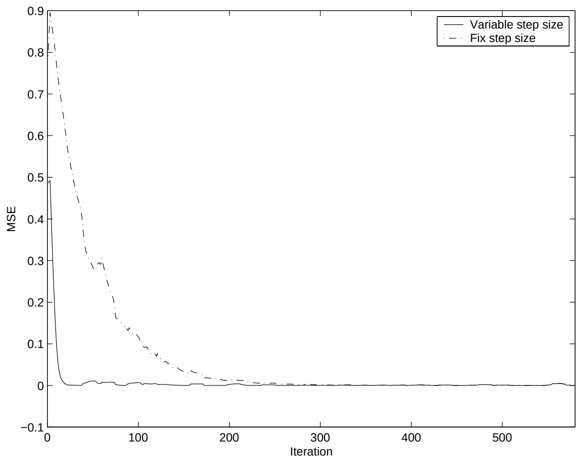

Fig. 5 Comparison of convergence properties of FSA ($\mu = 0.009$) and VSA on CD122 trace.

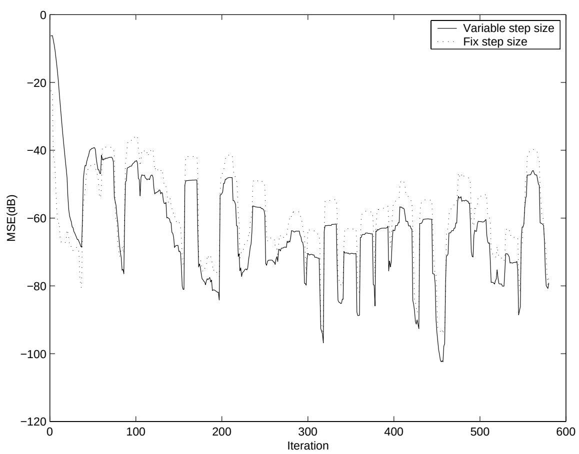

Fig. 6 Comparison of MSE (dB) of VSA and FSA ($\mu = 0.3$) on CD122 trace.

5.1 Dynamic Bandwidth Allocation Based on Predicted I Frames Using VSA

A single server FIFO queue is simulated. The assumed network service model is RCBR [9]. Let I_k be the size of the predicted I frame of the kth GOP, R be the transmission rate for the previous GOP and δ be a threshold, then the dynamic bandwidth allocation algorithm can be stated as follows:

- if $|I_k - R/N| < \delta$, then the transmission rate remains unchanged.
- if $|I_k - R/N| \geq \delta$, then $R = I_k \times N$,

where N is the number of frames needs to be transmitted per second.

The negotiation frequency can be reduced significantly because only I frames need to be checked. Since the I frame size in a GOP is the largest most of the time, the bandwidth allocated is very close to the largest one needed for transmission of frames in the GOP, and therefore CLR can be kept small. The CLRs for different values of δ for the sequence CD122 are shown in Fig. 7.

Figure 8 shows the performance of the video trace "Talk2" show using our prediction algorithm, where $\delta = 1000$, and the buffer size stands for the size of the buffer needed at the switch. The prediction performance is

$$SNR^{-1} = 0.0069.$$

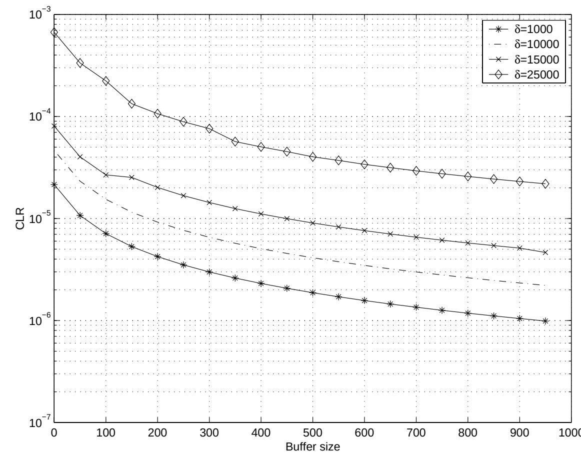

Fig. 7 CLR versus buffer size for different values of δ.

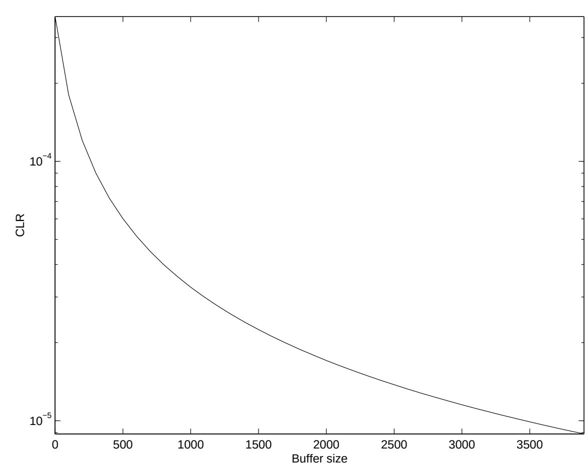

Fig. 8 CLR versus buffer size for the Talk2 show.

5.2 Impact of Autocorrelation to Queue Size

By reserving the bandwidth at least equal to the predicted values, only prediction errors need to be buffered. The autocorrelation of the prediction errors of the relative size change of I frames for video sequences

CD122 and Talk2 are shown in Figs. 9 and 10, respectively. Note that they resemble white noise, a rather uncorrelated process. Since traffic autocorrelation has great impact on the queuing performance, we need to understand how the queue responds to different autocorrelations.

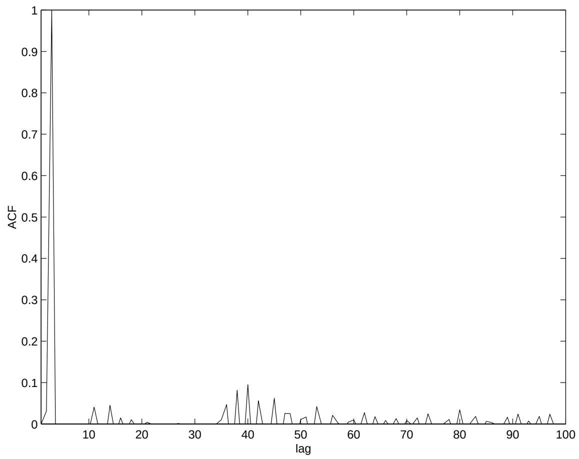

Fig. 9 Autocorrelation of the prediction error for CD122.

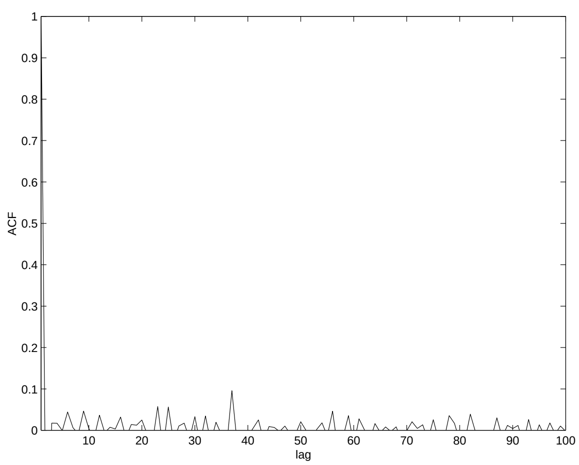

Fig. 10 Autocorrelation of the prediction error for Talk2.

Li and Hwang [10] analyzed this impact in the frequency domain. Consider a Markov modulated Poisson process. This process is described by an N state discrete time MC transition matrix P which is diagonalizable, and its associated input rate vector $\vec{\gamma}$. The input rate correlation is expressed as follows:

$$R(n) = \sum_{\lambda_l \in \Omega_r} \psi_l \lambda_l^{|n|} + \sum_{\lambda_l \in \Omega_C Im(\lambda_l) \geq 0} 2|\psi_l||\lambda_l|^{|n|} \cos(|n|\omega_l T + \theta_l T). \tag{10}$$

where λ_l is the l^{th} eigenvalue of P, ψ_l is associated with both the l^{th} eigenvector of P and the input rate $\vec{\gamma}$, $\omega_l = arg(\lambda_l)$, T is the time unit, Ω_r is the subset of the real eigenvalues of P, Ω_c is the subset of complex eigenvalues of P, and $Im(x)$ is the imaginary part of x.

The power spectral function of the input rate process is defined by the discrete-time Fourier transform of $R(n)$: $P(\omega) = \sum_{n=-\infty}^{\infty} R(n) e^{-\sqrt{-1} n\omega T}$, which is equal to

$$P(\omega) = \sum_l \frac{\psi_l (1 - \lambda_l^2)}{1 - 2\lambda_l \cos(\omega T) + \lambda_l^2} . \tag{11}$$

From Eqs. (10) and (11), essential characteristics of the input correlation (or its power spectrum) are captured by the input MC eigenvalues.

Li and Hong [10] analyzed the queue response to the input correlation or the power spectrum by an N-state periodic chain to match the input power spectrum. While in state i, the packet arrivals are characterized by a Poisson process with input rate γ_i. The model discussed here is a typical two dimensional MC in the discrete-time domain. The mean queue size is expressed as follows:

$$E[q] = \frac{\overline{\gamma}^2}{2(1-\gamma)} + \frac{1}{1-\bar{\gamma}} \sum_{j=0}^{N-1} \sum_{l=0}^{N-1} c_l \gamma(j,l), \tag{12}$$

with

$$\gamma(j,l) = \sum_{k=0}^{j} (\gamma_k - \bar{\gamma}) - \sum_{k=0}^{l} (\gamma_k - \bar{\gamma}), \tag{13}$$

$$\overline{\gamma} = \frac{1}{N} \sum_l \gamma_l. \tag{14}$$

c_l is the l^{th} element of the boundary vector $\vec{c}$ [11]. The first item in (12) is also equal to the mean queue size of the $M/D/1$ model, which in our case is equivalent to the queue response to the white noise input. The second moment of the queue is given by

$$E[q^2] = \frac{\bar{\gamma}^4 - \bar{\gamma}^3 + 3\bar{\gamma}^2}{6(1-\bar{\gamma})^2} + \frac{1 - \bar{\gamma} + \bar{\gamma}^2}{(1-\bar{\gamma})^2} \sum_{j=0}^{N-1} \sum_{l=0}^{N-1} c_l \gamma(j,l) + \frac{1}{1-\bar{\gamma}} c_l \gamma^2(j,l). \tag{15}$$

By examining the queue response to various input correlation properties on the basis of the input power spectrum in the discrete frequency domain, we conclude that the queuing behavior is dominated by input power in the low frequency band, and many high-frequency components existing in the input process can be replaced by a constant input rate, that have little impact on the queue response [10]. Thus, we can neglect the high frequency power in the power spectrum. This is true especially for multimedia traffic queuing analysis, where the input process often contains the dominant low frequency power as shown in Fig. 11. The larger the autocorrelation, the more input power is in the low

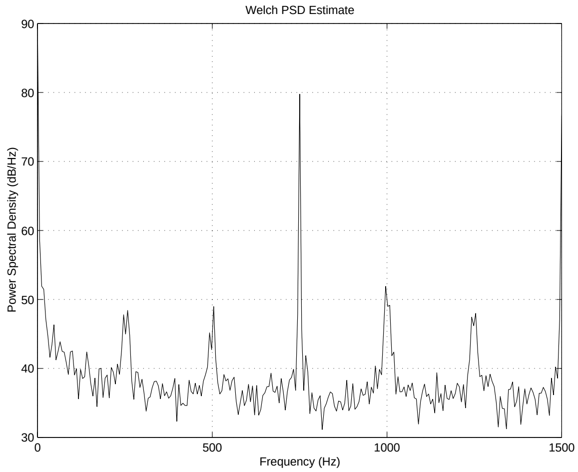

Fig. 11 The power spectrum of the Talk2 show.

frequency band and the longer the mean queue size is.

From the above analysis, we observe that the queuing performance is mainly dominated by input streams with high positive correlation. The higher the positive correlation in the time domain, the more the input power is in the low frequency band, and so the mean queue size will be longer. From the network perspective, the key point is to assign the link capacity to a given multimedia traffic in order to provide guaranteed quality of services. Video traffic is highly correlated, as shown in Fig. 1. The power spectrum as shown in Fig. 11, has spikes appeared at harmonic frequencies due to the strong autocorrelation of periodic I frames. More input power is located in the low frequency band, thus resulting in large queue size, and hence we cannot guarantee the quality of service. A large queue will introduce a long delay that cannot be tolerated for video delivery especially for real time videos. Here, we propose a bandwidth allocation scheme based on the predicted I frame size (retrieved from s_k) in which only prediction errors need to be buffered. From Figs. 9 and 10, we note that the prediction error resembles white noise, or at most short memory, a rather "uncorrelated" process resembling white noise; this has little impact on the queuing dynamics, and thus smaller buffers, less delays, and higher utilization can be achieved. Since for a given link, all the QoS parameters, namely, delay, loss and jitter for a stream depend on its queuing dynamics. Similar results are also derived for other sequences such as SoccerWM, News and Simpsons [12]. From the buffering standpoint, instead of buffering highly correlated input traffic directly, our proposed scheme buffers the residuals (errors) of the prediction, a rather "uncorrelated" process resembling white noise, thus requiring much smaller buffer space.

6. Conclusions

Our contribution discussed in this paper are four-fold:

- We have proposed to predict the relative size change of I frames, i.e. s_k. Owing to a smoother s_k, better prediction has been achieved.
- We have proposed to adopt a variable step size to improve the convergence of LMS for video traffic prediction, first using KVSA, and later VSA. Our simulations show that VSA not only incurs small prediction errors but more importantly also achieves fast convergence. This new algorithm converges faster, and hence, tracks scene changes better than FSA.
- We have justified analytically that the spectral density of the prediction error of s_k is rather uncorrelated, resembling white noise, and thus using much smaller buffer space.
- We have proposed a dynamic bandwidth allocation algorithm based on predicted I_k (derived from s_k) that has greatly reduced the renegotiation frequency with a small CLR.

The proposed algorithm is therefore applicable for online real time video transmission.

7. Acknowledgment

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

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

References

[1] ISO/IEC. Generic Coding of Moving Pictures and Associated Audio: Video. ISO/IEC Standard13818-2, 1995.

[2] S. Chong, S. Li, and J. Ghosh, "Efficient transport of real time VBR video over ATM via dynamic bandwidth allocation," IEEE J. Sel. Areas Commun., vol.13, pp.12–23, Jan. 1995.

[3] X. Wang, S. Jung, and J. Meditch, "Dynamic bandwidth allocation for VBR video traffic using adaptive wavelet prediction," Proc. IEEE Intl. Conf. on Communication, vol.1, pp.549–553, 1998.

[4] A.M. Adas, "Using adaptive linear prediction to support real-time VBR video under RCBR network service model," IEEE/ACM Trans. on Networking, vol.6, pp.635–644, 1998.

[5] W. Xu and A.G. Qureshi, "Adaptive linear prediction of MPEG video traffic," 5th Intl. Symp. on Signal Processing and its Application, pp.67–70, 1999.

[6] S. Haykin, Adaptive Filter Theory. Englewood Cliffs, NJ: Pretence Hall, 1991.

[7] M. Hayes, Statistical Digital Signal Processing and Modeling. New York: Wiley, 1996.

[8] R.H. Kwong and E.W. Johnston, "A variable step size LMS algorithm," IEEE Trans. Signal Processing, vol.40, pp.1633–1642, March 1993.

[9] M. Grossglauser, S. Keshav, and D. Tse, "RCBR: A simple and efficient service for multiple time scale traffic," ACM SIGCOMM, pp.219–230, 1995.

[10] S.Q. Li and C.L. Hwang, "Queue response to input correlation functions: Discrete spectral analysis," IEEE/ACM Trans. on Networking, vol.1, pp.522–533, Oct. 1993.
[11] S.Q. Li and H.D. Sheng, "Discrete queuing analysis of multimedia traffic with diversity of correlation and burstiness properties," IEEE Trans. on Communications, pp.1339–1351, April 1994.
[12] H. Zhao, N. Ansari, and Y.Q. Shi, "A fast non-linear adaptive algorithm for video traffic prediction," IEEE Intl. Conf. on Information Technology, Coding and Computing, April 2002.

Hong Zhao received her B.S. degree from Taiyuan University of Technology, and M.S. degree from Xian Jiaotong University, both in Electrical Engineering. After getting her Master degree, she joined Taiyuan University of Technology, Taiyuan, China, and did some research on Signal Processing and Microprocessor. She is now pursuing her Ph.D. degree in Electrical Engineering in New Jersey Institue of Technology. Her research interests include video delivery over network and bandwidth prediction.

Nirwan Ansari received the B.S.E.E. (summa cum laude), M.S.E.E., and Ph.D. from NJIT, University of Michigan, and Purdue University in 1982, 1983, and 1988, respectively. He joined the Department of Electrical and Computer Engineering, NJIT, in 1988, and has been Professor since 1997. He is a technical editor of the IEEE Communications Magazine, was instrumental, while serving as its Chapter Chair, in rejuvenating the North Jersey Chapter of the IEEE Communications Society which received the 1996 Chapter of the Year Award, currently serves as the Chair of the IEEE North Jersey Section, and also serves in the IEEE Region 1 Board of Directors and various IEEE committees. He was the 1998 recipient of the NJIT Excellence Teaching Award in Graduate Instruction, and a 1999 IEEE Region 1 Award. His current research focuses on various aspects of high speed networks and multimedia communications. He authored with E.S.H. Hou Computational Intelligence for Optimization (1997, and translated into Chinese in 2000), and edited with B. Yuhas Neural Networks in Telecommunications (1994), both published by Kluwer Academic Publishers.

Yun Q. Shi has been a professor with the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology, Newark, NJ since 87. He obtained his B.S. degree and M.S. degree from the Shanghai Jiao Tong University, Shanghai, China; his M.S. and Ph.D. degrees from the University of Pittsburgh, PA, all in Electrical Engineering. His research interests include visual signal processing and communications, digital multimedia data hiding, computer vision, applications of digital image processing and pattern recognition to industrial automation and biomedical engineering, multidimensional systems and signal processing. Prior to entering graduate school, he had industrial experience in a radio factory as a principal design and test engineer in numerical control manufacturing and electronic broadcasting devices. Some of his research projects are currently supported by several federal and New Jersey state funding agencies. He is an author/coauthor of more than 100 journal and conference proceedings papers in his research areas and a book on Image and Video Compression for Multimedia Engineering published in 99. He has been an IEEE senior member (93-), the chairman of Signal Processing Chapter of IEEE North Jersey Section (96-), an editorial board member of International Journal of Image and Graphics (99-), a member of IEEE Circuits and Systems Society's Technical Committee of Visual Signal Processing and Communications as well as Technical Committee of Multimedia Systems and Applications (01-). He is currently a co-general chair of IEEE 2002 International Workshop on Multimedia Signal Processing, an IEEE CAS Distinguished Lecturer (02-03). He was a formal reviewer of the Mathematical Reviews (87-99), an Associate Editor for IEEE Transactions on Signal Processing in the area of Multidimensional Signal Processing (94-96), the guest editor of the special issue on Image Sequence Processing for the International Journal of Imaging Systems and Technology, published as vol. 9.4 and vol. 9.5 in 98, one of the contributing authors in the area of Signal and Image Processing to the Comprehensive Dictionary of Electrical Engineering, published in 98. His biography has appeared in Marquis Who's Who in Science and Engineering.