AN IMPROVED TRAFFIC MODELING SCHEME FOR MPEG VIDEO OVER CONTENT DELIVERY NETWORKS

Chu-Chuan Lee [1,2], Chin-Wei Lin [1], and Pao-Chi Chang [1]

[1] Department of Electrical Engineering, National Central University, Chung-Li, Taoyuan, Taiwan 320, R.O.C.
[2] Telecommunication Laboratories, Chunghwa Telecom Co., Ltd., Yang-Mei, Taoyuan, Taiwan 326, R.O.C.
E-mail Address: {cclee, cwlin, pcchang}@vaplab.ee.ncu.edu.tw

Abstract - In this paper, we propose an improved video traffic modeling scheme, Delay-Tolerant Multi-Leaky-Bucket (DTMLB), for video streaming applications over Internet. Based on the original Multi-Leaky-Bucket modeling method (MLB), DTMLB parameterizes the actual traffic cumulative function instead of the empirical envelope function used by MLB. Moreover, DTMLB incorporates different delay tolerance characteristics among I-frames, P-frames, and B-frames of video to improve the overall bandwidth utilization. Simulation results show that DTMLB provides higher modeling accuracy and higher bandwidth utilization than the MLB and the peak rate modeling methods do. Furthermore, in our simulations, DTMLB generates the least burst traffic into the subsequent nodes of networks among all three methods tested.

Keywords - MPEG-4, traffic modeling, token-bucket.

I. INTRODUCTION

With the development of network technologies, it is now possible to transmit high quality video applications over Internet. Up to now, most of these video applications usually use unicast method to transmit the video packets. A unimaginable mass of video streaming flows can easily congest the core network. For resolving the problems and storing the contents more effectively, Content Delivery Networks (CDN) is introduced, and video server farms are distributed to the metro networks [1]. Consequently, the possible congestion bottleneck caused by transmitting video applications over Internet is also shifted to the border switch/router of metro server farms. Fig. 1 shows the CDN architectures with broadband access networks and metro networks.

To manage the available bandwidth more effectively, the border switch/routers need the accurate traffic descriptions and parameters provided by video sources. To date, a rich set of literatures have proposed various video traffic modeling methods to describe the characteristics of video sources, such as token-bucket model [2], self-similar model [3], Markov-modulated model [4], multi-leaky-bucket (MLB) model [5], and Deterministic Bounding Interval length Dependent (D-BIND) Model [6]. Basically, these schemes can be classified to two classes, the stochastic traffic modeling class and the deterministic traffic modeling class.

Stochastic approaches have the merit that they may be used to achieve higher bandwidth utilization by exploiting the statistical properties of video source. However, the implementation complexity that enforces the stochastic traffic characteristics into consideration is high. In contrast, developed under the worst-case concept and description, deterministic modeling methods can promise that no packet is discarded or delayed beyond the guaranteed delay bound. In addition, deterministic models can be parameterized easily such that the implementation is more practical than the stochastic approaches.

To date, however, all related deterministic and stochastic modeling literatures mentioned above have not included the fact that the I- and P-frames can tolerate extra frame delay as compared to B-frames. With these motivations, in this paper we try to re-design the deterministic traffic modeling schemes for improving its modeling accuracy.

The rest of this paper is structured as follows. In section II, we briefly review some related work on the video traffic modeling approaches. In section III, the proposed video traffic modeling scheme, delay-tolerant multi-leaky-bucket (DTMLB), is presented in detail. Section IV presents the simulation environment and results. Finally, section V concludes this paper.

II. RELATED WORK

In case of Integrated Services (Inteserv) of Internet,

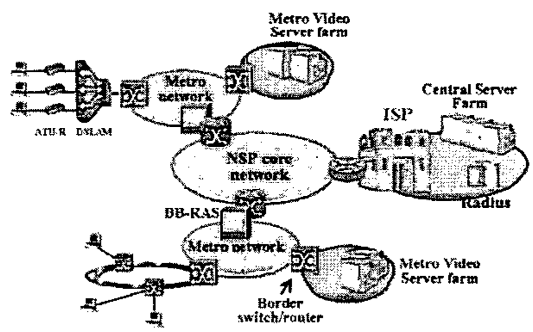

Fig. 1. CDN architectures with broadband access networks and metro networks.

0-7803-7510-6/02/$17.00 ©2002 IEEE

Resource reSerVation Protocol (RSVP) [7] utilizes the token bucket algorithm to describe traffic parameters corresponding to a specific flow of IP packets. The token-bucket model only uses one parameter pair to describe the IP flow. The parameter pair contains the average rate denoted as ρ and the bucket depth denoted as σ [2]. However, token bucket algorithm is not suitable for VBR video traffic. When the network resource is reserved according to the single (σ, ρ) pair, unacceptable delay or loss may occur because of unpredictable traffic burst.

To describe the video traffic characteristics more accurately, Wrege et al. proposed the MLB model which uses a set of parameter pairs $\{(\sigma_i, \rho_i)| \ i = 1, 2,..., n\}$ to characterize the video traffic [5]. Because the MLB model belongs to deterministic traffic modeling class, it uses worst-case description to get the empirical envelope E^* which is a time-invariant function, then it captures the set of parameter pairs from E^* in terms of piece-wise linear concave upper approximation. The concept of worst-case description is presented as follows. If the actual traffic given by a video source is described by an arrival cumulative function A such that $A[\tau, \ \tau+t]$ denotes the traffic arrives in the time interval $[\tau, \ \tau+t]$, an upper bound on A can be given by a traffic constraint function A^* if for all times $\tau \geqq 0$ and all interval lengths $t \geqq 0$, the following condition holds [5]:

$$A[\tau, \ \tau+t] \leqq A^*(t). \tag{1}$$

However, using the parameter pairs given by MLB, the traffic shaper of edge switch/router may need a deeper depth of token pool for the whole VBR video traffic. A deeper token pool may cause large and long traffic burst.

Peak rate description method is the most intuitive method that can provide the guarantee of the lowest transmission delay and no packet loss. However, its low bandwidth utilization also makes it impracticable to be implemented.

III. DELAY-TOLERANT MULTI-LEAKY-BUCKET MODEL

In this paper, an improved video traffic modeling method DTMLB that refines the original MLB model is proposed. The major differences between the proposed DTMLB and MLB are described as follows. 1) To describe the traffic more accurately, we determine and update the traffic parameters every GOP. 2) Because we focus on the video streaming applications, DTMLB determines the set of parameter pairs $\{(\sigma_i, \rho_i)| \ i = 1, 2,..., n\}$ according to the arrival cumulative function A directly, not the empirical envelope E^*. 3) The realistic situation that the delay tolerance characteristics are different among I-frame, P-frame, and B-frame is considered in DTMLB. With these properties, DTMLB can reduce the bandwidth requirement of each flow. We present in detail the DTMLB modeling scheme in the next two sub-sections.

A. Delay-Tolerant Arrival Cumulative Function

Fig. 2 shows the property that the delay tolerance is different among video frames. It presents the case where compressed video frames are transmitted in bit-stream order with arrows mapping frames in bit-stream order to the instants they are first needed in display order. The difference between display order and bit-stream order of video traffic shows that I- and P-frames can tolerate one extra frame delay as compared to B-frames during the transmission [8]. Hence, the video display at the receiver is still sustained even if some of the packets that belong to the I- or P-frame are transmitted in the next frame time of B-frame.

Using the delay tolerance properties mentioned above, the

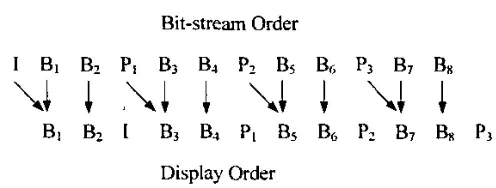

Fig. 2. The difference between display order and bit-stream order of video traffic.

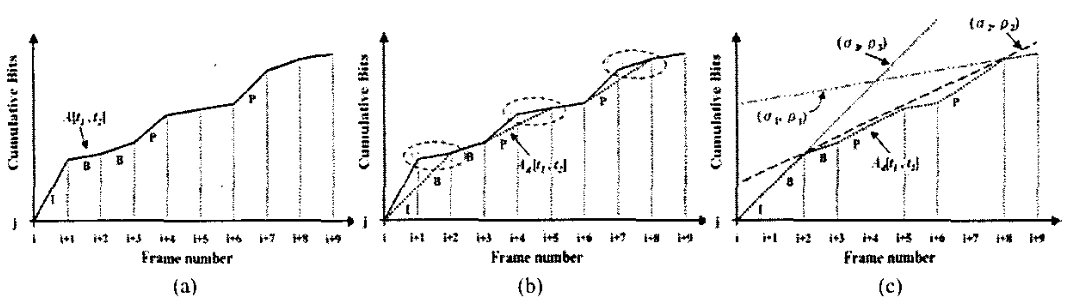

Fig. 3. An example of DTMLB modeling scheme; (a) The original arrival cumulative function A; (b) The proposed delay-tolerant arrival cumulative function A_d; (c) The piece-wise linear concave approximation for A_d.

improved arrival cumulative function denoted as A_d is proposed where $A_d[t_1, t_2]$ means the traffic arrives in the time interval $[t_1, t_2]$ with additional delay tolerant considerations. To describe the function in detail, Fig. 3 discusses a modeling example with one GOP which consists of nine frames. Fig. 3(a) depicts an arrival cumulative function $A[t_1, t_2]$ where $[t_1, t_2]$ is the duration of GOP in this example. Fig. 3(b) depicts the improved arrival cumulative function $A_d[t_1, t_2]$. Since the I-frames and P-frames can tolerate one extra frame time than B-frames, some of the I- or P-frame packets are permitted to transmit in the duration of the subsequent B-frame. Therefore, if the bit stream of any I-frame or P-frame with frame number k arrives up to time $k \times f$, where f is the reciprocal of frame rate, we can obtain $A_d[t_1, k \times f]$ by means of

$$A_d[t_1, k \times f] = \frac{A[t_1, (k-1) \times f] + A[t_1, (k+1) \times f]}{2} \quad (2)$$

B. Piece-Wise Linear Concave Approximation

After obtaining the improved arrival cumulative function, we approximate $A_d[t_1, t_2]$ by means of the piece-wise liner concave upper bound called traffic constraint function $A_c[t_1, t_2]$. The $A_c[t_1, t_2]$ is described in terms of a set of parameter pairs $\{(\sigma_i, \rho_i) | \ i = 1, 2, ..., n\}$ where the traffic constraint in an interval t is restricted to min $\{\sigma_i + \rho_i t\}$. The approximation algorithm used by DTMLB is presented in detail in Fig. 4.

In Fig. 4, V is defined as the total number of frames in a GOP. Moreover, though the required bandwidth of the last B-frame of a GOP is small, it sometimes needs an additional (σ, ρ) pair to describe the concave appeared at the last two B-frames. For reducing the unnecessary parameter pairs, we omit the last one frame time and get t^* in the operations of piece-wise linear concave approximation. The approximation is repeated until $t^* \leqq 0$. The final approximation result of the example that we have discussed early is shown in Fig. 3(c). We notice that the first parameter (σ_1, ρ_1) is only used by the last B-frame, therefore, the (σ_1, ρ_1) can be saved if the computation of t^* in Fig. 4 is introduced. For explaining the algorithm, we continue the example shown in Fig. 3(c). Finally, we obtain two parameter pairs, (σ_2, ρ_2) and (σ_3, ρ_3) in this example.

IV. SIMULATION RESULTS

With the trend of CDN architectures mentioned in Section I, the remote video servers are built in the metro server farm and connect to the same border switch/router. One of the major congestion bottlenecks is at the border switch/router. Therefore, We only focus on the border switch/router for the performance evaluations of video traffics and assume that the bandwidth of metro/access networks is enough such that the delay and loss performances of video traffics are satisfactory within the metro/access networks.

In the border switch/router, we set up the traffic management mechanisms which consist of token bucket traffic shaper [2], dynamic deficit round-robin (DDRR) scheduler [9], and video interleaving controller. According to the traffic modeling parameters (σ_i, ρ_i) provided by the video sources, token bucket traffic shaper smoothes the video traffics coming from video servers. DDRR is used to provide a fair, fast, and effective scheduling to the mass and independent video traffics. To avoid the situation that the I-frames belonging to different video sources arrive at the scheduler simultaneously, we use a simple video interleaving controller. The relationship among these traffic management elements mentioned above is shown in Fig. 5. In our simulations the link bandwidth that connects the border switch/router to the metro networks is assumed as 45 Mbps.

Find_A$_d$ _function ( )

FOR $k = 2$ **TO** $V\text{-}1$

IF *I-frame or P-frame* **THEN**

$$A_d[0, k \times f] = \frac{A[0, (k-1) \times f] + A[0, (k+1) \times f]}{2}$$

ELSE $A_d[0, k \times f] = A[0, k \times f]$

NEXT k

Return

Find_parameter_pairs (*GOP_duration* **)**

$$t^* = GOP\_duration - \frac{1}{frame\_rate}$$

$i = 0$

While $t^* > 0$ **Do**

$i = i + 1$

$$\sigma_i = \max_{0 \le \alpha < t^*} \left\{ \frac{\{t^* A_d[0, \alpha] - \alpha \ A_d[0, t^*]\}}{t^* - \alpha} \right\}$$

$$\rho_i = \frac{A_d[0, t^*] - \sigma_i}{t^*}$$

$$t^* = \min\{\alpha \mid \sigma_i + \rho_i \alpha = A_d[0, \alpha]\}$$

End While

Return

Fig. 4. The algorithms of piece-wise linear concave approximation

We utilize MoMuSys MPEG-4 VM 16 encoder to compress video sequences. We use three video sources, "foreman", "news", and a hybrid video sequence, which consists of "children", "Claire", "Miss America", and "salesman", for testing various scene change situations. However, all connections transmit the same sequence in

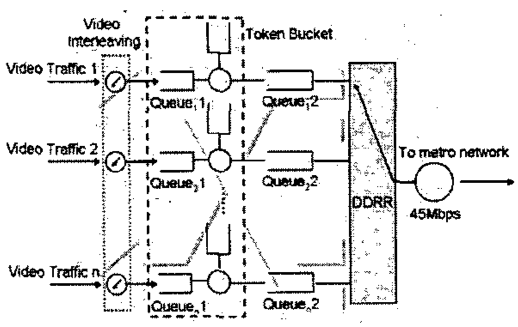

Fig. 5 The simulation environments for proposed DTMLB.

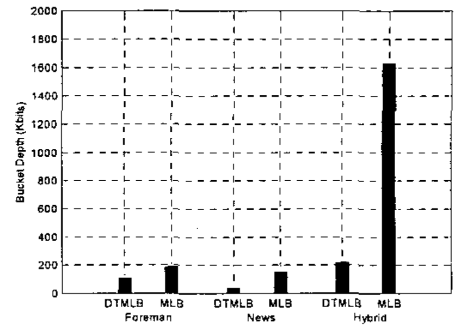

Fig. 6 The comparison of required maximum bucket depth using two modeling methods and three video sources.

each simulation case. All video sequences are in CIF format with512 kbps target rate and 30 frames per second frame rate. All video sequences have 2 B-frames between the I/P-frame and the subsequent P-/I-frame in a GOP. Foreman and news sequences are 300 frames each in length, and the hybrid sequence has a total of 750 frames.

For comparing the modeling accuracy, we introduce the accuracy deviation factor as follow

$$Deviation = \sum_{i=1}^{N} \left| M[(i-1)\times f, i\times f] - A[(i-1)\times f, i\times f] \right| \quad (3)$$

where N is the total number of frames of one video sequence, $A[t_1, t_2]$ is the arrival cumulative function defined in Section II for a video source, and $M[t_1, t_2]$ is the traffic constraint function defined in Section III-B for DTMLB scheme and defined in Section II for MLB method, respectively.

The bucket depth is first compared. Fig. 6 shows the required maximum bucket depth for various video sources, Foreman, News, and Hybrid sequences. It is apparent that the MLB approach always needs a deeper bucket depth than the DTMLB scheme in our simulations. The reason is that the MLB is a deterministic model using worst-case description. Therefore it needs deep bucket depth to pre-store the tokens for later packets. Moreover, when the frame complexity increases, e.g., the hybrid video sequence, the requirement of bucket depth for MLB increases significantly more than the proposed DTMLB scheme. Deep bucket depth may introduce serious traffic bursts to the subsequent nodes of networks. In our simulations, DTMLB are always with stronger capability to avoid unnecessary traffic burst into the networks.

To compare the modeling accuracy between DTMLB and MLB, the accuracy deviations using different video sources, Foreman, News, and Hybrid sequences are shown in Fig. 7. Notice that DTMLB always has smaller accuracy deviation than MLB model in the three simulation cases. Furthermore, when the frame complexity increases, DTMLB provides even more descriptive accuracy than MLB model. Because of the significantly large values, the accuracy deviations by using the peak rate description scheme are shown separately in Fig. 8. As expected, though Peak rate description method can provide the lowest transmission delay and the guarantee of no packet loss, the peak rate description model always has the largest accuracy deviation than the other two modeling approaches.

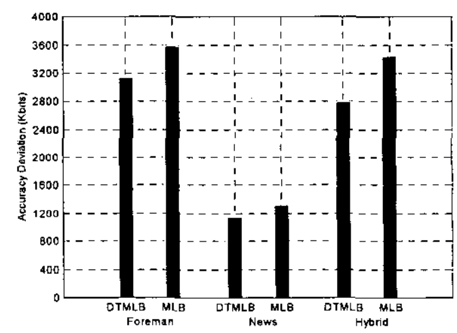

Fig. 7 The comparison of accuracy deviation using two modeling methods and three video sources.

The link utilization of various sequences by using the three different modeling schemes are shown in Fig. 9, Fig. 10, and Fig. 11, respectively. From Fig. 9 and Fig. 10, we observe that the link utilization of DTMLB is higher than MLB slightly. However, when the frame complexity increases such as the hybrid sequence, as shown in Fig.11, the link utilization of MLB decreases substantially. In contrast, the link utilization using proposed DTMLB is always stable in the three simulation cases. Moreover, peak rate description approach always has the lowest link utilization.

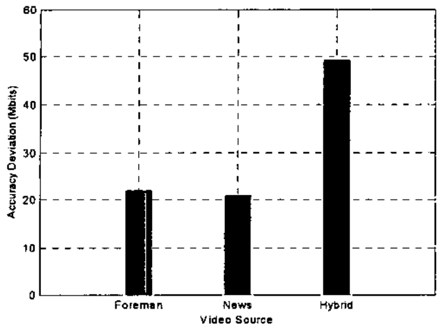

Fig. 8 Accuracy deviation for peak rate description with different video sources.

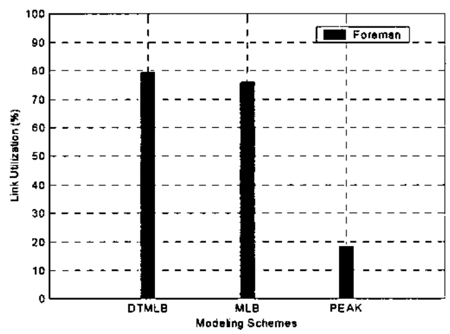

Fig. 9 The comparison of link utilization using the Foreman sequence and three modeling schemes.

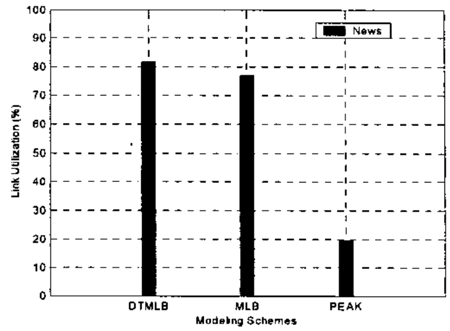

Fig. 10 The comparison of link utilization using the News sequence and three modeling schemes.

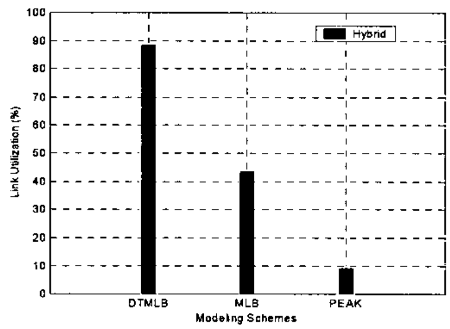

Fig. 11 The comparison of link utilization using the Hybrid sequence and three modeling schemes.

V. CONCLUSION

In this paper, with the consideration of different delay tolerance characteristics among I-frames, P-frames, and B-frames, we propose an improved video traffic modeling scheme, DTMLB, for video streaming applications over Content Delivery Networks. Simulation results show that DTMLB provides better traffic descriptive accuracy and higher bandwidth utilization, particularly for complex video sources, than the MLB and peak rate description modeling methods. Moreover, as shown in simulations, DTMLB provides strong capability to avoid possible additional traffic bursts into the subsequent nodes of networks.

VI. REFERENCES

[1]. Scot Hull, *Content Delivery Networks: Web Switching for Security, Availability, and Speed.* Osborne: McGraw-Hill, 2002.

[2] D. McDysan, *QoS & Traffic Management in IP & ATM Networks*. Osborne: McGraw-Hill, 2000.

[3] V. Frost and B.Melamed, "Traffic Modeling for Telecommunications Networks," *IEEE Commun. Mag.*, vol. 32, no. 3, pp. 70-81, Mar. 1994.

[4] I. Nikolaidis and R. Onvural, "A Bibliography on Performance Issues in ATM networks," *Comput. Commun. Rev.*, vol. 22, no. 5, pp. 8-23, Oct. 1992.

[5] D.E. Wrege, E.W. Knightly, H. Zhang, and J. Liebeherr, "Deterministic Delay Bounds for VBR Video in Packet-Switching Networks: Fundamental Limits and Practical Trade-Offs," *IEEE/ACM Transactions on Networking*, vol. 4, no. 3, Jun. 1996.

[6] E. W. Knightly and H. Zhang, "D-BIND: An Accurate Traffic Model for Providing QoS Guarantees to VBR Traffic," *IEEE/ACM Trans. on Networking*, vol. 5, no. 2, Apr. 1997.

[7] S. Shenker, J. Wroclawski, "General Characteriz- ation Parameters for Integrated Service Network Elements," *RFC 2215, IETF*, Sep. 1997.

[8] W.-T. Tan and A. Zakhor, "Packet Classification Schemes for Streaming MPEG Video over Delay and Loss Differentiated Networks," *Proceedings of Packet Video Workshop 2001*, Apr. 2001.

[9] K. Yamakosho, K. Nakai, E. Oki, and N. Yamanaka, "Dynamic Deficit Round-Robin Scheduling Scheme for Variable-Length Packets," *Electronics Letters*, vol. 38, no. 3, pp. 148-149, Jun. 2002.